Budapest University of Technology and Economics (BME)

Faculty of Transportation Engineering and Vehicle Engineering (KJK)

Department of Transport Technology and Economics (KUKG)

Optimization of Multimodal Travel Chains

PhD thesis

2016

Domokos Esztergár-Kiss

Supervisor: Dr. Csaba Csiszár
Nyilatkozat

Alulírott Esztergár-Kiss Domokos kijelentem, hogy ezt a doktori értekezést magam készítettem, és abban csak a megadott forrásokat használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos tartalomban, de átfogalmazva más forrásból átvertem, egyértelműen a forrás megadásával megjelöltem.
Budapest, 2016.05.12.

Esztergár-Kiss Domokos

Declaration

Hereby, I, Domokos Esztergár-Kiss declare that the present PhD Thesis is my own work, and I utilized only the sources indicated within. All parts taken from other works used word by word, or in a reedited way keeping the original contents, have been unambiguously marked by a reference to the source.
Budapest, 12.05.2016.

Esztergár-Kiss Domokos
# Table of content

1 MOBILITY RESEARCH INTRODUCTION  8
1.1 MOBILITY ISSUES  8
1.2 MOBILITY DEFINITIONS  10
1.3 MOBILITY SOLUTIONS  12
1.4 RESEARCH OBJECTIVE  14

2 EVALUATION METHOD DEVELOPMENT FOR MULTIMODAL JOURNEY PLANNERS  16
2.1 DEFINITION OF THE FRAMEWORK OF ASPECTS  17
2.2 DEFINITION OF THE USER GROUPS  19
2.3 ELABORATION OF THE EVALUATION METHOD  20
2.4 REVIEW OF THE JOURNEY PLANNERS  22
2.5 EVALUATION OF THE JOURNEY PLANNERS  26
2.6 EVALUATION OF HUNGARIAN ONLINE TRAVEL INFORMATION SERVICES  31

3 ANALYSIS AND CREATION OF USER GROUPS BASED ON PREFERENCES AND EVALUATION ASPECTS  33
3.1 DEFINITION OF THE SURVEY METHOD  35
3.2 ANALYSIS OF SURVEY RESULTS  36
3.3 CREATION OF WARD USER GROUPS  41
3.4 EVALUATION OF JOURNEY PLANNERS WITH USER GROUPS  46
3.5 ELABORATION OF A SURVEY METHOD USING FUZZY AHP  49
3.6 EVALUATION OF THE SURVEY METHOD BENEFITS  55

4 OPTIMIZATION METHOD DEVELOPMENT OF DAILY ACTIVITY CHAINS  57
4.1 DEFINITION OF ACTIVITY CHAINS AND CONSTRAINTS  58
4.2 ELABORATION OF THE OPTIMIZATION ALGORITHM  61
4.3 EXTENSIONS OF THE OPTIMIZATION ALGORITHM WITH POI SEARCH AND GENETIC ALGORITHM  66
4.4 EVALUATION OF THE SIMULATION RESULTS WITH EXTENSIONS  70
4.5 EVALUATION OF ACTIVITY CHAIN OPTIMIZATION ALGORITHM BENEFITS  76

5 MODELING THE INFORMATION SYSTEM OF ACTIVITY CHAIN OPTIMIZATION  81
5.1 DEFINITION OF CLASSIFICATION AND OPTIMIZATION PARAMETERS  83
5.2 ELABORATION OF THE UTILITY FUNCTION  89
5.3 MODELING THE SYSTEM ARCHITECTURE  93
5.4 DEVELOPMENT DIRECTIONS OF INFORMATION SERVICES  98

6 THESIS SUMMARY  101
6.1 THESIS 1. EVALUATION METHOD DEVELOPMENT FOR MULTIMODAL JOURNEY PLANNERS  101
6.2 THESIS 2. ANALYSIS AND CREATION OF USER GROUPS BASED ON PREFERENCES AND EVALUATION ASPECTS  102
6.3 THESIS 3. OPTIMIZATION METHOD DEVELOPMENT OF DAILY ACTIVITY CHAINS  103
6.4 THESIS 4. MODELING THE INFORMATION SYSTEM OF ACTIVITY CHAIN OPTIMIZATION  104
List of figures

Fig 1 Passenger transport demand tendency [130] ................................................................. 9
Fig 2 Mobility models on different levels ................................................................................. 11
Fig 3 Structure of the dissertation ......................................................................................... 15
Fig 4 User groups ................................................................................................................... 20
Fig 5 Process of the evaluation of multimodal journey planners ........................................ 21
Fig 6 Comparison of general evaluation numbers and average evaluation numbers ........... 29
Fig 7 Comparison of evaluation numbers of Hungarian journey planners ......................... 32
Fig 8 Spatial representation of the average evaluation numbers ............................................. 32
Fig 9 Analysis and definition of user groups .......................................................................... 34
Fig 10 Process of evaluation using survey results .................................................................... 34
Fig 11 Distribution of the user groups ..................................................................................... 36
Fig 12 Normalized ranking values of the main aspects for each user group ......................... 38
Fig 13 Comparison of original and ranked answer values for the user groups .................... 39
Fig 14 Answer values of single aspects in the original and in the ranked case ....................... 40
Fig 15 Answers given by the users for two highly correlated single aspects each ............... 42
Fig 16 Dendrogram of the Ward user groups ........................................................................ 44
Fig 17 Correspondence between the original user groups and Ward user groups .............. 44
Fig 18 Main aspects and Ward user groups ............................................................................ 46
Fig 19 Calculation of transportation shares of Ward user groups ......................................... 47
Fig 20 Evaluation of multimodal journey planners without user groups (general evaluation number) and with Ward user groups (average evaluation number) ........ 48
Fig 21 Process of evaluation using Fuzzy AHP method .......................................................... 51
Fig 22 Differences between original weights and AHP weights, respectively original weights and Fuzzy weights ................................................................. 55
Fig 23 Comparison of the results: original weights, AHP weights and Fuzzy weights ......... 56
Fig 24 Database structure of activity chains .......................................................................... 59
Fig 25 Temporal reachability of the demand points ............................................................... 61
Fig 26 Model of the activity chain optimization algorithm .................................................... 62
Fig 27 Presentation of the flexible demand points ................................................................. 65
Fig 28 Comparison of POI database and Overpass API search times ................................... 67
Fig 29 Optimization with inner GA and multiobjective GA ................................................... 68
Fig 30 Model of activity optimization chain with the extensions .......................................... 69
Fig 31 Example daily activity chain with fix and flexible demand points using car (visualization) ........................................................................................................... 71
Fig 32 Example daily activity chain with fix and flexible demand points using public transport (visualization) ...................................................................................... 72
Fig 33 Example daily activity chain with fix and flexible demand points combined (visualization) ........................................................................................................... 74
Fig 34 Processing times of the extended optimization algorithm ........................................... 75
Fig 35 Comparison of processing times of the original and the extended optimization algorithm .................................................................................................................. 76
Fig 36 Frequency of trips with particular purposes ............................................................... 78
Fig 37 Operational model of the information system of the activity chain optimization ........ 82
Fig 38 Grouping of parameters by parameter type ................................................................. 83
Fig 39 Identification of parameters by component type ......................................................... 84
Fig 40 Aggregated weights and related optimization parameters ......................................... 91
Fig 41 Elements of the utility function................................................................. 92
Fig 42 Model of the system architecture ............................................................ 93
Fig 43 Model of the menu structure of the mobile application .......................... 97
Fig 44 Development directions of information services ................................. 98
Fig 45 Spatial representation of the new organizational structure [106]............. 119
Fig 46 Results of single aspects (route planning services)............................... 122
Fig 47 Results of single aspects (booking and payment) ................................. 123
Fig 48 Results of single aspects (handled data)............................................... 123
Fig 49 Results of single aspects (comfort service information) ....................... 124
Fig 50 Results of single aspects (supplementary information) ......................... 124
Fig 51 Answers by the each user groups for two highly correlated single aspects ... 126
Fig 52 Centroids of sub-districts derived from Google Geocoding API ............ 130
Fig 53 Graphical representation of the graph .................................................. 131
Fig 54 Connection between the districts sorted by time and purpose ............... 132
Fig 55 Connection between the districts sorted by time and transportation mode..... 133
List of tables

Table 1 Framework of aspects regarding multimodal journey planners .................................. 17
Table 2 Classification of journey planners .............................................................................. 23
Table 3 Top solutions of journey planners .............................................................................. 25
Table 4 Evaluation of journey planners ................................................................................... 27
Table 5 Weights of main aspects according to user groups and transportation shares ............. 28
Table 6 Evaluation of journey planners with consideration to the user groups (%) ................. 29
Table 7 Result of the sensitivity analysis .................................................................................. 30
Table 8 Normality analysis of the single aspects ...................................................................... 37
Table 9 Features of Ward user groups ..................................................................................... 45
Table 10 Weights of main aspects according to Ward user groups and transportation shares ............................................................... 47
Table 11 Comparison of changes in the evaluation using estimated and Ward weights ............... 48
Table 12 Example comparison table with main aspects ............................................................ 52
Table 13 Calculation of AHP values ........................................................................................ 53
Table 14 Calculation of Fuzzy values ...................................................................................... 54
Table 15 Flexibility of priorities ............................................................................................... 63
Table 16 Structure of the POI database with example data ...................................................... 66
Table 17 Example daily activity chain input data ..................................................................... 70
Table 18 Example daily activity chain with fix and flexible demand points using car (data) ........................................................................................................................................ 70
Table 19 Pareto optimal solutions when using car .................................................................... 71
Table 20 Example daily activity chain with fix and flexible demand points using public transport (data) ................................................................................................................ 72
Table 21 Pareto optimal solutions when using public transport ............................................... 73
Table 22 Example daily activity chain with fix and flexible demand points using combined modes (data) ................................................................................................................... 73
Table 23 Pareto optimal solutions when using combined modes ............................................ 74
Table 24 Comparison of the simulation results ......................................................................... 74
Table 25 General statistics of the datasets ............................................................................... 79
Table 26 Description of classification parameters .................................................................... 85
Table 27 Description of general optimization parameters ....................................................... 86
Table 28 Description of comfort optimization parameters of trip component type .................. 88
Table 29 Description of comfort optimization parameters of location component type .......... 89
Table 30 Update frequencies of external databases ................................................................. 96
Table 31 Evaluation of Hungarian online travel information services ...................................... 121
Table 32 Evaluation of Hungarian online travel information services with user groups .......... 121
Table 33 Single aspect pairs of high correlation ...................................................................... 125
Table 34 Proportions and AHP values of main aspects for each user group .............................. 128
Table 35 Numerical comparison: original weights, AHP weights and Fuzzy weights .............. 129
Glossary

AHP: Analytic Hierarchy Process
ANOVA: ANalysis Of VAriance
API: Application Programming Interface
BMT: Balázs Mór Terv
CART: Classification And Regression Trees
CBA: Cost Benefit Analysis
EEA: European Environment Agency
EKFS: Unified Transport Development Strategy
FCD: Floating Car Data
GA: Genetic Algorithm
GTFS: General Transit Feed Specification
GTFS-RT: General Transit Feed Specification-Real-Time
ITS: Intelligent Transportation Systems
KPI: Key Performance Indicator
KSH: Central Statistics Office
MCA: MultiCriteria Analysis
MCDM: Multiple-Criteria Decision-Making
OSM: Open Street Map
NKS: National Transport Strategy
POI: Point Of Interest
SUMP: Sustainable Urban Mobility Plans
TSP: Traveling Salesman Problem
TSP-TW: Traveling Salesman Problem with Time Window
VRP: Vehicle Routing Problem
WTP: Workplace Travel Plans
XML: eXtensible Markup Language
1 Mobility research introduction

The growing mobility demands define new and innovative directions of transportation research, as they are both the preconditions and consequences of economic development. It is resulted in increased challenge towards passenger transport systems. The journey times of passengers should be decreased, while the quality of transportation services should be increased and the whole transportation related processes should be optimized.

In order to reach these aims, an integrated approach is to follow when carrying on transportation research. The fast development of Intelligent Transportation Systems (ITS) enables intervention in processes of transport, which contributes to the enhancement of service quality. To assist that strategic developments are realized in those areas, where they can provide most efficient functioning, a comprehensive evaluation of transportation information services has to be performed. Through the results of the evaluation the user preferences can be derived. In order to realize the requested features, optimization methods can provide information to passengers about how to plan and realize their journeys in the most personalized way. With the development of optimization methods the service quality can be significantly enhanced, especially in urban areas, where the increasing population and therefore the extensive use of transportation means tends to generate problems.

Having up-to-date ITS systems and a high amount of data does not implies automatically provide a better transportation information service. In order to use ITS systems and available data efficiently, results from more research fields have to be integrated (e.g. information technology, transportation infrastructure, passenger behavior, optimization methods. Considering this paradigm shift, smart systems are to be created, which provide comprehensive solutions through different aspects of transportation. A complex smart system includes personalization and adaptivity, which corresponds to the dynamically changing mobility demands of the users.

1.1 Mobility issues

Mobility in an urban environment plays an important role, as 74% of travelers in Europe live in cities, which value will even rise to 80% by 2030 [66]. This means that a significant proportion of the total mobility is and will be concentrated in urban areas. Although the modal-split has a more balanced situation in European cities with 60-80% share of sustainable modes, but the possibility of congestion and low services levels is higher, than in rural areas. Focusing on urban areas, the following options were formulated to improve transport in a report by the European Environment Agency [66]:

- avoid need of travel through efficient urban planning and demand management,
- shift demands to more sustainable transport modes,
- improve the environmental performance of vehicles.
The basis of the growing interest in mobility and its services can be derived from future trends, which could evolve to a serious problem, if not handled timely. Interestingly not the number of average daily trips is the source of problems, as it did not change significantly over the decades. This fact is supported by an international research of European driving patterns, however only for car trips [133]. The average number of daily trips was 2,5, which considering 80% the average modal-split of cars in Europe surveyed by the European Environment Agency (EEA) [67] results in 3,125 trips per day. Although the number of average daily trips and the time spent with travelling did not change significantly, the distance travelled did. According to an UK survey it has increased by 10% in the last 10 years [40]. This coincides with the report of EEA, where they claim that total passenger transport demand (pkm) increased by 11% between 2000 and 2011 [66].

Based on the prediction of the National Transport Strategy (NKS) document [130], the total passenger transport demand will increase in Hungary only slightly, which is caused by the negative tendency of population growth and by the aging society (Fig 1). But the demand growth is significant in case of cars, which is 15% by 2050, while public transportation modes (bus and railway) have to suffer from a drop of 8% and 5%. Similar trends are observed all over Europe [66], which leads to an undesired shift of modal-split.

This tendency can be reverted by applying optimized demand management strategies. Considering this general guidelines are provided by Sustainable Urban Mobility Plans (SUMP), which include best practice and tools to support development and implementation of mobility plans [68]. Based on the figures of the EEA report work and education motivated trips are 25-50% of all trips, which are handled by Workplace Travel Plans (WTP) [181]. WTPs launch travel surveys, promote sustainable transportation modes (e.g. walking and biking) and suggest flexible working practice. This means that at least 50% of trips are non-routine activities. These activities can be divided into short-distance trips (e.g. shopping, visits sport) and long-distance trips (e.g. business trip, holiday). In the case of non-routine activities the possibility of some flexibility is present, which represents a research potential.
1.2 Mobility definitions

In order to build a common vocabulary of technical terms, the most used ones were defined. The word “activity” means using a service in a predefined location. For movements between locations several terms are used in the scientific literature, as trip, journey, travel and locomotion. In case of changing locations between two consecutive activities the term “trip” is used, while the term “journey” appears as a synonym for trip mostly connected with the word “planner”. To describe several movements during a day for different types of travelers with various transportation modes the most appropriate term is “travel”. The term “activity chain” or “travel chain” can be defined as a set of trips realized by a traveler using public, private or soft transportation modes [78].

The term “mobility” represents the connection among activities in space and in time and is realized by activity chains. Mobility is a measure for the effectiveness of moving from one location to another, where an activity is realized. Therefore the aim of mobility is to gain access to a requested service, not the travelling process and location change itself.

The measurement of mobility is a complex issue, as many factors influence it, e.g. transportation network, service quality, network usage figures, user behavior and specific aspects of passengers using the transportation system. But it is essential to define and use mobility in the research, because it facilitates to determine patterns and activity chains of the passengers. Mobility has also an important social and economical impact, namely improved mobility means faster and more comfortable connections on the transportation network among people. Mobility demands are mainly influenced by following factors [66]:

- urban form (size, density and design),
- socio-economic factors,
- quality of transport infrastructure and services.

The easiest way of controlling the influencing factors is to enhance the quality of transportation. But in order to perform that, the elements of transport quality have to be defined, which is regulated by European standards [47]:

- service quality: availability, access,
- transport quality: information, travel times, handling passengers, comfort, safety,
- environmental quality: environmental effects.

In order to realize improvements in transport quality, mobility has to be handled in an abstract way. Mobility can be modeled on different levels, where land use and ownership models, travel demand models and traffic flow models can be defined (Fig 2).

- Land use models forecast the distribution of the population and activity locations in an area by modeling location choice processes. Ownership models determine the share of persons with car availability and with season tickets.
- Travel demand models reproduce the decision making processes of individuals leading to movements in the transport network. In passenger transport these decisions cover activity choice, destination choice, mode choice, departure time choice and route choice.
Traffic flow models simulate the flow of vehicles and pedestrians considering the interactions between the moving objects. In car transport traffic flow models replicate decisions of drivers describing the choice of travel speed, lane and preferred distance to the vehicle ahead. [78].

Fig 2 Mobility models on different levels

Considering travel demand models many research papers deal with the development of new modeling frameworks. Ben-Akiva and Lerman established the basics of current travel demand models, in his book both theoretical background and application possibilities were discussed [13]. Artenze and Timmermans presented a framework for modeling dynamic activity choices, which provides ideas for activity travel pattern analysis [8]. Bierlaire investigated theoretical aspects of discrete choice models, where some general assumptions were presented about the characteristics of the decision maker, the alternatives (possible options), attributes of the alternatives and the decision rules, which define the decision process [17].

When realizing mobility demands, journey planners are used to provide support for travelers. Again here the scientific literature and practical implementations use the following terms with similar content as synonyms: journey planner, trip planner, travel planner. I use the term “journey planner”, which provides information for passengers when planning their trips between two consecutive locations.
A good definition of multimodal travel planners is presented by Gentile and Nökel [78]: “Multimodal travel planners are front-end-back-end computer systems, which provide a traveller the best itinerary according to several parameters characterizing an intermodal passenger transport journey. Multimodal travel planners provide better modal integration and more sustainability by enabling travellers to select the most suitable combination of transport modes for the journey and could lead to an increase use of public transport, cycling or walking in urban environment.”

As the described real multimodal journey planning is only partially implemented in current transportation services, I define “multimodal journey planner” as a journey planner, which plans for several transportation modes, however not necessarily in a combined way.

Researchers in Hungary also proposed new ideas in this field. Kövesné [105] presented an evaluation method of urban passenger transportation systems based on service quality parameters, Csiszár [33] elaborated the model of integrated intelligent passenger information systems and Tóth et. al. [162] described main features of passenger information systems. Furthermore Csiszár and Földe provided analysis methods of integrated information systems [34], route planning evaluation [74] and traveler behavior [73]. Lindenbach et. al. [112] described the application possibilities of new ITS methods, Berki et. al. [15] presented an analysis method of public transport networks. On a more practical level Siegler [148] introduced a new multimodal route planner paradigm, which was implemented as Útvonalterv application.

1.3 Mobility solutions

In transport strategy documents of the European Union the idea of smart mobility solutions was also formulated. In 2011 in the White Paper general directives of ITS were published [176]. The aim of this document is fostering mobility, the integration of transportation networks and the decrease of emission. Furthermore the importance of ITS for vehicles and traffic control is emphasized, as one key factor of sustainable development. They say that “the paramount goal of European transport policy is to help establish a system that underpins European economic progress, enhances competitiveness and offers high-quality mobility services while using resources more efficiently.“ Furthermore the Green Paper on urban transport proposes the development of more intelligent transport systems to improve efficiency and intermodality [84].

Additionally, national strategic documents support the development of ITS systems in Hungary. The Hungarian Transport Policy (Magyar közlekedéspolitika) in 2003-2015 already mentions the establishment of more efficient transportation networks through ITS applications [92]. The Unified Transport Development Strategy (EKFS, Egységes Közlekedésfejlesztési Stratégia) in 2007-2020 defines its general aims, such as the optimization of passenger transport, the synergy of different transportation modes and providing comfortable and safe service using intelligent technologies [165]. The National Transport Strategy (NKS, Nemzeti Közlekedési Stratégia) defined its key strategic goals, which are the improvement of infrastructure and development of transport services [130].
Lately the Transport Development Strategy of Budapest (BMT, Balázs Móri Terv) in 2014-2030 put emphasis on integrated and sustainable transport development, the intelligent usage of the infrastructure and good information provisions for the passengers [21]. Furthermore a new strategy concerning public transport development directions has been announced in 2015 [170]. Efforts are made to unify service levels of operators and provide nationwide passenger information services, which are integrated and are focused on user experience.

Not only strategic documents describe the importance of advanced ITS systems, but the Horizon 2020 research framework programme also proposed calls concerning this issue [88]. An own area was dedicated to Intelligent Transportation Systems, where the optimal use of transport data as a basis for smart mobility solutions was one of the possible calls. Among others an expected impact of optimal route and schedule development was formulated. Many EU funded projects covered the issue of multimodal travel planners and smart mobility solutions. Furthermore in further calls innovative concepts, systems and services towards mobility as a service are requested, where multimodal information and planning systems should serve passengers’ demands.

Easyway ITS development project was running in 2007-2012, that has a pillar based on the development of travel information services, specially emphasizing the need of creating a multimodal journey planner and information portal [46].

In 2010-2013 i-TOUR: intelligent Transport system for Optimized URban trips [94] dealt with the promotion of public transport usage. An open framework was developed to provide intelligent, multimodal services, with the combination of external information (e.g. road conditions, weather). But the most interesting feature was the locations based service with personalized preferences, which provides rerouting possibility during trips.

Smart-way project developed Galileo based navigation in public transport systems with passenger interaction in 2010-2012, and provided location based services, real-time information and navigation for public transport networks [151]. The applications promises decrease of time loss caused by disturbances through a real-time and intelligent re-planning feature.

METPEX: MEasurement Tool to determine the quality of Passenger EXperience in 2012-2015 aimed to develop and evaluate a standardized tool to measure passenger experience across whole journeys [122]. The project collects the experience of passengers about service quality through surveys and interviews.

Furthermore a tender was announced in 2011 in correspondence with the long-term transportation strategy of the European Union, called the best “smart multimodal journey planner” [150]. But the question arises, on which basis can be the developed applications evaluated and compared to each other. What aspects should be chosen and how personalized user requirements should be taken into account.

The majority of listed publications in transportation research and several own papers are written in English, therefore it was decided to write the dissertation in English. This also provides a wider dissemination of the research results in Europe and may initiate a constructive debate with other researchers abroad.
1.4 Research objective

Technological development and changing mobility requirements, which are caused by societal changes, set new scientific challenges towards the researchers. Since the capacity of transport systems is limited, the increase of passenger transport performance is primarily obtainable by the enhancement of the modal share of public transportation and the introduction of new (telematics based) transportation services, which are more efficient in term of serving personalized needs. Travelers tend to choose public transport, if they experience high transport quality (e.g. clean and modern vehicles, punctuality, guaranteed transfers and personalized route-planning service). To reach this goal measures are to be taken, as soft measures (e.g. computer aided route planning, location-based information services during the journey) or hard measures. These measures are often called management or infrastructure type measures.

The aim of this research was to provide solutions in the field of soft measures, where the focus was set on the improvement of the most relevant transport service features, such as information and time. This can be obtained by the application of evaluation results of information provision services (e.g. online journey planners). The optimization is performed by developing optimization methods regarding activities and trips of travelers, where utility function is the total travel time. The order of the activities for each traveler is set with consideration to the constraints, the maximum number of activities, and the least travel time.

The aims correspond to management type strategic objectives, especially to the concept of NKS [130], which prefers the development of integrated planning of travel chains and the usage of cutting edge ITS solutions. Based on the discussed research objectives, the following research questions were defined:

- What kind of information services are needed by the users before and during traveling? How to provide this information to the travelers? How much multimodal journey planners correspond to the information needs and expectations of users?
- What differences can be revealed among user groups regarding journey planning aspects? How can be user groups created? What are the most common and different needs and expectations of specific user groups? How can new survey methods be applied for the evaluation of multimodal journey planners?
- How can users optimize their daily activity chains regarding travel time? How flexible are activities within an activity chain? How can the optimization algorithm provide services almost in real-time? How can the benefits of the optimization algorithm be measured?
- What parameters are to be applied during the optimization of activity chains? How can the utility function be defined? How can be the optimization algorithm realized in a real system? What are the main development directions of information systems and services?
In order to answer the research questions in integrated and systematic approach, the following structure was created (Fig 3). The numbers in the figure represent chapters of the dissertation.

- My aim was to elaborate a comprehensive method for the evaluation of multimodal journey planners, to compare and evaluate several relevant journey planners.
- In order to obtain a detailed insight into preferences of user groups, I aimed to analyze differences and similarities of user groups regarding the evaluation aspects of multimodal journey planners. As well as to create such new user groups, whose members evaluate aspects in the most similar way. I also aimed to investigate the applicability of new survey methods to calculate weights regarding the user groups and main aspects.
- My goal was to develop such an optimization method for daily activity chains, which can reorganize activities and thus total travel times are reduced. With the development of the optimization algorithm I aimed to provide a service, which considers several constraints and optimization parameters at the same time. I also intended to analyze the benefits of the optimization of activity chains.
- I aimed to define and classify parameters of activity chains. I wanted to identify the utility function of activity chain optimization and to model such a system architecture, which includes the operational model of an application to be realized. Finally, I outlined the future development directions of information services.
2 Evaluation method development for multimodal journey planners

In order to provide high service quality in passenger transportation systems, the comparability of the information services has to be enabled. For that the aspects of comparison and an evaluation methodology has to be elaborated. Travel information can influence the decisions of the passengers, who want to reach their destinations in the most favorable way according to their personal preferences. This usually means the least possible travel time, the minimum number of transfers or using only low-floor vehicles in case of disabilities. Travelers need user friendly information systems, which can provide information before and during the trip considering combined transportation modes. These information services, which are multimodal journey planners, are nowadays available on the internet. The quality of these services can be evaluated and compared, if a framework is present, which contains all necessary aspects of journey planning and information provision. Therefore some recent research work was reviewed, so that a comprehensive framework of aspects and evaluation method can be developed.

In the broader context Lupo [114] pointed out that quality level of services has to be constantly controlled and evaluations should reflect the viewpoints of the users. Hüging et. al. [91] dealt with the benefits of sustainable mobility measures, and provided a comprehensive review of assessment methods. Shang et. al. [147] described a comprehensive and flexible evaluation method for selection of transportation projects. Browne and Ryan [19] compared different evaluation methods, where they found that MCA (MultiCriteria Analysis) methods are the most suitable for quantitative assessments. Beira et. al. [14] tried to combine two evaluation methods (MCA, MultiCriteria Analysis and CBA, Cost Benefit Analysis) in order to assess sustainable mobility.

In the paper of Mátrai et al. [119] transport management measures were described in order to enhance sustainable transportation modes. Their solutions helped a more effective way of transport organization. One measure is the integration of information of all transport modes, which enhances the quality of public transportation. Csiszár et. al. [33] gave guidelines, how to create an integrated telematics system, which can influence passengers and provide them actual information.

According to Grotenhuis et. al. [85] travel information is one factor, which contributes to the quality of public transport. They found that especially pre-trip information is essential for the passengers, and they consider travel time as the most important parameter of quality, but interchanges, waiting times, alternate routes and maps are also crucial. Leviakangas [109] claimed that the value of information contributes to high quality ITS services. The importance of real-time and event driven information was emphasized. A framework of attributes was established, which contains different kinds of information for different providers (e.g. content provider, service provider).

Several concepts have been published which contain dynamic data handling, alternative route planning and personalized recommendations. Dia and Panwai [41] elaborated an agent-based approach in order to present a real-time transport advice system. The aim was to influence traveling behavior with information based on the
personal characteristics (e.g. age, gender) of the passengers, route attributes (e.g. travel time and distance) and trip characteristics (e.g. work, shopping). In the paper of Kim et al. [102] a dynamic, real-time data based route planning is performed. The value of travel information is emphasized, where event-based (e.g. accidents) and weather information was used in order to develop a decision making procedure. Nadi and Delavar [129] realized a personalized route planning model, which presents a decision-making tool. The different routes were pairwise compared, and weighted according to personal preferences (e.g. travel time and distance).

Zhang, Artenze and Timmermans [162] described a framework of a personalized multimodal traveler information system, which includes real-time information, personal preferences and multimodal route planning. Their modular system can give advice to the passengers and can handle unforeseen events. WiseTrip application [154] uses personalized information for multimodal journey planning. The passengers may set their preferences, and would get not only specific routes, but notifications and alerts of possible disturbances of the trip. Thus they may reschedule their trip and choose an alternative route. Many valuable developments are to be found, but no comprehensive aspects regarding journey planners or the evaluation of these services was present in the literature.

2.1 Definition of the framework of aspects

As a first step the information services were identified, which are mostly important for the passengers. To the adequate functioning of a multimodal journey planner a well manageable, user friendly interface and personalized adjustments (e.g. maximal number of transfers or preferred transport mode) are needed, which can provide some benefits for the users (e.g. time saving). In order to realize these prerequisites reliable and actual data are needed. Also the parameters have to be determined, which influence the „goodness“ of the offered journey plans. These are beside the personal comfort the social and environmental effects.

<table>
<thead>
<tr>
<th>1. Route-planning services</th>
<th>2. Booking and payment</th>
<th>3. Handled data, operational features</th>
<th>4. Comfort service information</th>
<th>5. Supplementary information</th>
</tr>
</thead>
<tbody>
<tr>
<td>ways of data input</td>
<td>tariff information</td>
<td>static data</td>
<td>services at the stations/stops</td>
<td>environmental impacts</td>
</tr>
<tr>
<td>planning aspects</td>
<td>method of booking and payment</td>
<td>semi-dynamic data</td>
<td>services on board</td>
<td>information in foreign languages</td>
</tr>
<tr>
<td>displayed data</td>
<td>payment options</td>
<td>dynamic and estimated data</td>
<td>additional services</td>
<td>customer service</td>
</tr>
<tr>
<td>perspicuity of displayed data</td>
<td>personal data</td>
<td></td>
<td>information of equal opportunity</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Framework of aspects regarding multimodal journey planners
I have defined the most important aspects of multimodal journey planners considering KPIs (Key Performance Indicator) of the ITS Action Plan [65] and features being most probably required by the users (Table 1). The detailed scoring of the aspects is presented in the following list in brackets. Maximal 10 points could be given for each single aspect. Considering the certain features of the aspects in most cases:

- 2 points were given, if the feature was perfectly implemented and well usable,
- 1 point was given, if the feature was somehow implemented or partially usable,
- 0 point was given, if the feature was not implemented or was unusable.

1. Route-planning services:
   - **ways of data input**: address (2), name of stop (2), institutions and service facilities (2) (e.g. museums, restaurants, swimming pools, cinemas, offices, schools), GPS coordinates (2), pointing out on the map (2);
   - **planning aspects**: departure and arrival time (2), duration and costs (2), number of transfers and walking distance (2), other aspects (4) (e.g. preferred transportation mode, P+R, B+R, crowding);
   - **displayed data**: travel duration and distance (2), transfer information and location plans (2), waiting times (2), walking time and distance (2), alternative routes (2);
   - **perspicuity of displayed data**: compact design (2), easy understanding (2), visualization on the map (6) (e.g. details, zoom function, displaying transport lines).

2. Booking and payment:
   - **tariff information**: zones (2), prices (2), reduced fares separately (2) (e.g. tabular view), fee of the entire travel chain (2), calculation method (2);
   - **method of booking and payment**: way of data input (4), in how many steps (2), what kind of data is needed (2), possibility of choosing seats (2);
   - **payment options**: types of bank cards (2), payment per mobile phones (2), transaction fees (2), types of vouchers (4) (e.g. SMS, code per e-mail, paper ticket printed at home/at the station/sent by post).

3. Handled data, operational features:
   - **static data**: timetable for a given route and/or date (6), travel conditions and rules (2) (e.g. animals, luggage); export features (2) (e.g. PDF, printing);
   - **semi-dynamic data**: information about deviations (2), list of planned restrictions (4), visualization of planned restrictions (4);
   - **dynamic and estimated data**: information about current and extraordinary traffic situations (4) (e.g. weather conditions, accidents), usage of crowd sourcing data (2), deviation from the timetables (2), calculation of the probable impacts of the extraordinary traffic situations (2) (e.g. alternative routes);
   - **personal data**: creating a profile (2), setting personal preferences (4), saving searches and favourites (2), personalized offers (2).
4. Comfort service information:

- **services at the stations/ops:** Wi-Fi (2), luggage storage (2), other services (6) (e.g. newsagent’s, bakery, car sharing);
- **services on board:** Wi-Fi (2), electrical supply (2), information about dining opportunities (2), other services (4);
- **additional services:** weather forecast (2), booking a room (2), car rental (2), sightseeing (2), opening time of the shops (2).

5. Supplementary information:

- **environmental impacts:** degree of air pollution (6) (e.g. CO₂), energy consumption (2), comparison of transport modes and travel chains (2);
- **information in foreign languages:** how much of the homepage is translated (6), number of foreign languages (2), automatic language choice based on IP address (2);
- **customer service:** requesting information via e-mail and/or telephone (4), feedback opportunities (2), opinion about travels or services (2), forum (2);
- **information of equal opportunity:** routes for disabled passengers (4), information about vehicles (2), webpage for visually impaired people (4).

2.2 Definition of the user groups

Since the appreciation of the information service depends significantly on the personal features of the passengers, different user groups were created from the passengers by their age, mobility features and their motion abilities (Fig 4). The user groups were formed using the combination of these three points of view [46].

- **Student – younger, work-motivated and leisure time based, without any problem:** Members of this group are open to new solutions, are interested in the comfort services and use dynamic data, but supplementary information is not so important for them.
- **Worker – older, work-motivated and leisure time based, without any problem:** Daily travel in the city is their usual characteristics, that is, why they prefer the route planning options using actual data. Comfort services have the lowest priority for them.
- **Tourist – younger and older, leisure time based, handicapped:** During leisure activities the non-well-known routes are mostly used, therefore tourists (e.g. with big luggage) belong here. For them route-planning services and payment are the most important, while dynamic data and supplementary information the less.
- **Businessman – older, work-motivated and leisure time based, without any problem + handicapped:** Mostly those passengers belong here, who want to travel in the most convenient way and therefore are more interested in comfort services and supplementary information.
- **Pensioner – older, work-motivated and leisure time based, handicapped:** Members of the elderly generation belong here, who can orientate and move mostly with difficulties. For them the emphasis is on supplementary information.
2.3 Elaboration of the evaluation method

The evaluation was performed in two main steps (Fig 5). First the journey planners were compared (scoring) to each other based to the elaborated framework of aspects, which resulted in the general evaluation number. Then user groups, their characteristics and their transportation share were defined. Taking preferences of the user groups into account (weighting) the average evaluation number for each journey planners was calculated. Finally based on this new measure the multimodal journey planners can be compared to each other.

For the scoring step the compensational Multi-Criteria Analysis (MCA) was adapted and implemented, because it produces clear and well-comparable results. MCA methods are easy to compute and widely used [115], [167], [39], [156]. An interval based scoring method was used, where the aim was to provide comparison among the journey planners. The weighting was introduced in order to take the different preferences of the user groups according to the aspects into account. The method has the following advantages:

- The methodology is easy and understandable, the process is simple.
- Simplification of complex situations and provision of a solution transparently.
- MCA is rational, which means that it provides a homogeneous result for a range of aspects, therefore the results can be easily compared.
Fig 5 Process of the evaluation of multimodal journey planners

2.3.1 Scoring

The multimodal journey planners \( (j) \) were evaluated on a 0-10 valued scale according to the single aspects \((i)\). To each journey planner belong together \(I\) pieces of evaluation numbers and \(J\) pieces of single aspects. From the evaluation numbers an \(I \times J\) sized evaluation matrix is defined, in which the elements are signed by \(p_{ij}\). Summing up the evaluation numbers for each single aspect, a general evaluation number can be provided for the multimodal journey planners \((u)\).

\[
u_j = \sum_{i=1}^{I} p_{ij}
\]  

- \(i\) – single aspects, \(i = 1, \ldots, I\),
- \(j\) – multimodal journey planners, \(j = 1, \ldots, J\),
- \(p_{ij}\) – elements of the evaluation matrix,
- \(u_j\) – general evaluation number for the \(j\) multimodal journey planner.

2.3.2 Weighting

The general evaluation number is already useful by itself, but it does not take into account the different preferences of certain user groups, which are defined later. The solution is presented by using normalization and weighting.

To all user groups \((k)\) and aspects \((i)\) belong weighting numbers, so called preference values \((s_{ki})\), which form a \(K \times I\) sized weight matrix. The values of the elements in this matrix can be determined by exploring the preferences of the user groups.
From the evaluation matrix and the weight matrix a $K \times J$ sized qualifier matrix can be generated, that takes into account the different preferences of the user groups. Its elements, which are the qualifier values ($u_{kj}$) for a certain multimodal journey planner and a certain user group, are to be calculated in the following way. The summed product is generated from the elements of the $j.$ column of the evaluation matrix ($p_{ij}$) and from the elements of $k.$ row of the weight matrix ($s_{ki}$), which value then is divided by the summed product of the maximal given evaluation numbers ($p_{i \text{max}}$) and the corresponding weights. This step implements the compensation in the method.

\[
u_{kj} = \frac{\sum_{l=1}^{l=1} s_{kl} * p_{ij}}{\sum_{l=1}^{l=1} s_{kl} * p_{l \text{max}}}\]

- $k$ – user groups, \( k=1,..,K, \sum_{l=1}^{l=1} s_{kl} = 1, \)
- $s_{kl}$ – elements of the weight matrix,
- $p_{ij}$ – the maximal given evaluation number according to the $i.$ aspect,
- $u_{kj}$ – elements of the qualifier matrix.

Knowing the qualifier values ($u_{kj}$) for the multimodal journey planners and the transportation share ($r_k$) of the user groups the average evaluation number ($u_j^*$) can be determined, which is referred to all passengers and user groups at the same time.

\[
u_j^* = \sum_{k=1}^{K} r_k * u_{kj}\]

- $r_k$ – transportation share of the $k.$ user group,
- $u_j^*$ – average qualifier number for the $j.$ multimodal journey planner.

### 2.4 Review of the journey planners

Such journey planners were selected, which are widespread, popular, important in a region, have leading ideas or extraordinary solutions. Their strengths and weaknesses were also analyzed.

#### 2.4.1 Classification of the journey planners

14 journey planners have been analyzed (Table 2). The selected systems were grouped by spatial coverage (urban, regional, international) and handled transport modes (public transport, car, rail, air, multimodal).

Considering urban and regional journey planners the BKK Futár from Budapest, the TfL (Transport for London), the Austrian AnachB, the 9292 from the Netherlands, the German Bayerninfo and the Hungarian Útvonalterv were selected, which provide basically journey planning for public transportation, but the service is usually extended to other transportation modes (e.g. bike, car, rail).
Some journey planners with international coverage were also investigated, as Google maps, Rome2Rio, TripGo, Eu-Spirit and RouteRank. These services usually cover specific areas (e.g. Europe, Australia) in details, however, they can plan between main destinations, too.

The basically railway or airplane based journey planners, which are mainly operator dependent, are the following: German railways (DB), Austrian railways (SCOTTY), and Lufthansa.

A few journey planners (BKK Futár, AnachB, Bayerninfo, Rome2Rio, TripGo, Eu-Spirit, RouteRank) provide multimodal services (combining different transportation modes within a trip), and in most cases they only combine a few modes (e.g. public transport with bike sharing).

<table>
<thead>
<tr>
<th></th>
<th>BKK Futár</th>
<th>TFL</th>
<th>AnachB</th>
<th>9292</th>
<th>Bayerninfo</th>
<th>UrbanRail</th>
<th>Google</th>
<th>maps</th>
<th>Rome</th>
<th>2Rio</th>
<th>TripGo</th>
<th>Eu-Spirit</th>
<th>RouteRank</th>
<th>DB</th>
<th>Scotty</th>
<th>Lufthansa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spatial coverage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>urban</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>international</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Handled transport modes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bike</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>car</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rail</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>air</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>multimodal</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Classification of journey planners

BKK Futár was developed by the Transport Centre of Budapest in 2014 and covers the area of Budapest in Hungary. The data source is GTFS (General Transit Feed Specification) data, which is extended with current traffic information both for private and public transport. The timetables are monthly updated, in case of special events even more frequently.

TFL was developed by Transport for London, which is responsible for transport system in Greater London in England. It covers the underground lines, buses, light railways, rails, trams, taxis, cycling, and river services. An API is available for checking timetables, current status, disruptions and fares.

AnachB works in the Vienna Region relying on Traffic Information Austria (VAO) since 2009. It provides traffic information service across the entire country. Several data sources are combined, as the bike-sharing and the car-sharing system, the Park and Ride facilities and transport authorities.

9292 has provided travel information on public transport in the Netherlands for over 20 years. It covers all transport operators in the country considering train, bus, tram and metro.
Bayerninfo went online already in 1995 and collects traffic information from the following sources: traffic sensors and cameras on the motorways, real-time FCD data, incident information from traffic authorities, weather conditions, public transport data from the operators.

Útvonalterv was developed by the company Topolisz in 2003 and it provides besides public transport information actual traffic information based on camera pictures and FCD data.

Google maps is possibly the most well known journey planner, which was launched in 2005. It offers an API for route planning and covers most countries in the world. Google maps offers traffic data in real-time using crowd sourcing data obtaining GPS locations of mobile phones. It also provides public transport data using GTFS (General Transit Feed Specification) to exchange transit information.

Rome2rio was launched in 2011 and it uses data from several types of sources including API data feeds, GTFS data and OSM maps. The service is based on a unique repository of airplane, train, bus and ferry routes with thousands of transport operators represented in the database.

TripGo covers ca. 200 cities with up-to-date public transport time tables and also real-time data for selected public transport providers.

Eu-Spirit was developed in the framework of a European Union project in 1998 and is based on existing local, regional, and national travel information systems, which data are linked through interfaces.

RouteRank was developed in 2006 and it integrates road, rail and air options for journeys in Europe and the USA. RouteRank searches the websites of ca. 700 airlines and 200 online travel agencies, and provides routing data through ViaMichelin.

DB relies not only on its own data sources regarding train schedules, but covers whole Europe. It also provides private transport route information provided by ADAC (General German Automobile Club).

Scotty was launched in 2007 and is based on the route planning information system of Hafas. Its main routing information source originates from its owner, ÖBB (Austrian Railways).

Lufthansa provides route planning for its 200 destination worldwide and other services through external service providers (e.g. hotel booking through booking.com).

2.4.2 Outstanding functions and top solutions

The innovative features of the journey planners were highlighted (Table 3). Progressive development has been realized by several journey planners, but the most outstanding ones are: BKK Futár, DB, Rome2Rio and Scotty. Concerning route-planning services a wide range of features was already implemented by many journey planners. The developments mainly focused on handled data and supplementary information, while concerning booking and payment and comfort service information no real novel solutions were presented.
In Bayerninfo not only bike routes are recommended, but also their height profile and supplementary services (e.g. tourist attractions, B+R, GPS track) are considered. Scotty has a unique function with its motorailtrains planning module (traveling by train with own car on board), and it also draws attention on crowded vehicles on specific days.

Rome2Rio and DB can calculate prices for different transportation modes.

BKK Futár provides a very comprehensive dynamic information service covering both private and public transport, which information is both listed and visualized on the map. Furthermore it presents a map of actual traffic information based on crowd sourcing (data from Waze application). Google maps provides actual traffic situation based on crowd sourcing from mobile phones.

The journey planner of TfL displays planned restrictions during the route planning; furthermore there is the possibility to check the traffic situation of all transport means in the city on a map with text information and camera pictures.

The most innovative function of Scotty is that it can show real-time position of trains on the map and also their delays.

AnachB can continuously update searches with current traffic information, furthermore it can show the P+R parking lots, the citybike and car-sharing stations with their actual capacity. It provides a highly personalized bike route planning service. AnachB gives information about the traffic situation, also by live images of the road network and estimates it up to 1 hour in advance.

DB can calculate new, alternative routes according to current traffic situation. It completes the car journey plans with traffic jam forecast based on historical data and with estimated arrival time based on real-time data.

AnachB and TfL systems are outstanding in data handling, namely they provide detailed personalized settings. In TripGo a slider for trip planning aspects (cost, time, environment and convenience) is built in, which represents personal preferences in details.

### Table 3 Top solutions of journey planners

<table>
<thead>
<tr>
<th></th>
<th>BKK Futár</th>
<th>TfL</th>
<th>AnachB</th>
<th>9392</th>
<th>Eu-Spirit</th>
<th>Bayerninfo</th>
<th>Lehonalterv</th>
<th>Google maps</th>
<th>Rome2Rio</th>
<th>TripGo</th>
<th>RouteRank</th>
<th>DB</th>
<th>Scotty</th>
<th>Lufthansa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Route-planning services</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Booking and payment</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Handled data, operational features</strong></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Comfort service information</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Supplementary information</strong></td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>-</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

- 25 -
Rome2Rio, DB, Scotty and Lufthansa provide information about several additional services (e.g. hotel, car rental). Furthermore Rome2Rio provides sightseeing information about the cities, as well as DB and Scotty inform about WiFi access on the trains and in the stations.

DB, RouteRank, 9292 and TripGo attach the estimated CO₂-emission to their journey plans. Furthermore in some cases calculation methods are available and travel costs are also presented, which are compared to the individual transport.

In BKK Futár a direct feedback opportunity about failures of the system is realized. Rome2Rio and 9292 provide a special option regarding customer contact on Twitter.

The Pedroute application of the Útvonalterv offers detailed setting opportunities for disabled (e.g. degree of slope, pavement type, edges, crossing points) and visually impaired (e.g. font size, colour scheme) persons. Eu-Spirit also provides barrier free routes with options of vehicle accessibility and walkways (eg. ramp, escalator, stairs). A unique help for blind people is the reading out and dictation function of 9292.

2.5 Evaluation of the journey planners

Using the multi-criteria analysis the evaluation numbers and the weights were determined. Finally journey planners were ranked by the average qualifier numbers.

2.5.1 Scoring

The evaluation numbers are presented and summarized in Table 4. The rows represent the aspects \((i)\) and the columns represent the journey planners \((j)\). The general evaluation numbers of the journey planners \((u)\) are calculated by the summation of the evaluation matrix’s elements \((p_{ij})\) by columns. Not only the evaluation numbers, but the values in percentage are also shown. The maximal obtainable value was 180, where 10 points for each single aspect was the maximum. Values were assigned to the journey planners based on how much the journey planner met the requirements of the aspects. The detailed evaluation of the journey planners regarding features of the single aspects can be found in Annex 1.

The urban and regional journey planners perform better regarding route-planning services and dynamic data. The international journey planners tend to apply real multimodal solutions and provide balanced results in case of all main aspects except from booking and payment and customer service information. Journey planners of rail and airway service providers attained better results for booking and payment and customer service information.

It has been found, that even the best journey planner did receive only 64% of all possible points and the journey planners received 41% on average, which confirms that there is still much development opportunity for the operators of journey planners to fulfill user expectations.
According to the evaluation the journey planner of DB provides the best services and widest range of information for passengers. This system is especially outstanding in the main aspects of booking and payment and comfort service information, but its dynamic and supplementary information is also very detailed (e.g. environmental impacts and services at the station). However the route planning module of DB is not one of the bests.

BKK Futár attained the second place because of its comprehensive features for route-planning services and handled data. All possible data input modes (address, name of stop, POI, GPS coordinates and map) and many planning aspects and data display opportunities with detailed information and visualization have been implemented in the system. As BKK Futár is an urban journey planner, booking and payment and comfort service information are not comprehensively developed.
AnachB and Scotty were very close to each other running for the third place. Both journey planners performed well regarding handled data and supplementary information, however AnachB has better results for route-planning services and dynamic data, while Scotty is better in booking and payment and customer service information.

In the middle section Google maps, TfL, Útvonalterv and 9292 are to be found. All these journey planners have outstanding results for some main aspects (e.g. Google maps for route-planning services or 9292 for supplementary information), but their overall performance is not very spectacular.

Eu-Spirit, Lufthansa, Bayerninfo, Rome2Rio, TripGo and RouteRank received points below average, however they possess some outstanding functions, which are not implemented in other journey planners (e.g. Rome2Rio and Lufthansa in customer service information or Bayerninfo in handled data).

2.5.2 Weighting

As a first approach estimated weights were assigned to the main aspects based on typical user group preferences and engineering considerations. In Table 5 the weight matrix is illustrated, in the rows the user groups \((k)\) and in the columns the main aspects \((i)\). The weights of the single aspects \((s_{ki})\) are calculated by the equal distribution of the main aspect values.

<table>
<thead>
<tr>
<th></th>
<th>Route-planning services</th>
<th>Booking and payment</th>
<th>Handled data, operational features</th>
<th>Comfort service information</th>
<th>Supplementary information</th>
<th>Transportation share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>0,2</td>
<td>0,15</td>
<td>0,3</td>
<td>0,25</td>
<td>0,1</td>
<td>0,3</td>
</tr>
<tr>
<td>Worker</td>
<td>0,3</td>
<td>0,2</td>
<td>0,25</td>
<td>0,1</td>
<td>0,15</td>
<td>0,3</td>
</tr>
<tr>
<td>Tourist</td>
<td>0,25</td>
<td>0,3</td>
<td>0,15</td>
<td>0,2</td>
<td>0,1</td>
<td>0,15</td>
</tr>
<tr>
<td>Businessman</td>
<td>0,25</td>
<td>0,1</td>
<td>0,15</td>
<td>0,3</td>
<td>0,2</td>
<td>0,1</td>
</tr>
<tr>
<td>Pensioner</td>
<td>0,3</td>
<td>0,1</td>
<td>0,1</td>
<td>0,2</td>
<td>0,3</td>
<td>0,15</td>
</tr>
<tr>
<td>Average</td>
<td>0,26</td>
<td>0,17</td>
<td>0,19</td>
<td>0,21</td>
<td>0,17</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5 Weights of main aspects according to user groups and transportation shares

The values of transportation share \((r_k)\), which are shown in the last column of the table, are based on the results of the National Traffic Data Survey [126]. This is a general value that may differ among countries, however it still provides some guidelines.

The qualifier matrix (Table 6) can be calculated by weighting the journey planners according to the user group expectations and normalizing these values \((q_{kj})\). The values of the qualifier matrix were calculated considering how close they are to the maximal given value of single aspects. From these values the average evaluation number \((n_j^*)\) can be obtained by multiplying them with the transportation share \((r)\). On the basis of the formulas the evaluation of multimodal journey planners was updated. In the rows of the table the evaluation results considering user groups are present (in per cent), while in the last row the average evaluation number \((n_j^*)\) is shown referred to all passengers.

- 28 -
Table 6 Evaluation of journey planners with consideration to the user groups (%)

<table>
<thead>
<tr>
<th></th>
<th>BKrk Futur</th>
<th>TIL</th>
<th>AnachB</th>
<th>9292</th>
<th>Eu-Spirit</th>
<th>Bayerninfo</th>
<th>Utvornaltiev</th>
<th>Google maps</th>
<th>Rome2Rio</th>
<th>TripGo</th>
<th>RouteRank</th>
<th>DB</th>
<th>Scotty</th>
<th>Lufthansa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>59</td>
<td>49</td>
<td>53</td>
<td>42</td>
<td>36</td>
<td>43</td>
<td>51</td>
<td>38</td>
<td>40</td>
<td>19</td>
<td>75</td>
<td>59</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Worker</td>
<td>68</td>
<td>56</td>
<td>62</td>
<td>48</td>
<td>41</td>
<td>51</td>
<td>59</td>
<td>39</td>
<td>45</td>
<td>24</td>
<td>75</td>
<td>58</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Tourist</td>
<td>56</td>
<td>49</td>
<td>50</td>
<td>41</td>
<td>31</td>
<td>39</td>
<td>47</td>
<td>39</td>
<td>41</td>
<td>24</td>
<td>78</td>
<td>59</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Businessman</td>
<td>55</td>
<td>44</td>
<td>50</td>
<td>40</td>
<td>34</td>
<td>45</td>
<td>49</td>
<td>41</td>
<td>41</td>
<td>25</td>
<td>73</td>
<td>58</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Pensioner</td>
<td>61</td>
<td>48</td>
<td>55</td>
<td>46</td>
<td>38</td>
<td>53</td>
<td>54</td>
<td>41</td>
<td>45</td>
<td>30</td>
<td>73</td>
<td>56</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Average qualifier number</td>
<td>61</td>
<td>50</td>
<td>55</td>
<td>44</td>
<td>37</td>
<td>46</td>
<td>53</td>
<td>39</td>
<td>43</td>
<td>24</td>
<td>75</td>
<td>58</td>
<td>42</td>
<td></td>
</tr>
</tbody>
</table>

2.5.3 Comparison of the general and weighted evaluation

In order to obtain the differences between the general evaluation and weighted evaluation with user groups, the two evaluation numbers were compared to each other and the differences were highlighted.

Fig 6 Comparison of general evaluation numbers and average evaluation numbers

The average evaluation numbers were by 6.7% higher compared to the general evaluation numbers (Fig 6). This relative significant modification is caused by the compensational feature of the applied MCA method. Because in case of 12 out of 18 single aspects the journey planners did not meet completely the features of the certain single aspect (namely the values were lower than the maximal 10 points). Consequently the average evaluation numbers are higher.

The other cause for modification is the distribution of values given for the main aspects and requirements of the user groups. Namely, if all user groups prefer route-
planning services and in general features of route-planning services are implemented, the average evaluation number is higher. But in case most user groups prefer comfort information services and comfort information services are not realized by the journey planners, the average evaluation number is lower (even less than the general evaluation number).

In our case the distribution of user group preferences for main aspects were quite balanced and similar to the features implemented in the journey planners, this factor did not modify the final result of the average evaluation number. The biggest change was obtained in case of DB and Scotty (10.6% and 9.9%), as they performed well regarding route-planning services and comfort service information, which were the most preferred main aspects (with 0.26 and 0.21) in average (see Table 5).

The elaborated method presented a quantified evaluation and ranking of multimodal journey planners, which helps the operators to rate and compare their provided services from the viewpoint of the passengers. The ranking provides information for the service operators and decision makers about possible development directions, so that they can deliver more advanced solutions for the passengers.

2.5.4 Sensitivity analysis

The sensitivity analysis was performed to present the sensitivity of single aspects regarding journey planners. Sensitivity is a measure, if any value of a single aspect is modified, how much this affects the final result of the evaluation of the journey planners.

It was preconditioned that the single aspects are independent from each other, therefore the sensitivity analysis needs to be performed only for an arbitrarily chosen single aspect per journey planner. That is why the results can be obtained in Table 7 in vector form, not matrix.

The average sensitivity was 1.45%, which means that in general the journey planners depend much on the values given for the single aspects. The highest sensitivity was observed in case of RouteRank with 2.44%, while in case of DB with 0.85% the change of a value had less effect. This phenomenon can be directly derived from the overall evaluation. The more points were given for a certain journey planner, the sensitive for modifications it is.

<table>
<thead>
<tr>
<th>Sensitivity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTR</td>
</tr>
<tr>
<td>1.11</td>
</tr>
</tbody>
</table>

Table 7 Result of the sensitivity analysis
2.6 Evaluation of Hungarian online travel information services

After the survey of international journey planners, I put emphasis on national online travel information services, primarily on the systems of bus transport operators, as the others had been already included in the international evaluation. Prior to this study no such a quantitative comparison was available in Hungary, which deals with functionalities and other features of online travel information services. The results of the analysis can be used by the operators to improve their service according to passengers’ requirements.

The comparison was realized for the old organizational structure of Volán operators in order to reveal differences among operators in a more detailed way. Meanwhile the organizational structure of the bus companies was reorganized (creating larger companies with their integration), however the information handling and journey planning functionalities did not change significantly. The integration did not cover online travel information services yet. In order to provide comparability to the new structure, the new organizations are also represented in the evaluation.

The description and features of Hungarian online travel information services, and the details of the evaluation can be found in Appendix 1.

According to the evaluation it can be stated that Kisalföld Volán offers the highest level of services with its outstanding journey planning function and handling of dynamic data. Not far behind are the Vértess Volán and the Kunság Volán, which use the same journey planner. The latter journey planner obtained a higher rank because of its chip card pass system and foreign language opportunity.

The Borsod Volán, the Bakony Volán, the Somló Volán, the Vasi Volán and the Zala Volán reached middle level points. These operators enable the online ticket purchase through a link to menetrendek.hu, but do not have journey planning service, foreign language information or equal opportunity information (except for Borsod Volán).

Concerning the further operators in many cases even static maps are not available, or there is no information about the planned restrictions, thus the results of these operators are obviously weak. Only the Pannon Volán reaches up to the middle level because of the foreign language information.

In case of route-planning services, booking and payment some online travel information services performed well, but concerning the maximal obtainable values all online travel information services were far behind. Therefore the weighting of the results caused big differences, because of the many single aspects, where none of the online travel information services were given the maximal available points. It can be observed that best operator (Kisalföld Volán) attained 66 per cent, which means that for some single aspects there was another operator, which preformed better. Concerning the average evaluation number (Fig 7) the ranking did not change.

Based on the evaluation the weaknesses and development directions of online travel information services were revealed, which is important to provide standardized and integrated services in the new operational structure.
The spatial representation of the results was also presented. The information service levels of the online travel information services are very different considering spatial distribution (Fig 8). The darker the color is the better the result. The operators located close to the capital, in the western and surprisingly in the southern area of Hungary provide much higher service level than operators from the north-eastern part of Hungary. The results do not correlate with the average GDP or number of inhabitants of the area, because for example in the southern counties more people live, than in the service area of Kisalföld Volán. Concerning the disparities within the new operational structure the biggest differences are to find in case of ÉNYKK, KNYKK and DAKK.
3 Analysis and creation of user groups based on preferences and evaluation aspects

Preferences regarding the evaluation aspects of multimodal journey planners were analyzed in order to provide more precise information about the user groups. The aim was to define and create realistic weights of the user groups. Considering user group preferences big differences were expected among them. In order to investigate this hypothesis, a survey was created and statistical analysis methods were applied. The methods were customized for the specific task to provide well comparable results.

In the field of passenger preference analysis many papers presented valuable results and contribution. In the paper of Wang et. al. [171] different weighting approaches are described for MCA method to obtain rational results. However, Garmendia and Gamboa [76] criticized weighting processes for aggregating various decision maker priorities into an average weight. Moreover, based on the research of Rogers and Seager [140] participants might be reluctant to reveal their preferences.

Longo et. al. [113] tried to understand the preference structure of passengers. They created a survey, which pointed out information handling features and preferences of different user groups. The work aimed to define the most suitable transportation mode for the users based on user preferences. Some general evaluation aspects are present in the paper of Campos et. al. [25]. They proposed a procedure to evaluate sustainable mobility in urban areas. A set of indicators based on user needs were defined, as environmental, economical, and social aspects. The evaluation is based on an index calculated through a weighted multi-criteria combination procedure. Winkler [78] modeled the decisions of passengers in urban environment. The aim was to go beyond the reliability of stated preference surveys. A new survey method was build, which represents the real decisions of passengers based on their personal preferences.

Identifying patterns in answers of users can be performed by different statistical and clustering methods. Hierarchical clustering is one of the most used method for that, where a set of answers is present without any pre-classified grouping [155]. Therefore it is also called unsupervised classification. Clustering algorithms are used in many different fields, also in transportation. In the paper of Chiang et. al. [31] airline passenger behavior was investigated from the economical viewpoint. Market clusters were established in order to enable companies to focus their marketing strategy on similar user groups.

An analysis of service quality in transportation is conducted by de Ona et. al. [37], who defined qualitative aspects of public transportation services and created user groups based on age and usage habits. A CART (Classification And Regression Trees) method was used, which includes classification and regression trees. The analysis was focused on revealing the suitable public transportation modes for the established user groups. Becker et. al. [12] considered indicators for public transport service quality, as convenience and cleanliness of stations, safety of vehicles, ticket sale opportunities. Travel information and conditions (e.g. schedule, punctuality) are also taken into account. The users are grouped based on their age and trip purpose (work, leisure, shopping, services). The aim of their study was to define the reliability of the given answers by the users.
Although, several papers dealt with the topic of service quality and passenger information, it was rarely used directly for user groups. Therefore current research aimed to provide valuable statements about user group preferences (Fig 9). First a survey was created to analyze preferences of the original user groups. With the preferences regarding the main aspects, the evaluation of multimodal journey planners was performed. In order to find connection between single aspects, a correlation analysis was assessed. Finally to create user groups, which answers are similar to each other within the groups, but different among the groups, Ward method was applied to create new user groups.

Fig 9 Analysis and definition of user groups

An expanded figure describes the process of the evaluation using survey results. The first step is the introduction of survey content and the method of the statistical analysis. Then the survey results are analyzed, where values of original and ranked aspects are calculated. Based on these values the weighting and evaluation of multimodal journey planners was performed.

Fig 10 Process of evaluation using survey results
3.1 Definition of the survey method

The key of collecting reliable data was a survey based on the aspects, which is necessary to get realistic weights of the user groups. Therefore a survey of 12 question groups was created, which is available in Annex 2.

3.1.1 Survey content

The first part of the survey contained questions about the users’ age, their occupation (student, worker, pensioner), their health situation (whether they are handicapped or not), and the reason why they use journey planners (work, leisure, tourism). According to these data I have identified to what user group the participants belonged. The users were asked to rank the main aspects according to how important are these services in general for them. The rest of the survey was divided into 7 question groups, each part focusing on a category of journey planning features.

All together 57 thematic questions had to be answered, which were connected to the single aspects, in some cases more questions belonged to a single aspect. A 10-grade scale was defined for each answer, where the possible value set consisted of integers. 1 stand for the least importance, 10 stand for the most importance. This enabled to get a wide and accurate range of answers, which represents the variety of users’ requirements.

3.1.2 Method of statistical analysis

Having the results of the survey a statistical analysis was performed [70], [152] in order to reveal differences between the user groups. The discussed methods are standard and most widely used statistical methods, which were suitable to define differences among user groups. In the first step the mean values and variances were calculated concerning each single aspect and user group.

In the next step the Bartlett test was performed, which examines the variances of the user groups (i.e. whether the answers of the user groups are similar to each other or not). In the Bartlett test the number of the samples (user groups) is \( r \), which contains \( n_1, n_2, \ldots n_r \) users. The mean of the samples (\( \bar{X}_i \)), and the standard deviation (\( s_i \)) was defined, where \( i \) is a certain user group. According to the null hypothesis (\( H_0 \)) the Bartlett test assumes that variances are equal among the user groups. If the result of the Bartlett test is lower, than the critical value, the test is accepted. If higher, then the test is rejected, which means that the variances are different at least for two groups (i.e. there is a difference among the answers of the user groups). Another parameter of the Bartlett test is the critical value (based on \( \chi^2 \) distribution), which depends on the degree of freedom \( (n_i - 1) \), which is the member’s number of the user groups, and the significance level \( (\alpha) \). In our case \( \alpha \) was 0.05. The deviances between variances of selected user groups (\( s_1, s_2 \)) were pairwise tested using the F-test. If the F-test is lower than the critical F value, the variances are similar. Therefore no special requirements of the user groups are present. Because of the high number of observations the critical value for F was chosen 2.2. [152]

\[
F = \frac{s_1^2}{s_2^2}
\] (4)
If the Bartlett test was accepted, the ANOVA (ANalysis Of VAriance) method was used to determine whether there are any significant differences between the user groups. If a difference is revealed, it can provide information about development directions. If there is no difference, then all user groups have the same needs. The null hypothesis ($H_0$) of the ANOVA assumes that mean values are equal among the user groups. The ANOVA method calculates the sum of squares, where the square contains values between and within groups. In cases the ANOVA test was refused, a two-sample pairwise t-test was performed in order to reveal, which groups differ from each other concerning mean values and variances. The critical value ($t_c$) for the t-test depends also on the significance level ($\alpha$) and the degree of freedom. If the t-test is higher than the critical value, the $H_0$ hypothesis is rejected, which means that opinions of the two user groups are different [152].

The results of t-test can be calculated as following:

$$t = \frac{x_1 - x_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

(5)

$x_1, x_2$ – mean values of the populations, where $x_1 - x_2 > 0$,

$n_1, n_2$ – numbers of the populations.

The $s_p$ formula is the following:

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

(6)

$s_1, s_2$ – standard deviations,

$n_1, n_2$ – numbers of the populations.

3.2 Analysis of survey results

The research was conducted online, where 133 participants filled in the survey in a time period of 8 weeks. The significant part of the participants came from Hungary, but some answers arrived from abroad. Concerning the distribution of the user groups most of them were students and tourists, however the number of pensioners was not very high (only 9%), which is due to the online type of the survey (Fig 11).
The prerequisite of the application of presented statistical analysis methods is the examination of the normal distribution of the results, which is called normality analysis [23]. The normality of a distribution can be verified by the analysis of two typical parameters, the kurtosis and the skewness.

The kurtosis is the measure, how much the distribution of the answer values of the participants are peaked or flat regarding an ideal normal distribution. If the value of the kurtosis is high, then distribution of the answer values has a peak near the mean value and declines rapidly. If the value of the kurtosis is low, then distribution of the answer values has a rather flat top.

The skewness is a measure of the lack of symmetry. If the distribution of the answer values has a similar shape in the left and in the right of the mean value, then the distribution is symmetric.

In case of a perfectly normal distribution the value of both parameters is 0. Small deviations from this ideal value do not cause any problem in the statistical calculations, therefore as a general rule it is enough to be checked, whether the absolute value of the parameters is less than 2 [35]. In case it is, then the statistical analysis can be seamlessly performed. In case of higher values a problem only occurs, if the number of participants is small or the number of participants in user groups is highly inhomogeneous.

Performing the normality analysis for each single aspect it can be obtained that most of the single aspects correspond to a normal distribution (Table 8). Only in the case of 6 single aspects is the value higher than expected. Considering address, trip duration, visualization and timetable, all these single aspects were valued extremely high among the users with less exception, and this caused a shift in the normal distribution (high kurtosis and negative skewness). In the case of mobile phone and other payment the shift was caused by the low evaluations by the users (high kurtosis and positive skewness).

However the average of the single aspects is between -2 and 2, therefore it can be claimed that the survey results represent a normal distribution, therefore further investigations regarding the analysis of characteristics of main and single aspects can be executed.

<table>
<thead>
<tr>
<th>Normality analysis of the single aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td>route-planning services</td>
</tr>
<tr>
<td>address</td>
</tr>
<tr>
<td>skewness</td>
</tr>
<tr>
<td>kurtosis</td>
</tr>
</tbody>
</table>

Table 8 Normality analysis of the single aspects
3.2.2 Analysis of ranking values

The survey results were used to analyze the preferences of the user groups. The users were asked to rank the main aspects (1 is the least important, 5 is the most important main aspect). Their answers were normalized (where the value 1 means that the certain aspect is of average importance). Therefore if a main aspect gains higher values, it has to be considered with a higher weight in the evaluation (Fig 12).

Route planning is definitely the most important aspect for each user group, while handled data reached rank 2. Booking and payment is of average importance, while the comfort service and supplementary information seem to be the less important to all passengers.

Considering differences between the user groups the biggest standard deviation is present for comfort service information and the most agreement is shown for route planning. Businessman and students gave the most stress on route planning, while students and pensioners obtained higher scores for booking and payment, and surprisingly concerning handled data pensioners are mostly interested.

![Fig 12 Normalized ranking values of the main aspects for each user group](image)

3.2.3 Weighting the main aspects with ranking values

The answer values (given for the main aspects by the users) were weighted with the ranking values of main aspects, so that the differences among user group preferences could be shown. Some aspects will be even more important, while other less. As a result of the ranking the route planning and handled data strengthened their positions, and increased from 23% to 33% (from original to ranked values) and from 27% to 31%, while the role of the other aspects respectively decreased, especially the supplementary information from 17% to 10%. Surprisingly handled data were the most important for the users, which is in contrary to the answers to the ranking question.

The averages of the answers were calculated for the main aspects for each user group. In the original case the average values of the main aspects resulted in a ranking of the following order (Fig 13): handled data (7,69), route planning services (6,57), booking and payment (5,05), supplementary information (4,84) and comfort service information...
Students valued most handled data and route planning, while tourists gave lower points for all aspects in general. The biggest differences among user groups were present for supplementary information. However it can be obtained that the averages are quite similar to each other, all user groups behaved similar and prefer generally the same features.

Weighting the main aspects with the ranking values highlights the differences among the preferences of the user groups. The average values after ranking are the following: route planning services (9,79), handled data (9,20), booking and payment (4,73), comfort service information (2,99), supplementary information (2,97). Some aspects are close to the value 10, which is the consequence of the ranking, as it considers some aspects with higher values and others with lower than the original. As expected the values for route planning service were mostly increased due to its rank 1 in the ranking question. Also the values for handled data are significantly higher due to rank 2. Booking and payment did not change that much due to its rank 3. Comfort service and supplementary information values decreased as they reached rank 4 and 5. The biggest differences among user groups were present for handled data. The interest of workers for supplementary information is stronger represented after the ranking.

Fig 13 Comparison of original and ranked answer values for the user groups
3.2.4 Weighting the single aspects with ranking values

In the last and longest part of the survey answers of the single aspects were evaluated. A detailed analysis of each single aspect for all user groups averaged is shown in Fig 14, where the original and the weighted values are presented for all user groups. The weighting with ranking values highlights differences in preferences, and can result in ranked values above 10, which then represent more important single aspects. Respectively ranked values lower than the original values mean less importance. A detailed statistical analysis of single aspects in presented in Annex 3, where to all single aspect the means and variances regarding the user groups are shown. Furthermore the results of pairwise F-tests (to compare variances) and t-tests (to compare means) are represented, where extreme values are denoted by red. Original and ranked values were calculated in the last part of the Annex, where only extreme values are shown.

The cause for rank 2 of route-planning services in the original case is revealed. Because some single aspects got high values (e.g. address based, travel time and distance, visualization), but many aspects were valued lower (e.g. POI, GPS based), and this resulted in a lower average for route-planning services. Possibly having fewer questions for this main aspect could have resulted in higher values.

Fig 14 Answer values of single aspects in the original and in the ranked case

Only small values were given by the passengers considering GPS-based input, however most of the users prefer address-based data input. Name of the stop was equally important among almost all user groups except for businessman. The travel time and duration was more important for the users than costs. The visualization on the map seems to be highly interesting for all user groups. The fares were equally relevant for all user groups. The most appreciated payment option was the credit card, then cash. Information about timetable is far the most needed feature, but concerning the deviation there are differences. The travel conditions have less importance for all user groups with a high variance. Information about weather conditions was higher valued, than other single aspects. In the last main aspect group it was revealed that users are less concerned about environmental questions. The detailed calculations of the single aspects regarding the user groups can be obtained in Appendix 2.
3.3 Creation of Ward user groups

The results of the user group analysis showed that all original user groups (student, worker, tourist, businessman, pensioner) behaved similarly and preferred generally the same features. Therefore Ward method was used to create with clustering such user groups that have similar requirements regarding single aspects and more divergent requirements regarding main aspects.

3.3.1 Correlation analysis

Correlation analysis is a statistical method to define connections between single aspects [32]. In my research it is applied to determine the consistency and similarity of answers of user groups regarding single aspect pairs. With the revealed connections latent requirements are identified (i.e. if a single aspect is highly correlated to another, then knowing the importance of one predicts the importance of the other). This helps developers to find out needs of the user groups regarding single aspect pairs.

The mean values of the answer for single aspects were compared. The correlation coefficient is the proportion of the covariance and product of standard deviations of answer values of single aspects (s_x, s_y), where x and y are single aspects. Correlation can have values from -1 to +1, where -1 means negative (inverse) correlation and +1 means positive correlation. Values close to 0 represent independency between the single aspects.

$$ corr(x, y) = \frac{cov(x,y)}{s_x \times s_y} $$ (7)

In order to reveal connections among single aspects, the survey results were investigated with correlation analysis. When comparing the single aspects no negative correlations (x < -0.3) were found, which means that no clearly opposite opinions are present. Positive correlation (x > 0.5) could be obtained for a few questions.

Generally high correlations were observed for comfort service information and handled data. This means that a connection is present between requirements of the user groups regarding these two main aspects. The highest correlation was found in case of questions about Wi-Fi at the stations and Wi-Fi in the vehicles (correlation: 0.87; average values of the single aspects: 4.00/3.44), which is quite obvious that users are interested in Wi-Fi connectivity during their whole journey. Other single aspects are also closely related, such as showing alternative routes during the planning phase and information about alternative routes during the journey in case of a traffic jam (0.71; 7.92/7.98), planned restrictions and information about current traffic situation (0.72; 7.53/7.32), Wi-Fi connectivity in the vehicles and electric supply in the vehicles (0.72; 4.00/3.71), information for disabled people and low floor vehicles (0.70). Some less stronger connections were discovered between information about actual traffic situation and estimated delays (0.63), and between showing prices during the planning phase and general information about prices (0.60). These connections are in line with general expectations, which mean that the users provided unambiguous answers.

Considering the answers of the users for highly correlated single aspect pairs (Fig 15) it can be stated that the majority of the answers are close to the main axis (one green
rectangle represent an answer of a user regarding the two single aspects, which values are between 1 and 10). However for a) the value distribution between the axis is more uniform (some users are interested in both single aspects, some are not, but at least each user consistently), while for b) and c) the values are shifted to the higher segment (users are very interested in both single aspects), and finally for d) the values are present in the lower segment (users are not interested in both single aspects).

Fig 15 Answers given by the users for two highly correlated single aspects each

- a) Wi-Fi at the stations Wi-Fi in the vehicles
- b) alternative routes during the planning phase and information about alternatives
- c) planned restrictions and information about current traffic situation
- d) Wi-Fi in the vehicles and electric supply in the vehicles

More detailed results of the correlation analysis regarding all user groups can be found in Appendix 2. The calculations of the correlation matrices can be found in Annex 4, where correlations above the threshold are shown in red.

Using correlation analysis some connections between user’s requirements regarding a service could be explored. Knowing particular answers for single aspects, the importance of other single aspects can be derived, which helps operators and developers to determine feature development strategies.

3.3.2 Ward method description

The Ward method was used for classification by all aspects of users and creating new user groups [174]. This analysis provides as an output a number of clusters or groups, in which the users are classified. In an ideal case these groups show high similarity considering the answers of their members.

Ward is a hierarchical clustering method, where the users are partitioned into a dedicated number of clusters in many steps. The number of steps may reach from 1 to n. In case of 1 only one single cluster contains all elements, while in the case of n all elements form an own cluster.

Hierarchical clustering methods can be agglomerative or divisive [69], [97]. Agglomerative methods collect elements into groups (clusters), while divisive methods separate the elements successively into groups. I have reviewed the most important and widely used clustering methods, as single linkage and centroid clustering [72]. For this specific task Ward suits best, as Ward method is conservative, monotonic and creates about same big groups, but is sensitive to outliers [2], [3].
Ward method is agglomerative, thus first each element is independent, and then step by step more elements will be ordered to a cluster. At each step the method includes those elements, which are the “closest” (according to a metric) to the existing clusters. Once a unified cluster is created as a result of a step, the elements of the new cluster cannot be separated again.

Usually because of the possible scale differences among the questions (j) the data should be normalized. In our case the maximum of each question was 10, thus this step could be skipped.

\[ m_{kj} = \frac{x_{kj} - x_{mj}}{s_j} \]  

\( x_{kj} \) – answer value (x) of the questions (j) concerning the users (k),  
\( x_{mj} \) – average (m) of the answers regarding a question (j),  
\( s_j \) – standard deviance (s) regarding a question (j),  
\( m_{kj} \) – normalized answer value (m).

The distance \( d \) between two users or clusters (k and l) was calculated with the quadratic Euclidean distance using the normalized values for the total number of questions (q).

\[ d(k, l) = \sum_{j=1}^{q} (m_{kj} - m_{lj})^2 \]  

Users or clusters of minimal distance to each other will be unified into a new cluster (k+l). If the new cluster exists, its distances have to be redefined towards all other users or clusters (a). Different clustering methods use different algorithms for calculation of new distances. In case of single linkage method simply the smallest possible distance is chosen.

\[ d(a, k + l) = \min \{d(a, k), d(a, l)\} \]  

Ward method calculates the optimal minimum distance taking into account the number of users in the clusters.

\[ d(a, k + l) = \frac{(N_a + N_k) \cdot d(a, k) + (N_a + N_l) \cdot d(a, l) - N_a \cdot d(k, l)}{N_a + N_k + N_l} \]  

\( N_a \) – number of users (a) in the cluster,  
\( N_k \) – number of users (k) in the cluster,  
\( N_l \) – number of users (l) in the cluster.

3.3.3 Creation of new user groups using Ward method

The clustering algorithm was implemented in MatLab working with the original survey answers, which code can be found in Annex 5.
When running the code those answers of the users had to be deleted, which have exactly the same values for each question. As Ward is a clustering method with undetermined number of groups, clustering with different number of Ward groups were run to identify the optimal solution. The aim was to create groups of same sizes and similar opinions. In order to provide good evaluation possibility among the user groups, clusters from 3 to 10 were tested. Many possible solutions were assessed, and finally 5 groups were created, where 4 were clustered by Ward and other smaller groups with distinct answers were assigned to group nr. 1. The dendrogram was also presented for the Ward user groups (Fig 16), which shows the 4 big user groups and the smaller groups.

Fig 16 Dendrogram of the Ward user groups

The connection between the original user groups and Ward user groups were investigated (Fig 17). The darker the colours, the more users from the original groups were assigned to the specific Ward user group. 36% of the students were assigned to group 2, while their distribution to other groups was quite similar. Workers mainly are find in group 4 (35%), but also group 2 and 5 (25% and 25%). The favorite group of tourists was group 4 (30%), while businessmen were more equally distributed, only group 1 was less preferred (5%). Pensioners went mainly to group 3 and group 4 (33% and 25%). In general some trends are present, but no definite connection could be found, therefore new group names were identified based on user preferences in the groups.

<table>
<thead>
<tr>
<th>Original user groups</th>
<th>Ward user groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>student</td>
<td>11</td>
</tr>
<tr>
<td>worker</td>
<td>10</td>
</tr>
<tr>
<td>tourist</td>
<td>18.5</td>
</tr>
<tr>
<td>businessman</td>
<td>11</td>
</tr>
<tr>
<td>pensioner</td>
<td>8</td>
</tr>
</tbody>
</table>

Fig 17 Correspondence between the original user groups and Ward user groups
In order to present typical preferences of Ward user groups, those single aspects were chosen, which had the lowest standard deviation among the questions (considering this per user group). The lower the deviation, the more the users of a group agree on a single aspect. The grouping helps also to reveal connections between the transportation process and information handling process. With the stated preferences of the users they can be classified by single aspects. Therefore the single aspect with the lowest deviation can be chosen as a typical preference of the group.

Standard deviation values were relatively high in case of group 1, which was expected, as this group is a merged group of other smaller groups. Undoubtedly the lowest single aspect was alternative route suggestion during the journey (average: 8,80; standard deviation: 1,97). For group 2 much lower values were present, the lowest was visualization on the map (9,18; 1,04). For group 3 planned restrictions (9,97; 0,94) and estimated delays (8,82; 1,00) were most coherent. For group 4 mobile payment was similarly not important (1,33; 1,15). For group 5 WiFi at stations (1,54; 0,9) and WiFi in the vehicles (1,69; 0,93) were equally not relevant. Actually travel time and distance, and timetable had one of the lowest deviations for all user groups, but these single aspects could be not considered as a distinctive feature. Based on these single features new names were assigned to the Ward user groups: alternative route planning, visualization on the map, dynamic info group, no mobile payment, no WiFi interested group.

The 5 most important single aspects were collected for the Ward user groups to demonstrate their features (Table 9). A more detailed analysis of similarities within (highlighted with yellow) and differences among (highlighted with orange) the user groups is presented in Annex 6. In the last part of the Annex the standard deviations and mean values among user groups were presented. The red cells highlight extreme values.

<table>
<thead>
<tr>
<th>alternative route suggestions</th>
<th>visualization on the map</th>
<th>estimated delays</th>
<th>mobile payment</th>
<th>WiFi at stations</th>
</tr>
</thead>
<tbody>
<tr>
<td>planning with alternative routes</td>
<td>travel time and distance</td>
<td>travel time and distance</td>
<td>travel time and distance</td>
<td>WiFi in the vehicles</td>
</tr>
<tr>
<td>travel time and distance</td>
<td>timetable</td>
<td>visualization on the map</td>
<td>timetable</td>
<td>address based input</td>
</tr>
<tr>
<td>address based input</td>
<td>alternative route suggestions</td>
<td>planned restrictions</td>
<td>address based input</td>
<td>electric supply in the vehicles</td>
</tr>
<tr>
<td>estimated delays</td>
<td>planning with alternative routes</td>
<td>number of transfers</td>
<td>luggage storage</td>
<td>timetable</td>
</tr>
</tbody>
</table>

Table 9 Features of Ward user groups

The preferences of the Ward user groups are presented concerning the main aspects (Fig 18). The general preferences did not change, handled data (28%) and route planning (24%) are the most important main aspects, the others are ranked as followed, booking and payment (18%), supplementary information (17%), comfort service information (13%).
The alternative group is equally interested in booking and comfort service information (19% and 19%), which is an extreme difference for the latter main aspect. This is caused by the various composition of this specific group. Therefore service providers should consider feature development (e.g. WiFi information and weather forecast) for this group. The map group provided values closest to the averages, thus they represent the average user. Handled data and route planning is far more interesting to dynamic users (32% and 27%) than to other groups, so these features should be highlighted. For the no payment group supplementary information is much more important, than comfort service information (21% and 9%), applications including customer service features and information for disabled people should be developed. The no WiFi group is mainly interested in route planning and least in comfort service information (26% and 10%). Based on these figures the main requirements of Ward user groups could be identified, which can help operators to provide specific feature development for different user groups.

3.4 Evaluation of journey planners with user groups

In order to apply the results of the Ward user groups, the evaluation of the multimodal journey planners was performed. The same evaluation method was applied for the same journey planners introduced in the previous chapter.

In the case of original user groups the transportation share of the defined groups was available through the national survey. In the case of Ward user groups such statistics do not exist, therefore this value had to be calculated (Fig 19). During the calculation rounding was applied. I used the connection table between the original user groups and the Ward user groups and based on the distributions the transportation shares were defined: 12% for alternative group (1), 27% for map group (2), 14% for dynamic group (3), 26% for no payment group (4), 21% for no WiFi group (5).
Fig 19 Calculation of transportation shares of Ward user groups

Based on the averages of the answer values concerning main aspects, the weights (based on the average answers of the users) were assigned to each Ward user group (Table 10). In the rows the Ward user groups, in the columns the main aspects, while in the last column the transportation shares are represented. The differences of the values among main aspects of Ward user groups are remarkable higher, especially for handled data and comfort service information, however the values are very similar in case of booking and payment. This means that in the latter case the expectations of the users are not different, and a feature development for booking and payment will appreciated equally for each user group. However in case of other main aspects the user groups will differently appreciate a feature development.

<table>
<thead>
<tr>
<th>Ward user groups</th>
<th>%</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>student</td>
<td>3.3</td>
<td>10.8</td>
<td>4.2</td>
<td>5.8</td>
<td>5.8</td>
<td>30</td>
</tr>
<tr>
<td>worker</td>
<td>3.0</td>
<td>7.5</td>
<td>1.5</td>
<td>10.5</td>
<td>7.5</td>
<td>30</td>
</tr>
<tr>
<td>tourist</td>
<td>2.8</td>
<td>3.3</td>
<td>1.7</td>
<td>4.4</td>
<td>2.8</td>
<td>15</td>
</tr>
<tr>
<td>businessman</td>
<td>1.1</td>
<td>2.6</td>
<td>1.9</td>
<td>1.9</td>
<td>2.6</td>
<td>10</td>
</tr>
<tr>
<td>pensioner</td>
<td>1.3</td>
<td>2.5</td>
<td>5.0</td>
<td>3.8</td>
<td>2.6</td>
<td>15</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12</strong></td>
<td><strong>27</strong></td>
<td><strong>14</strong></td>
<td><strong>26</strong></td>
<td><strong>21</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Table 10 Weights of main aspects according to Ward user groups and transportation shares

Having the new weight matrix on the survey results, the evaluation of multimodal journey planners have been updated (Fig 20). The values represent in percent (%), how close the journey planners are to the theoretical optimal journey planner. In the weighted case (average evaluation number) some aspects gained more importance in contrast to other aspects based on Ward user preferences. Respectively journey planners, where aspects of higher importance are implemented, are also higher valued.

The results show an average change of 8.16% in the evaluation numbers with Ward user groups. The biggest changes (more than 10%) can be observed in case of BKK, TfL, AnachB, Google maps and DB, which means that such features were implemented in the journey planners, which are important for the Ward user groups.
Considering the difference between the evaluation with estimated weights and Ward weights (Table 11) it can be obtained that in most cases the change was positive, which means that Ward user groups generally require such features, which are present in current journey planners. The biggest change is present for BKK, AnachB, 9292 and Google maps (with more than 3%). Only some journey planners (Rome2Rio, Scotty, Lufthansa) did not possess such implemented features, which were really important for the Ward user groups. In these cases a negative difference could be observed, which means that in these journey planners the functionality of their service according to the user requirements uncovered by the survey should be improved.

Table 11 Comparison of changes in the evaluation using estimated and Ward weights
3.4.1 Future research perspectives

Knowing the highly correlated single aspect pairs and creating new user groups with Ward method provides important, but limited information about possible user needs. If the importance of a single aspect is known, the high correlation can only refer to the importance of another single aspect. The aim is to extend this knowledge to more single aspects, in order to find correspondences among more single aspects.

The research question is how to enable a reduction of questions to be answered by a new user, while keeping to maximum the clustering correctness of new users to the existing groups. This would help the effectiveness of the survey and raise the willingness of filling in the questions.

In order to define latent demands of the users those questions have to be chosen, which have the biggest difference among the user groups. This means that the answers for the chosen questions are the most different, thus users can be easier separated to the existing groups. The difference of the answers can be defined by the ratio of the absolute difference of the averages and the standard deviation. By summing up some answer differences a value is present for each user groups. Finally the new user will be assigned to that user group, which has the minimum average difference.

To test the validity of group assignments, the original survey data has to be divided into learning and test sets. The learning set will represent existing user groups, and will be the basis for calculating values for answer differences. Users of the test set will represent new users, who have to be assigned to an existing user groups, therefore their answers will be tested by the algorithm and assigned based on the answer differences. The soundness of the algorithm will be measured by the number of correct assignments. This correctness could be improved by including more questions in the calculation of answer differences. In an ideal case the new users could be assigned to existing user groups based on 5 to 7 questions.

3.5 Elaboration of a survey method using Fuzzy AHP

The definition of the weights of the user groups is a cardinal issue of the evaluation of journey planners. The weights can be obtained through surveys among the users, however it is not easy to determine the weights of the main aspects in many cases. Analytic hierarchy process (AHP) can facilitate the decision making of users by reducing complex decisions to a set of pairwise comparisons (instead of evaluations on a scale of 1-10). A new survey method can be introduced by application of the Fuzzy AHP method, and accordingly weights for the main aspects can be defined in a more consistent way. Thus the results regarding preferences of user groups can be “smoothened”, namely similar results can be obtained as in the case of the original survey.

There has been a growth in the number of multiple-criteria decision-making (MCDM) methods to evaluate several alternatives to achieve a certain goal. The analytic hierarchy process (AHP) was developed by Saaty [144],[143] and is one of the most widely used methods, as it can solve complex decision problems, e.g. in transport infrastructural, economic and survey evaluation problems. However this method is often criticized for its uncertainty and imprecision associated with the mapping of users’
perception. This is because user judgments are represented as exact or crisp numbers, while in many practical cases preferences are uncertain or users might not be able to assign exact numerical values to the comparison of aspects [125]. There are several methods to obtain priorities in a softer way, e.g. using a fuzzy toolkit. Among others Xu [180] introduced a method, which estimates weights of aspects by least squares from a fuzzy matrix. Buckley [20] presented a geometric mean method and Wang [173] implemented a two stage logarithmic programming application. The fuzzy AHP approach introduced by Deng [38] can tackle qualitative problems of the pairwise comparisons, which provides a more efficient decision making process in terms of uncertainty of human behaviour, which can cause imprecision in the evaluation. Mikhailov [124] presented a fuzzy method, which can derive priorities using fuzzy pairwise judgments, and also implemented a programming method for this [123]. Van Laarhoven and Pedrycz [169] extended the fuzzy hierarchical analysis using comparison matrices with triangular fuzzy numbers. Also Wang et. al. [172] discussed the extension of the AHP method with fuzzy comparison matrices. Gumus [86] combined fuzzy AHP and TOPSIS methods in order to rank aspects of waste transportation firms. Farkas [71] investigated the consistency of AHP pairwise ranking matrices.

A very detailed review of AHP applications is written by Vaidya and Kumar [166]. Some studies consider problems of urban mobility and transportation issues in the broader context, as Campos et. al. [25] is dealing with multi-criteria analysis of sustainable mobility evaluations, where environmental, economic and social aspects are weighted. Markovits-Somogyi [117] applied data envelopment analysis (DEA) to evaluate efficiency of decision making, where the aim was to provide full ranking of transport companies. In many cases DEA is integrated with AHP. Kaur and Kumar [98] cover transportation problems using raking function, and Teng and Tzeng [158] rank urban transport investment alternatives using AHP. Duleba et. al. [44] considered quality of public bus transportation by using AHP on a strategic level. Their solution supports the decision making process for public transport developments. Sivilevicius and Maskeliunaite [149] calculated ranking values of the quality of passenger transportation focusing on railway, which helps taking into account the interests of passengers.

Also Longo et. al. [113] investigate an issue of mobility management at a university campus, where user groups are defined, a survey is created and the general AHP method is applied in order to understand user preferences regarding transportation modes. In the paper of Shafabakhsh et. al. [146] the appropriate public transport system was chosen using a fuzzy decision making method.

Many application fields are present, even for transportation issues, but the fuzzy AHP method was not really used for weight definition in evaluation of multimodal journey planners from the user perspective. Looking for useful applications the approach introduced by Chang [29] is used in this research, where triangular fuzzy numbers are generated. This method was also presented by Catak et. al. [26] for a problem of database management system selection for national identity card management. The steps of the Fuzzy extend analysis are followed, while AHP values are defined for the special problem, and Fuzzy weights are computed for the evaluation of journey planners.
The Analytic Hierarchy Process (AHP) is an effective tool for dealing with complex decision making problems, and helps to set priorities among different alternatives (e.g. journey planners). The results of AHP (which are in our case the weights for the user groups) are provided by the users on the basis of pairwise relative evaluations of the aspects. While in the original survey all aspects had to be considered at the same time, and it could happen that the user focused first only on a few aspects, and would have evaluated differently, if knowing all aspects. The method is also able to check the consistency of the pairwise comparison values, which reduces the possibility of contradictory results.

Fig 21 Process of evaluation using Fuzzy AHP method

The Fuzzy AHP method is a more advanced analytical method, which is developed from the AHP. The fuzzy comparison ratios should be able to tolerate imprecision and overcome uncertainties of AHP. However the experience of the user is crucial, as the results of the pairwise comparisons high dependent on the answers of the users. In order to reduce the effect of individual experience of the users, averaged values derived from the answers of the user groups were taken into account during the calculations. In our case the AHP and the Fuzzy AHP methods are used only to calculate the weights of the main aspects regarding the user groups, the evaluation of the journey planners is performed based on the compensatory multi criteria evaluation method (Fig 21).

3.5.1 Pairwise comparison

In the course of the comparison the user has to decide between two aspects, which one is more important, and how much (for that importance levels are defined). Usually 5-6 importance levels are created, in our case they are: equally, slightly, more, strongly and absolutely important. The AHP method generates a value for each main
aspect according to the pairwise comparison results. The higher the value, the more important is the corresponding aspect. This calculation process is performed for each user group separately. An example comparison table is presented containing the main aspects of the evaluation (Table 12).

<table>
<thead>
<tr>
<th>Main aspect</th>
<th>Aspects of comparison</th>
<th>Importance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>route-planning service</td>
<td>X booking and payment</td>
<td>X</td>
</tr>
<tr>
<td>route-planning service</td>
<td>X handled data</td>
<td>X</td>
</tr>
<tr>
<td>route-planning service</td>
<td>X comfort service information</td>
<td>X</td>
</tr>
<tr>
<td>route-planning service</td>
<td>X supplementary information</td>
<td>X</td>
</tr>
<tr>
<td>booking and payment</td>
<td>X handled data</td>
<td>X</td>
</tr>
<tr>
<td>booking and payment</td>
<td>X comfort service information</td>
<td>X</td>
</tr>
<tr>
<td>booking and payment</td>
<td>X supplementary information</td>
<td>X</td>
</tr>
<tr>
<td>handled data</td>
<td>X comfort service information</td>
<td>X</td>
</tr>
<tr>
<td>handled data</td>
<td>X supplementary information</td>
<td>X</td>
</tr>
<tr>
<td>comfort service information</td>
<td>supplementary information</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 12 Example comparison table with main aspects

This step of the method may require a large number of comparisons by the user, especially for problems with many aspects. Although the single evaluation is simple, since it only requires comparing two aspects to each other, the load of the evaluation task may become unreasonable. The number of pairwise comparisons grows quadratically with the number of aspects.

3.5.2 Calculation of AHP values

In order to compute weights for the main aspects, following the AHP method, a value matrix $A$ is created, which contains the AHP values. This matrix has a size of $m \times m$, where $m$ is the number of main aspects. Each element $(a_{ij})$ of the matrix represents the importance of the $i^{th}$ main aspect relative to the $j^{th}$ main aspect. If $a_{ij} > 1$, then the $j^{th}$ main aspect is more important, than the $k^{th}$ main aspect, while if $a_{ij} < 1$, then the $j^{th}$ main aspect is less important than the $k^{th}$ main aspect. If the two main aspects have the same importance, then the element $a_{ij}$ is 1. The elements $a_{ij}$ and $a_{ji}$ satisfy following constraint: $a_{ij} \times a_{ji} = 1$. Therefore it is enough to fill the upper triangle of the matrix, as the evaluation of the main aspects is symmetric. Furthermore the main aspects have the same importance regarding themselves, therefore $a_{ij} = 1$ for every main aspect $j$.

$$A = \begin{bmatrix}
1 & a_{12} & \cdots & a_{1m} \\
a_{21} & 1 & \cdots & a_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{12} & \cdots & 1
\end{bmatrix}$$ (12)
The possible values for AHP values are set from 1 to 6. Value 1 represents the case, when the same main aspects are considered. Value 2 means that the difference between the two main aspects is negligible. Value 3 stands for a slight difference, values 4 and 5 represent more important differences, while value 6 means an absolute big difference between the two analyzed main aspects.

In our case the results of the online survey were used, where participants were assigned to the user groups, and answered questions connected to the aspects. Based on the answers the importance among the main aspects has been derived, which value is represented in percentage. The survey did not contain direct questions for the pairwise comparisons, therefore the AHP values were calculated pairwise from the proportions of the examined main aspects (Table 13). Considering the values for proportion more combinations were sampled, and the one was chosen, which provided the closest results to the original weights.

<table>
<thead>
<tr>
<th>Proportion</th>
<th>AHP value</th>
<th>Explanation</th>
<th>Proportion</th>
<th>AHP value</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>j and k are the same</td>
<td>1</td>
<td>1</td>
<td>k and j are the same</td>
</tr>
<tr>
<td>1,5</td>
<td>2</td>
<td>j and k are equally important</td>
<td>0,6667</td>
<td>1/2</td>
<td>k and j are equally important</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>j is slightly more important than k</td>
<td>0,5</td>
<td>1/3</td>
<td>k is slightly more important than j</td>
</tr>
<tr>
<td>2,5</td>
<td>4</td>
<td>j is more important than k</td>
<td>0,4</td>
<td>1/4</td>
<td>k is more important than j</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>j is strongly more important than k</td>
<td>0,3333</td>
<td>1/5</td>
<td>k is strongly more important than j</td>
</tr>
<tr>
<td>100</td>
<td>6</td>
<td>j is absolutely more important than k</td>
<td>0,01</td>
<td>1/6</td>
<td>k is absolutely more important than j</td>
</tr>
</tbody>
</table>

Table 13 Calculation of AHP values

3.5.3 Calculation of AHP weights

Once the matrix A is built, it is possible to derive from it the normalized pairwise comparison matrix \( \bar{A} \). The matrix elements are computed as follows, where \( m \) is the number of the main aspects. Finally, the weights for the main aspects (\( w_j \)) are calculated by averaging the elements on each row of \( \bar{A} \).

\[
\bar{a}_{jk} = \frac{a_{jk}}{\sum_{k=1}^{m} a_{tk}} \tag{13}
\]

\[
w_j = \frac{\sum_{j=1}^{m} \bar{a}_{jj}}{m} \tag{14}
\]

3.5.4 Calculation of Fuzzy weights

The first step of converting AHP values is the fuzzification of the original AHP elements (\( a_{jk} \)), which generates the fuzzy elements (\( \bar{a}_{jk} \)). The membership function (\( \mu \)) defines the degree of truth, which indicates how much the value is included in the
category [182]. The membership function covers the interval \([0,1]\), and can have more definitions [142]. For fuzzification the most popular definition, the triangular membership function, is used. In this function each element has a lower limit \((l_{jk})\), an upper limit \((u_{jk})\) and the point \((m_{jk})\), where the membership function \((\mu)\) is 1.

\[
\tilde{a}_{jk} = (l_{jk}, m_{jk}, u_{jk})
\]  \hspace{1cm} (15)

The fuzzification was performed by using triangular fuzzy scales with values shown in Table 14. The same categories were used as for the AHP values.

<table>
<thead>
<tr>
<th>Proportion</th>
<th>Fuzzy value</th>
<th>Explanation</th>
<th>Proportion</th>
<th>Fuzzy value</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1,1,1)</td>
<td>(j) and (k) are the same</td>
<td>1</td>
<td>(1,1,1)</td>
<td>(k) and (j) are the same</td>
</tr>
<tr>
<td>1,5</td>
<td>(1/2,1,3/2)</td>
<td>(j) and (k) are equally important</td>
<td>0,6667</td>
<td>(2/3,1,2)</td>
<td>(k) is slightly more important than (j)</td>
</tr>
<tr>
<td>2</td>
<td>(1,3/2,2)</td>
<td>(j) is slightly more important than (k)</td>
<td>0,5</td>
<td>(1/2,2/3,1)</td>
<td>(k) is slightly more important than (j)</td>
</tr>
<tr>
<td>2,5</td>
<td>(3/2,2,5/2)</td>
<td>(j) is more important than (k)</td>
<td>0,4</td>
<td>(2/5,1/2,3)</td>
<td>(k) is more important than (j)</td>
</tr>
<tr>
<td>3</td>
<td>(2,5/2,3)</td>
<td>(j) is strongly more important than (k)</td>
<td>0,3333</td>
<td>(1/3,2/5,1/2)</td>
<td>(k) is strongly more important than (j)</td>
</tr>
<tr>
<td>100</td>
<td>(5/2,3,7/2)</td>
<td>(j) is absolutely more important than (k)</td>
<td>0,01</td>
<td>(2/7,1/3,2/5)</td>
<td>(k) is absolutely more important than (j)</td>
</tr>
</tbody>
</table>

Table 14 Calculation of Fuzzy values

As Fuzzy AHP provides smoother results the calculation of weights were also processed with this method. The same proportions were used to provide a good basis of comparison of the AHP weights and Fuzzy weights. The calculation method of Fuzzy weights \((w_j)\) is similar to the AHP method, but the processing of Fuzzy operations requires more complicated calculations, which are described in details in Appendix 4.

3.5.5 Consistency check

The last step of the AHP method is checking the consistency of AHP values. This step should reveal inconsistent answers by the users during the pairwise comparisons. As for the calculation of both the AHP weights and the Fuzzy weights the same AHP values were used, it is enough to perform the consistency check once for both methods.

The consistency index \((CI)\) is calculated by obtaining the maximal eigenvector \((x)\) of the value matrix \(A\), and then subtracting the number of main aspects \((m)\). This value is divided by \(m-1\).

\[
CI = \frac{x - m}{m - 1}
\]  \hspace{1cm} (16)

The best consistency is achieved, when the value of CI is 0. The answers of the users can be considered as consistent, if the ratio of the consistency index \((CI)\) and random index \((RI)\) are less, than 0.1. In this case reliable results can be expected from the AHP method.
The RI represents a value, when the elements of the value matrix \( A \) are totally random [138]. In our case the number of main aspects is 5, which means that the respective RI value is 1.12. The consistency check is beneficial to filter different answers, which may often arise, when many pairwise comparisons are performed. In our case the AHP values were calculated from the survey results, therefore a really low consistency value was expected.

3.6 Evaluation of the survey method benefits

First the consistency was checked, which resulted in \( \frac{CI}{RI} = 0.021-0.025 \) for the user groups. This is under the defined limit \( \frac{CI}{CR} = 0.1 \), therefore it can be claimed that the results are consistent and further investigations can be proceeded.

The AHP weights and Fuzzy weights were calculated and compared to the original weights. The original weights come from the results of the survey, where the answers for each single aspect were averaged and the weights of the main aspects were calculated for each user group. The calculations were performed by using MatLab software and the code of Catak [26], who developed it for both the AHP and the Fuzzy AHP weight calculations.

In general it can be observed (Fig 22) that the AHP generated more divergent weights compared to the other two possibilities, while Fuzzy weights resulted in a smoother distribution among the main aspects. Therefore using Fuzzy weights is advisable, as they represent closer results to the original weights. More detailed calculations regarding the main aspects and user groups can be obtained in Appendix 4.

![Fig 22 Differences between original weights and AHP weights, respectively original weights and Fuzzy weights](image)
The differences were also analyzed regarding the user groups (Fig 23), so that differences among the user groups can be detected. It has been found that the order of the main aspects unchanged in all the three cases. The differences among the user groups are slightly higher for Fuzzy weights, especially in case of route-planning services, handled data, comfort services; however this does not cause big changes in the preference of the user groups. Based on the results it can be claimed that Fuzzy weights are good approximations for the original weights. The application of surveys with pairwise comparisons produces good results. Also the inconsistent answers of users are to be filtered, therefore more reliable survey results can be obtained.

Fig 23 Comparison of the results: original weights, AHP weights and Fuzzy weights
4 Optimization method development of daily activity chains

The most important main aspects for the travelers were route planning and handled data, more precisely regarding single aspects, travel duration, visualization and alternative routes. Knowing these preferences, my aim was to provide an optimization method of their activity chains during the day. The development of the method can be performed by applying LBS (Location Based Services) technology, which uses positioning and geographical data of a trip in order to utilize demanded services of the passengers.

The novel positioning solutions, the penetration of social networking and mobile devices produce huge amount of location information. Collecting this information generates mobility patterns. The analysis of these patterns is a research topic since a long time [83],[161], but a significantly enhanced interest can be observed in the last decade.

Concerning daily activity chains previous studies claim that particular person’s daily routes are really similar to each other, and according to González, Hidalgo and Barabási [83] the chosen routes show a high degree of spatial and temporal regularity. That means the passengers are very likely to use the same routes and visit the same places frequently, which allows providing reliable forecast based on mobility data from the past. In the paper of Phitakkitnukoon et al. [135] a mobility analysis can be found. The main goal was to describe daily activity patterns of travelers, who usually visit the same working area. Using correlation analysis the results were shown on an activity map. Individuals have similar activity patterns, if they are close to each other. In the study of Bradley and Vovsha [18] the interactions between members of the households were analyzed. A group decision making process was taken into account, and patterns were derived for the individuals. Furthermore a statistical analysis of interactions and a choice model based on the utility (with person specific components) of travelers was presented.

Keynon and Lyons [100] wrote about a concept of integrated multimodal traveler information, which is exactly the key of the optimization of journeys. With different amount of provided information different decisions were made concerning the route of the journey. The research showed that the passengers choose their routes based on their habits, and do not optimize them according to any decision model, however it would result in optimized routes. Golob [81] described the Structural Equation Modeling (SEM) technique for travel behavior. SEM provides an accurate estimation of travel behavior and with it dynamic travel demand modeling can be performed.

Timmermans et al. [160] reported an international comparison of travel patterns. The aim of the research was to understand the connection of urban structure and travels of individuals. They used data from different data collections with different methods and compared them applying a unified methodology. The results indicated that in general travel pattern is independent from urban structure. Song et al. [153] investigated the limits of predictability in human behavior. They also used routes based on mobile phone information, and were able to predict journeys. They claim that there are many different travel patterns for the passengers, but they are well predictable, furthermore the results are independent from the distance.
The organization of the daily activity chains has been analyzed in several articles [160], [127] and books [161]. Analyzing the regularity of chains some periodically repeated activity can be revealed (e.g. going to the office), which depends on the demographical [101], on the spatial situation [22] and on the personal characteristics of the user [96], [7]. Several measurements were conducted in order to define the visited points, the average travel distance and time [95], and in general the way of organizing the chains [131], [42]. Furthermore articles were written in topics related to spatial and temporal solutions [43], to dynamic planning [141], [132], [116] and to resolving possible scheduling conflicts [10].

For sorting and ordering activities the TSP (Traveling Salesman Problem) method offers a solution [139], [6], whose popular version is often called VRP (Vehicle Routing Problem) [163], [80]. The solution for this problem was developed already in the 1960s, and since then numerous versions were implemented. The basic problem is that an order has to be set among the points to be visited according to a specified aspect. This aspect could be travel distance, cost, number of transfers or the combination of these.

Basically the TSP method is used in logistics systems, but here an application for passenger transportation is proposed. Having demand points of activities for the passengers instead of demand points of goods causes no significant changes in the methodology. The main optimization parameter was time. In our case an extra constraint was defined, because the opening times of the shops and institutions have to be considered. Many optimizing algorithms were considered, but TSP has the widest literature and application, therefore the proposed algorithm is based on the TSP-TW (TSP with Time Window) version. Numerous articles deal with the problems and solutions of the TSP-TW method [11], [104], [45], [145], [103], [79], [36].

Considering extensions genetic algorithms are widely used to solve complex problems in travel behavior and optimization. Kazemi et. al. [99] applied genetic algorithm for supply chain management, where the production and distribution problem was solved. This approach handles location and time constraints of freight transport vehicles. Pattanik et. al. [134] described a method for transit routes for buses in urban environment. The routes of buses and frequencies were optimized with a cost function of overall transport costs. Also activity plans were generated by genetic algorithms according to Charypar and Nagel [30]. They developed daily activity plans, where they focused on activity based transport demand generation. Opening times of the shops and priorities were also taken into account, but the demand points were not really flexible.

4.1 Definition of activity chains and constraints

In order to build the optimization algorithm, the activity chains have to be defined, the basic structure of the graph model has to be described and the most important constraints have to be specified.

4.1.1 Definition of a database structure of activity chains

The first step of modeling of activity chains is the creation of a database structure to store the data (Fig 24). The model contains a Traveler table, where data about the traveler are stored (e.g. age, gender, preferences). Information about the demand points is
contained in the Demand point table, which includes the name of the demand point, its location (latitude and longitude), the opening hours and other information (e.g. description, picture). These demand points are classified into POI types (Point Of Interest, e.g. restaurant or post office), which facilitate finding alternative demand points, as travelers often do not search for specific locations, but for types of activity. The most important table is the Activity table, where travelers can construct their daily activity chains. To each activity belong a POI type, a priority (P), a processing time (TP) and a demand time window (TD) with earliest start time and latest end time.

Fig 24 Database structure of activity chains

The source of the data input has to be also defined. Information of the Traveler and Activities table is provided by the traveler. It is assumed that the traveler knows the activities of the certain day in advance and prepares a list of them. Meanwhile information of Demand point and POI type table are retrieved from an external database. The source of these data was the POI information provided by OSM (Open Street Map).

4.1.2 Description of the Traveling Salesmen Problem

The optimization algorithm of daily activity chains relies on the solution of the Traveling Salesman Problem [118], which problem can be defined as an undirected weighted graph (G). The vertices (V) are the demand points and the edges (E) are the trips between demand points. The cost matrix (C) is defined on E, and its elements are travel times between the demand points, but can also be represented as a utility function with an abstract measure.

\[ G = (V, E) \]  \hspace{1cm} (18)

\[ V = \{1, \ldots, n\} \]  \hspace{1cm} (19)

\[ E = \{(i, j), \text{ where } i, j \in V \text{ and } i \neq j\} \]  \hspace{1cm} (20)

\[ C = C(i, j) \]  \hspace{1cm} (21)

The goal of the optimization with TSP is to minimize the cost of the set of trips, where \( x_i \) is a binary variable with the value 1, if the trip is part of the optimal activity chain and the value 0, if it is not.

\[ \min \sum c_{ij} \cdot x_{ij} \]  \hspace{1cm} (22)
4.1.3 Constraints of the TSP-TW

The following constraints were defined when solving the basic TSP problem:

- the traveler leaves every demand point,
  \[ \sum_{i=1}^{n} x_{ij} = 1, \text{where } i \in V \text{ and } i \neq j \]  

- the traveler arrives to every demand point,
  \[ \sum_{i=1}^{n} x_{ji} = 1, \text{where } j \in V \text{ and } i \neq j \]  

- disjunctive partial activity chains are not allowed, therefore the traveler must always visit a demand point, which was not visited before,

- the elements of the cost matrix satisfy the triangle inequality,
  \[ c_{ij} \leq c_{ik} + c_{kj}, \text{for } \forall i, j, k \]  

- the elements of the cost matrix are not negative.
  \[ c_{ij} \geq 0, \text{for } \forall i, j \text{ and } i \neq j \]  

Furthermore the following constrains have to be fulfilled when using the graph model with Time Windows (Fig 25):

- real time window (TR) is defined as a difference of the time window (TW), which is composed by the opening times of the demand points, and the processing times (TP). Processing time is the time needed for travelers to execute operations and use the provided services (e.g. shopping). The processing times are static, therefore queues or delayed services are not considered.
  \[ TR = TW - TP \]  

- each time window (TW) is at least as long as the processing time (TP),
  \[ TW \geq TP \]  

- a demand point is valid, if the real time window (TR) and the processing time (TP) suits in the demand time window (TD) of the traveler,

- each demand point has to be reachable during the travel time (TT) between demand points,

- the total travel time (T) is the sum of all travel times (TT) and potential waiting times (\( T_{wait} \)), which can occur when arriving earlier to a demand point.
  \[ T = \sum TT + \sum T_{wait} \]
4.2 Elaboration of the optimization algorithm

The nowadays available methods, which are defined as basic TSP, are based on a transportation network and demand points, which have to be explored and cost functions, which describe the values among the demand points. A TSP method is considered flexible, when according the subjective demands of the travelers demand points can be arbitrarily replaced with another point with the same function. Therefore using the flexible TSP method a solution could exist, which would not exist using the basic TSP. The added value of the proposed method is the extension of TSP-TW with flexible demand points, which are variable in time and space, the implementation of the algorithm in Matlab environment, and the application of a genetic algorithm. The concept of the method was modeled and consists of the following steps (Fig 26).

- definition of daily activity chains and constraints,
- prioritization of activities,
- calculation of travel times,
- solving the basic TSP-TW,
- solving the flexible TSP-TW,
- optimization and visualization.

4.2.1 Prioritization of activities

Activity chains contain all regular (e.g. school, workplace) and non-regular (e.g. eating out at a restaurant) activities of a traveler. The spatial and temporal parameters of regular activities are usually fixed, while in many cases non-regular activities are flexible. To each activity a priority is assigned, which represents its importance. The prioritization is realized according to personal decisions of the traveler.
Fig 26 Model of the activity chain optimization algorithm
Priority can have the following values (Table 15):

- 1: fix, which has to be definitely arranged in the predefined location and demand time window,
- 2: spatially flexible, which means that the demand time window is fix, but the location is not set,
- 3: temporally flexible, where the location is fix, but the demand point can be visited anytime during the day,
- 4: totally flexible, which means that the activity could be shifted to another day, if necessary (e.g. if reaching the demand point would take more than a predefined value of travel time).

<table>
<thead>
<tr>
<th>Priority value</th>
<th>Priority name</th>
<th>temporal flexibility</th>
<th>spatial flexibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>fix</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>spatially flexible</td>
<td>-</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>temporally flexible</td>
<td>x</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>totally flexible</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Table 15 Flexibility of priorities

In case of priority 1 the location, earliest start time and latest end times are set, respectively the demand point is fixed and has to be visited according choice of traveler.

In case of priority 2 all nearby demand points are searched by the algorithm, which are within a predefined distance (e.g. 1000 m) in straight line and have the same POI type (e.g. bar). The alternative demand points were searched by Overpass API (Application Programming Interface) based on OSM (Open Street Map). The Overpass API provides location coordinates of nearby alternative demand points of the same POI type. These alternative demand points are searched around every fix demand point. In case of closely located fix demand points same alternatives are filtered. From the examined alternative demand points those are considered, which average distance is minimal to the fix demand points. This measure was implemented because the chosen alternative demand point shall be more likely close to routes among other demand points. In the algorithm the number of chosen alternative demand points can be set. Having more demand points of priority 2, all alternative combinations are calculated.

In case of priority 3 only temporally flexibility was considered, where the end time was left empty by the traveler. This parameter is set by the algorithm to 23:59 of the chosen day. After that from the algorithmic point of view, the demand point is handled as a fix demand point of priority 1.

In case of priority 4 the algorithm considers traveler preferences (e.g. the traveler would like arrive home before a certain time). If the preference is not fulfilled, the algorithm skips the totally flexible activity (this activity drops out from the actual daily activity schedule and is postponed for another day), and recalculates the whole daily activity chain. Iteratively totally flexible activities can be skipped until the preference is fulfilled or totally flexible activities are run out.
4.2.2 Calculation of travel times

In order to realize activities in different locations, travel time matrices between demand points were defined. The location changes can be realized by 3 different transportation modes: car, public transport and combined (public transport with car-sharing opportunity).

**Car mode:**

Considering car mode travel times are provided by Google API, where an average walking time (e.g. 5 minutes) to the parking spot is added. As the setting of arbitrary start times in Google API is not enabled, actual traffic situation was not included in the travel time calculation process. A possible solution for this could be the usage of time dependent additions to the calculated travel time matrices. However, this would present only an average value, which would not be specific for the chosen route.

**Public transport mode:**

In case of public transport mode travel time calculation is performed also by Google API using actual start times. The data source for this is GTFS (General Transit Feed Specification), which input is provided by local transport authorities. In Budapest GTFS data are available and regularly updated. However, these data are not real-time, but in the future GTFS-RT (GTFS-Real-Time) should be implemented and widespread used, which would allow calculation with real-time schedules. The criterion of the chosen route is the minimal travel time. Currently no other criteria, as walking distance or number of transfers, are taken into account.

**Combined mode:**

Calculating with public transport and car-sharing combined, travel times for each route between two demand points are requested through Google API, both for public transport and for car-sharing mode. Then for each route the better option is chosen (the transportation mode with less travel time). Regarding car-sharing mode an extra walking time is added to the travel time of the car mode, which is needed to arrive to the closest free car to use. In case there is no available car in a predefined distance, the algorithm uses the public transport mode. In Budapest there is no widely used car-sharing service in operation yet, thus the algorithm assumes an ideal case, and calculates with an average of 5 minutes walking distance to an available car.

4.2.3 Solving the basic TSP-TW

The originally chosen demand points of the activity chain are the inputs for the basic TSP-TW method. During the optimization process all possible combinations are considered and the optimization function is the minimization of total travel time. The algorithm calculates an order of the demand points and respectively a total travel time. This basic scenario is the reference for comparisons to the proposed method with flexible demand points.
4.2.4 Solving the flexible TSP-TW

In case of flexible demand points, the activities can be reached in more locations (e.g. instead of the point C, also in the point C'). A new set of alternative demand points of the same POI type can be considered, which results in new versions of the daily activity chains (Fig 27).

![Fig 27 Presentation of the flexible demand points](image)

The activity chains with flexible demand points are formulated for all possible alternatives. Regarding car mode the travel time matrix is calculated between each demand point for all possible activity chains. Regarding public transport mode no travel matrix is created, but the actual travel times are requested during the activity chain generation. Regarding combined mode (public transport and car-sharing) values of both transportation modes are calculated and compared.

All combinations are tested for the defined constraints (e.g. opening time). Then those activity chains are selected, which correspond with the constraints. In this step are demands of the travelers (demanded time windows) assigned to supplies of the demand points (e.g. opening times). Arriving by a predefined time interval (e.g. 15 minutes) before the requested start time at the demand point is acceptable. If this time interval is longer, the solution will not be considered.

In case of totally flexible demand points (which may be shifted to another day), the traveler preferences are checked. If the preference is not fulfilled, the calculation of possible activity chains is repeated with skipping a totally flexible demand point.

For each valid version (where a new flexible demand point was chosen) the TSP-TW problem is solved with the algorithm and the total travel time is calculated.

4.2.5 Optimization and visualization

After creating all valid activity chains the one with the least total travel time will be chosen as optimal. Finally the set of optimal demand points and total travel time is presented as an output. This is calculated for all transportation modes. Also visualization of the optimal daily activity chain is realized using Google maps. Thus the result of the basic scenario can be compared to the best version of the flexible TSP-TW.
4.3 Extensions of the optimization algorithm with POI search and genetic algorithm

The original algorithm provided good results with some limitations, because the high complexity and long processing times of solving a TSP-TW problem with many flexible demand points. Two extensions were introduced to speed up the processes. A POI search algorithm enabled to search demand points in advance and store them in an offline database. Furthermore a genetic algorithm (GA) was applied and customized to realize lower optimization times.

4.3.1 Extension of the optimization with POI search algorithm

In order to find more alternative demand points a POI search algorithm implemented by Čerticky et. al. [28] was inserted into the elaborated algorithm. The POI search algorithm can retrieve POIs from different databases through Foursquare API, OSM Overpass API and Google API.

The POI search algorithm provides demand points of a predefined area, which could even be a whole city. First POIs through the Foursquare API were collected, where each POI comes with a set of metadata, from which the following types were used: POI name, POI type, latitude, longitude, city, opening times for each day (Table 16). Then additional POIs are added through Overpass API in XML (eXtensible Markup Language) format. The data are extracted using a Java parser, then the new data are merged to the database. Additionally some missing data, usually opening times, are filled using Google API.

<table>
<thead>
<tr>
<th>POI name</th>
<th>POI type</th>
<th>latitude</th>
<th>longitude</th>
<th>city</th>
<th>opening times</th>
</tr>
</thead>
<tbody>
<tr>
<td>IKEA</td>
<td>Furniture</td>
<td>47.5055407</td>
<td>19.1384143</td>
<td>Budapest</td>
<td>0900-2200</td>
</tr>
<tr>
<td>Árkád</td>
<td>Mall</td>
<td>47.5027648</td>
<td>19.1388273</td>
<td>Budapest</td>
<td>0650-2200</td>
</tr>
<tr>
<td>Szentmihalyi Panzlo</td>
<td>Bed &amp; Breakfast</td>
<td>47.5199980</td>
<td>19.1554760</td>
<td>Budapest</td>
<td>0000-2359</td>
</tr>
<tr>
<td>Aquaworld Resort &amp; Spa</td>
<td>Water Park</td>
<td>47.6003696</td>
<td>19.1122208</td>
<td>Budapest</td>
<td>0600-2200</td>
</tr>
<tr>
<td>TESCO Extra</td>
<td>Supermarket</td>
<td>47.5047942</td>
<td>19.1109405</td>
<td>Budapest</td>
<td>0600-2200</td>
</tr>
</tbody>
</table>

Table 16 Structure of the POI database with example data

Search times of the originally used Overpass API and POI database were compared. Using the POI database to search alternative demand points, the search time is constantly 0,14 s, while using Overpass API it depends on the search radius (Fig 28). Even for a radius of 20 m is Overpass API search slower (0,32 s), than the offline database. Considering a more realistic radius (e.g. 1000 m), the value is even higher (0,43 s), because searching in a database is faster, than sending requests through an API every time. For higher radiuses the difference is even bigger, e.g. for 10 000 m 2,6 s.
Considering the number of found alternative demand points, generally 4x more alternative demand points are found using the POI database (for a radius of 1000 m). Using the POI search algorithm for the city of Budapest a database of ca 54,000 demand points was created.

4.3.2 Extension of the optimization with genetic algorithm

GA is an evolutionary algorithm producing heuristic solutions [179], [128], thus it does not provide the optimal solution necessarily, but it produces useful results close enough to the optimal one. The closeness can be defined by setting the parameters of the algorithm.

When using a GA first a random population of predefined size is created [77], where each member of the population represents a possible solution for the TSP-TW problem. This population is evolved by creating new possible solutions. The direction of evolution is determined by the cost or fitness function (which is in our case travel time). Those solutions are chosen, which are closer to the minimum of the cost function. The process is iterative, where each step is called a generation.

For the creation of new solutions GA uses mutation, crossover and selection. In the selection phase those solutions are chosen with higher probability, which are closer to the minimum value of the cost function. For creating new solutions two existing solutions are chosen and their parameters are mixed together. The algorithm terminates when it reaches the maximal number of generation (generation limit), approximates the predefined fitness level or no more improvement of the cost function is possible within a predefined time interval (stall time limit).

The elaborated extension of the optimization algorithm consists of 2 GA-s, an inner GA and a multiobjective GA, which produce Pareto optimal solutions (Fig 29).

**Inner GA:**

The inner GA optimizes a scenario with flexible demand points regarding the cost function, which is in our case total travel time. If the scenario is not fitting the constraints, an infinite penalty is set for the total travel time. Therefore this scenario will
never be optimal. The inner GA works with the following parameters: 100 generations, population size of 60, stall generation limit of 10. The parameters were set to provide short running times and near optimal results.

**Multiobjective GA:**

The multiobjective GA was introduced, because it can handle integer parameters, which are the number of totally flexible demand points. If a demand point can be postponed to another day, its value is 0, if not, its value is 1. The other optimization parameter for the multiobjective GA is the result of the inner GA. The cost function therefore is a combination of the number of postponed demand points and total travel time.

**Pareto optimal result:**

As a result of the optimization a Pareto optimal front is built, as there are more optimal solutions. The Pareto parameters are the number of postponed demand points and the minimal total travel times. Because of using integer parameters more optimal solutions with the same values can be present. The solution, which occurs most often, was chosen as the final optimal scenario.

![Fig 29 Optimization with inner GA and multiobjective GA](image-url)
The extension caused some changes in the structure of the model of the original optimization algorithm (Fig 30).

A crucial change to the original algorithm in step 2 (Prioritization of activities) is that now all possible demand points are collected into the POI database in advance. Therefore the application of Overpass API during every request is not needed, as the alternative demand points can be searched from the offline database.

The calculation of travel times (step 3) for cars and public transport is performed in advance, as for running the genetic algorithm (GA) values for travel times are needed. In step 4 the basic scenario is calculated by running the extended optimization algorithm with fixed demand points. The obtained travel times are the basis of comparison with other scenarios, where flexible demand points are applied.

The extension of the optimization algorithm combines steps 5 and partially 6 (Solving the flexible TSP-TW and Optimization). It deals with the formulation of activity chains, with the selection of activity chains based on predefined criteria, with totally flexible demand points and with the optimization process.

In step 6 the Pareto optimal solutions are shown as an output and also the set of optimal demand points of the final optimal scenario and the related total travel time is presented. Also visualization of the optimal daily activity chain is realized using Google maps.
4.4 Evaluation of the simulation results with extensions

The simulations were performed on arbitrarily chosen test networks in Budapest using MatLab. In order to simulate a real daily activity chain, in the test network 7 activities were used with 5 flexible activities. A table of a daily activity plan (Table 17) was filled containing the type of the activity, the processing time, the priority, the opening and closing times of the demand point (which defines the time window), and the earliest start and latest end time (which defines the demanded time window of the traveler).

<table>
<thead>
<tr>
<th>Type of activity</th>
<th>processing time</th>
<th>priority</th>
<th>opening time</th>
<th>closing time</th>
<th>earliest start time</th>
<th>latest end time</th>
</tr>
</thead>
<tbody>
<tr>
<td>workplace</td>
<td>480 min</td>
<td>1</td>
<td>07:00</td>
<td>18:00</td>
<td>08:00</td>
<td>18:00</td>
</tr>
<tr>
<td>hairdresser</td>
<td>45 min</td>
<td>4</td>
<td>10:00</td>
<td>21:00</td>
<td>17:00</td>
<td>21:00</td>
</tr>
<tr>
<td>post office</td>
<td>15 min</td>
<td>2</td>
<td>8:00</td>
<td>18:00</td>
<td>16:00</td>
<td>21:00</td>
</tr>
<tr>
<td>ATM</td>
<td>5 min</td>
<td>2</td>
<td>0:00</td>
<td>23:59</td>
<td>7:00</td>
<td>18:00</td>
</tr>
<tr>
<td>gym</td>
<td>60 min</td>
<td>2</td>
<td>9:00</td>
<td>19:00</td>
<td>16:00</td>
<td>20:00</td>
</tr>
<tr>
<td>restaurant</td>
<td>30 min</td>
<td>3</td>
<td>10:00</td>
<td>23:00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pub</td>
<td>120 min</td>
<td>4</td>
<td>12:00</td>
<td>18:00</td>
<td>18:00</td>
<td>23:00</td>
</tr>
</tbody>
</table>

Table 17 Example daily activity chain input data

4.4.1 Results using car

In case of car it was assumed that the traveler uses own car during the whole day without significant traffic jams. In the basic scenario with fix schedule the total travel time is 1 hour 55 minutes, which decreases in the optimized scenario with 9 minutes (Table 18). This value highly depends on the chosen example, therefore it may result in much higher time savings.

In the flexible schedule the main changes of the daily schedule are noticeable in the afternoon. As the restaurant is temporally flexible, it could be scheduled before the other activities. The post office is spatially flexible, therefore a new location for this activity was found, which is closer to the hairdresser. Although the gym was flexible, it was close enough to previous activities, thus a change in its location was not necessary. The pub was reached earlier, and so home, too.

<table>
<thead>
<tr>
<th>fix schedule</th>
<th>end time</th>
<th>flexible schedule</th>
<th>end time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>start time</strong></td>
<td><strong>end time</strong></td>
<td><strong>start time</strong></td>
<td><strong>end time</strong></td>
</tr>
<tr>
<td>1. ATM</td>
<td>07:39</td>
<td>1. ATM</td>
<td>07:43</td>
</tr>
<tr>
<td>2. workplace</td>
<td>07:58</td>
<td>2. workplace</td>
<td>07:59</td>
</tr>
<tr>
<td>3. gym</td>
<td>16:14</td>
<td>3. restaurant</td>
<td>16:22</td>
</tr>
<tr>
<td>4. post office</td>
<td>17:30</td>
<td>4. post office</td>
<td>17:09</td>
</tr>
<tr>
<td>5. restaurant</td>
<td>17:57</td>
<td>5. hairdresser</td>
<td>17:31</td>
</tr>
<tr>
<td>7. pub</td>
<td>19:48</td>
<td>7. pub</td>
<td>19:44</td>
</tr>
<tr>
<td>8. home</td>
<td>21:56</td>
<td>8. home</td>
<td>21:52</td>
</tr>
<tr>
<td><strong>Total travel time</strong></td>
<td><strong>1 hour 58 min</strong></td>
<td><strong>Total travel time</strong></td>
<td><strong>1 hour 49 min</strong></td>
</tr>
</tbody>
</table>

Table 18 Example daily activity chain with fix and flexible demand points using car (data)
Concerning the visualization (Fig 31) of the daily activities (car travel coloured with magenta) it can be observed that more direct trips were performed and the locations of the activities are closer to each other.

The Pareto optimal solutions mean that a set of solutions are created, where the number of postponed demand points and total travel times are optimized. This provides more solutions for the user, which has to be displayed as possible alternatives, so that the user can decide, whether the travel time or the number of demand points is more important in the actual situation.

As a result of the optimization 4 Pareto optimal solutions are presented (Table 19). From the Pareto optimal solutions “solution 1” with the maximal number of demand points (where no activity is postponed) and a total travel time of 1 hour and 49 minutes was shown in the example. In case of “solution 2” and “solution 3” one activity is postponed (in the former case one, in the latter case the other totally flexible activity), therefore a total time reduction of 13 and 14 minutes is achievable. In case of “solution 4” two activities are postponed, which is 22 minutes less as in the first case.

<table>
<thead>
<tr>
<th>pareto optimal solutions</th>
<th>number of demand points</th>
<th>number of postponed demand points</th>
<th>total travel time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>0</td>
<td>1 hour 49 min</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>1</td>
<td>1 hour 36 min</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>1</td>
<td>1 hour 35 min</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>2</td>
<td>1 hour 27 min</td>
</tr>
</tbody>
</table>

Table 19 Pareto optimal solutions when using car

4.4.2 Results using public transport

In case of public transport usage the POI types, the set start and end times are summarized in Table 20. It is assumed that only public transportation and walking is used during the trips. Here the difference was even more significant, as ca. 11% decrease of total travel time was realized. On Fig 32 the routes of public transportation modes are shown.
The location and the order of the post office and the ATM have been changed. The algorithm found a post office closer to the workplace, and as earlier start time is enabled, it was put right after work. The location of the gym was not changed, but the activity was performed earlier. As the ATM was flexible, this activity was executed on the way to the hairdresser. The restaurant was flexible in time, but its location was fix, accordingly its starting time was slightly postponed to 19:03.

<table>
<thead>
<tr>
<th>fix schedule</th>
<th>flexible schedule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>start time</td>
</tr>
<tr>
<td>1. workplace</td>
<td>07:48</td>
</tr>
<tr>
<td>2. ATM</td>
<td>16:11</td>
</tr>
<tr>
<td>3. post office</td>
<td>16:31</td>
</tr>
<tr>
<td>4. restaurant</td>
<td>17:03</td>
</tr>
<tr>
<td>5. gym</td>
<td>18:06</td>
</tr>
<tr>
<td>7. pub</td>
<td>20:27</td>
</tr>
<tr>
<td>8. home</td>
<td>22:49</td>
</tr>
</tbody>
</table>

Table 20 Example daily activity chain with fix and flexible demand points using public transport (data)

Concerning transportation modes (Fig 32) after the optimization more walking was chosen, as the new demand points were close enough to each other. The colours represent the following transportation modes: walking (black), bus (blue), tram (orange), metro (green).

Fig 32 Example daily activity chain with fix and flexible demand points using public transport (visualization)

Considering the Pareto optimal solutions (Table 21) “solution 1” with the maximal number of demand points was presented in the detailed example. Postponing 1 demand point is resulted in “solution 2” and “solution 3”, with less total travel time (2 hour 14 minutes and 2 hour 12 minutes). Regarding the total travel time the minimum value is presented by “solution 4” with 1 hour 56 minutes, but in this case two demand points were postponed. The difference between “solution 1” and “solution 4” is 25 minutes.
Table 21 Pareto optimal solutions when using public transport

<table>
<thead>
<tr>
<th>pareto optimal solutions</th>
<th>number of demand points</th>
<th>number of postponed demand points</th>
<th>total travel time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>0</td>
<td>2 hour 21 min</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>1</td>
<td>2 hour 14 min</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>1</td>
<td>2 hour 12 min</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>2</td>
<td>1 hour 56 min</td>
</tr>
</tbody>
</table>

4.4.3 Results using combined mode

In case of combined modes it is possible to use public transportation and a car-sharing system available at any preferred location. As a result the order of visited demand points was quite similar (Table 22), as in case of car usage. This is because most of the trips were realized by car. In the fix schedule post office and restaurant are placed after work, and these activities are followed by gym. The remaining order is the same, as for cars. Therefore starting times and total travel times are also similar. The difference between fix and flexible schedule is 17 minutes, which is 14%.

Table 22 Example daily activity chain with fix and flexible demand points using combined modes (data)

<table>
<thead>
<tr>
<th></th>
<th>start time</th>
<th>end time</th>
<th>mode</th>
<th>start time</th>
<th>end time</th>
<th>mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>fix schedule</td>
<td></td>
<td></td>
<td></td>
<td>flexible schedule</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. ATM</td>
<td>07:44</td>
<td>07:49</td>
<td>car</td>
<td>1. ATM</td>
<td>07:41</td>
<td>07:46</td>
</tr>
<tr>
<td>2. workplace</td>
<td>07:59</td>
<td>16:00</td>
<td>car</td>
<td>2. workplace</td>
<td>07:50</td>
<td>16:00</td>
</tr>
<tr>
<td>3. post office</td>
<td>16:19</td>
<td>16:34</td>
<td>6</td>
<td>3. restaurant</td>
<td>16:22</td>
<td>16:52</td>
</tr>
<tr>
<td>4. restaurant</td>
<td>16:46</td>
<td>17:16</td>
<td>car</td>
<td>4. post office</td>
<td>17:11</td>
<td>17:26</td>
</tr>
<tr>
<td>5. gym</td>
<td>17:34</td>
<td>18:34</td>
<td>car</td>
<td>5. hairdresser</td>
<td>17:26</td>
<td>18:11</td>
</tr>
<tr>
<td>8. home</td>
<td>22:00</td>
<td>-</td>
<td>car</td>
<td>8. home</td>
<td>21:52</td>
<td>-</td>
</tr>
<tr>
<td>Total travel time</td>
<td>1 hour 59 min</td>
<td></td>
<td></td>
<td>Total travel time</td>
<td>1 hour 42 min</td>
<td></td>
</tr>
</tbody>
</table>

Concerning transportation modes (Fig 33) mainly car (magenta) was chosen, which is reasonable, as actual traffic situation and parking problems are not taken into account. In the basic scenario 1 tram usage (orange) and in the optimized case a walking (black) was inserted. In a more realistic situation probably more public transport usage could have been observed.
Fig 33 Example daily activity chain with fix and flexible demand points combined (visualization)

Looking at the Pareto optimal solutions the available alternatives can be observed in Table 23. The case of “solution 1” was presented in the example. The difference in total travel time is slightly smaller, than in the other scenarios. In case of “solution 2” and “solution 3” 6 and 7 minutes, while in case of “solution 4” only 15 minutes. The time difference is caused by postponing demand points. With the presented Pareto optimal solutions the user can decide, which solution fits best his or her needs. Thus if the user wants to arrive home early, he or she will choose the solution with the least travel time. But if it is more important for the user to finish all activities, then he or she will choose the solution, where not demand points are postponed.

<table>
<thead>
<tr>
<th>pareto optimal solutions</th>
<th>number of demand points</th>
<th>number of postponed demand points</th>
<th>total travel time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>0</td>
<td>1 hour 42 min</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>1</td>
<td>1 hour 36 min</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>1</td>
<td>1 hour 35 min</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>2</td>
<td>1 hour 27 min</td>
</tr>
</tbody>
</table>

Table 23 Pareto optimal solutions when using combined modes

4.4.4 Comparison

Comparing the simulation results (Table 24), combined modes provided the optimal choice with 1 hour 42 minutes total travel time, which was 7 minutes better, than car and 39 minutes better, than public transport. In relative terms the biggest reduction was observed in case of combined modes with 14,3%.

<table>
<thead>
<tr>
<th>transportation mode</th>
<th>total travel time with fix schedule</th>
<th>total travel time with flexible schedule</th>
<th>relative total travel time reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
<td>1 hour 58 min</td>
<td>1 hour 49 min</td>
<td>7,6%</td>
</tr>
<tr>
<td>public transport</td>
<td>2 hour 37 min</td>
<td>2 hour 21 min</td>
<td>10,2%</td>
</tr>
<tr>
<td>combined</td>
<td>1 hour 59 min</td>
<td>1 hour 42 min</td>
<td>14,3%</td>
</tr>
</tbody>
</table>

Table 24 Comparison of the simulation results
4.4.5 Calculation of processing times

Processing times of the extended optimization algorithm were analyzed for all transport modes in the case of only fix (basic scenario) and flexible demand points (optimal scenario). Comparing basic and optimal scenarios the optimization time is only doubled (while in the original case the difference was 6-10 x), which is very promising considering the high number of flexible points. As public transport needs more queries, than car, because of parallel alternative routes, its computing times are respectively higher (36,7 s and 28,5 s). However the difference disappears in case of flexible points (71,2 s and 69,3 s). In case of combined transport mode both computing times are higher (57,6 s and 111,6 s), as queries for car and public transport are also required.

Processing times of the original and extended optimization algorithms were also compared (Fig 35). The basic scenario (when all demand points are fixed) and the optimal scenario (when there are flexible demand points) were calculated with both algorithms. The difference is obvious, in general the extended algorithm over performs the original algorithm by 90% (resulting in 10% processing time).

In case of fix demand points the difference is smaller for public transport (original and extended algorithm: 96 s and 19 s) and combined transportation mode (147 s and 29 s). The performance of the extended algorithm is even worse for cars (original and extended algorithm: 7 s and 16 s). This is because the GA generates huge populations, which requires computing time, while in the original case only a few combinations were present. In case of flexible demand points for cars the difference is quite big (71 s and 23 s), but for the other transportation modes it is remarkable (556 s and 34 s; 861 s and 41 s).
4.5 Evaluation of activity chain optimization algorithm benefits

After the detailed discussion of simulated activity chains, real activity chains were analyzed to assess the benefits of the optimization algorithm. The benefits were measured in terms of the average travel time saving per person per day. The evaluation was run on a test dataset and then it was generalized to a small and a big dataset. The general statistics of the datasets contains the following parameters:

- number of users in the dataset, who participated in the survey,
- journeys of all users made during a day,
- average number of journeys per user during day,
- average duration of journeys during a day,
- areas of activities, which can be predefined districts or exact GPS coordinates,
- departure and arrival times, which are exact timestamps,
- transportation mode, which can also include more modes during a trip,
- trip purpose, which is usually grouped into 7-10 purposes containing work, school, shopping, food and entertainment,
- flexibility of the activity, which corresponds to the introduced flexibility definition in this chapter.

The small and the big dataset contains several individual activity chains and the relating socio-economic information of the users (e.g. age), which helps the classification of users into user groups, however it does not contain all required inputs for the optimization algorithm (e.g flexibility of the activities).

The test dataset contains all required inputs for the optimization algorithm including flexibility of the activities, therefore the evaluation could be run with exact values and produced accurate results.
4.5.1 Big dataset description

The dataset is based on the “Mobilität in Städten – SrV 2008” survey conducted by a researcher group of TU Dresden in Germany [3]. The aim of this survey was to determine users’ needs to analyze daily travels and to monitor changes of behavioral patterns.

Generally the survey is a household interview elaborated in German cities, which is repeated every 5 years. The dataset consists of approximately 110,000 records and it contains data of 41,000 persons from Berlin in 2008. The interviewees have been chosen to represent the demographic diversity of Berlin based on statistics reports. On each day, through 6 months, the interviewer randomly chose households and each person from each household had to report all of his or her trips on that day. They reported on their daily mobility pattern with departure and destination district, purpose of the trip, the vehicle they used and other attributes of the trip and about themselves.

These answers with household’s information and metadata were stored and organized in a database, which consists of 3 tables:

- The household table contains the location (city, district, street), basic information about households (number of persons, number of cars), and accessibility of public transportation modes (walking time to the next bus/tram/metro stop).
- The personal table stores personal information (age, gender, type of occupation, qualification, income) and other special attributes (visually handicapped, physically disabled).
- The travel table contains the starting location and time, purpose of the trip, arrival location and time, transportation mode and weather. During a trip more transportation modes may be used. The location information is represented by location codes, which are 195 statistical area codes of Berlin, which correspond to the borders of the districts and it reflects social and economic aspects.

In order to provide a basic pattern of travel behaviour, the trips were sorted by trip purpose and visualized (Fig 36). On the horizontal axis the hours of the day (hour) can be obtained, and the vertical axis shows the number of trips (frequency), which were performed during the indicated time interval. The different trip purposes are distinguished by colors. The number of trips build on each other, thus the value of one selected purpose graph is the summary of all graphs below.

The aim of this representation was to present not only the purposes separately, but the summary and ratio among the purposes. In the graph two peak hour periods can be detected, the first between 7 and 10 o’clock and the second between 15 and 18 o’clock. This corresponds to usual travel behavior. Most of the trips at early peak are the travels to the workplace, to the school, university or taking kids to the school/kindergarten. The frequency of these trips decreases as time passes. Late afternoon and at night the services and entertainment trips have a majority role. The least number of trips can be observed during late night hours.
Fig 36 Frequency of trips with particular purposes

The stored information in the database allows graphical representation of the trips between the districts of Berlin, and therefore enables the analysis of typical mobility patterns. A more detailed analysis and visualization of Berlin dataset can be obtained in Appendix 5.

4.5.2 Small dataset description

A similar survey was conducted in Hungary in 2014 by Center for Budapest Transport (BKK). The focus of the survey was on the analysis of mobility patterns and structure of daily activity chains.

In this case the survey was also based on a household interview with inhabitants of Budapest and its agglomeration, which is repeated every 7 years. The dataset contains data from 5000 households in Budapest and 2000 households in the agglomeration of Budapest. The size of the sample was set by the data of census of KSH (Central Statistics Office). The interviewees have been chosen randomly, and they had to report all trips of a specific day. The model contains 922 areas in Budapest, where the areas were defined based on the number of inhabitants. During the interview information about the departure and destination area, travel times, purpose of the trip, and used vehicles were asked.

4.5.3 Test dataset description

A survey was performed in Budapest in 2015 with university students. The aim of the survey was to collect data about daily activity chains with all required information, which is needed by the optimization algorithm.

The dataset contains data from about 93 users and 315 activity chains. The survey was designed to receive as exact data about mobility patterns, as possible, thus exact GPS coordinates of the activities were asked. During the interview information about the departure and destination, travel times, purpose of the trip, used vehicles and flexibility of the activity was also collected.
4.5.4 Evaluation of the optimization benefits

Daily activity chains of travelers were obtained and evaluated. As in the test dataset the students (test data users) mainly used public transport, only the transportation mode of the optimization algorithm was analyzed.

From the general statistics (Table 25) it was observed that test data users are traveling by ca. 30% more (average journeys per user: 3,39), than big and small data users (average journeys per user: 2,69 and 2,55). The difference is caused by the more active lifestyle of younger people, as in the test dataset students were surveyed, while in the big and small dataset representative samples of all ages were considered. It has a positive effect on the efficiency of the optimization algorithm, because more activities mean more possibility of applying flexible demand points and thus travel time decrease.

Considering average duration of journeys, the value of the test dataset corresponds to the big and small dataset. This means that test data users are processing many activities during a day within rather small distances. It has a negative effect on the efficiency of the optimization algorithm, because if many activities are close to each other, it is harder to find such flexible demand points, which cause significant travel time saving.

<table>
<thead>
<tr>
<th>name</th>
<th>big dataset</th>
<th>small dataset</th>
<th>test dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of users</td>
<td>41 050</td>
<td>8573</td>
<td>93</td>
</tr>
<tr>
<td>Journeys of all users</td>
<td>111 228</td>
<td>21 844</td>
<td>315</td>
</tr>
<tr>
<td>Average journeys per user</td>
<td>2,69</td>
<td>2,55</td>
<td>3,39</td>
</tr>
<tr>
<td>Average duration of journeys</td>
<td>24,9 min</td>
<td>37,5 min</td>
<td>29,24 min</td>
</tr>
<tr>
<td>Areas of activities</td>
<td>195</td>
<td>922</td>
<td>GPS coord.</td>
</tr>
<tr>
<td>Departure and arrival time</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Transportation mode</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Trip purpose</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Flexibility of the activity</td>
<td>-</td>
<td>-</td>
<td>x</td>
</tr>
</tbody>
</table>

Table 25 General statistics of the datasets

When running the optimization algorithm on the test dataset, only those activity chains were taken into account, which consist of at least 3 activities and 1 flexible activity, and where the travel time to the flexible activity was at least 10 minutes. These constraints were applied, because only in the selected cases provides the optimization of activity chains feasible results.

From the 93 activity chains performed by the users in the test dataset 29 activity chains corresponded to the constraints. The average travel time saving for the selected activity chains (where an optimization was feasible) was 21,6%. Extrapolating the results to all activity chains of the test dataset the average travel time saving drops to 6,7%.

Generalizing the results to the big and small dataset it was calculated that an average travel time saving of 4,5 and 6,4 minutes per person per day (considering the average duration of journeys and average journeys per user) can be achieved by the application of the optimization method for activity planning.
The results provide only an approximation of exact travel time savings, which depend on the travel habits of the user, on the chosen day, on the season, on the possible routes and transportation network opportunities, on the specificities of the city and several other factors. However, it presents a quantitative result of the benefits of the optimization algorithm in general, which helps understanding influence opportunities of travel behaviour and facilitates the planning of activity chains.

4.5.5 Future development directions

Using mobility analysis on individual level, daily activity chains of users can be obtained and the patterns of their movements can be determined. The predictable routes of users can be beneficial for transport operators, as they can create such timetable schedules of buses and trams that represent almost real travel demand of the passengers.

The theoretical model and simulation results have to be followed by the elaboration of validation scenarios using big amount of real passenger observation data. Understanding and describing mobility patterns enables the simulation of the city life that can be useful for urban planners and decision makers.

In the model it was assumed that the passenger’s start and end point is the same. In the most cases it is valid, because usually the daily activity chains start from home and after finishing all activities the destination is home again. But in some cases the destination can be different, which requires the application of another TSP method.

The travel times are now fix values, thus among each point during every time of the day the optimization algorithm calculates with the same travel times. Considering the actual traffic situation, the elements of the travel matrix could be changed, even predicted. Also the calculation of processing times could be extended by defining an average TP value, which can be modified to a minimal TP value, if a delay occurs. The minimal TP value would be the minimum time, which has to be spent at the given demand point.

Introducing predictive TSP the latent demands of the passengers could be served. These latent demands of the passengers could be derived from the requested demands of passengers with similar characteristics.

To enhance the dynamics and actuality of the model, changes in the activity plan during the day could be taken into consideration, as the appearance of a new demand (respectively a new demand point) or the occurrence of a delay at a demand point. Thus the daily activity plan has to be re-planned and recalculated according to the new situation.
5  Modeling the information system of activity chain optimization

In order to reveal wider context of optimization, my aim was to define all possible optimization parameters and highlight the connections between them. I have created the operational model of the related information system. Accordingly, the user requirements are taken into account and the personalized settings are realized.

Recent literature in field of activity chain optimization is related to subtopics like activity-based trip analysis, mode choice modeling, travel demand management and exploration flexible mobility options. Islam and Habib [93] classified the literature dealing with activity-based travel modeling into the following groups: definition of activity patterns, investigation of the effect of socio-demographic characteristics and prediction of activity chains.


Many aspects of activity chains have been studied, as the comparison of travel behaviour of men and women by McGuckin and Murakami [121], the study of activity chains in specific regions by Subbarao and Krishna Rao [157], also its assessment by age groups was presented by Golob and Hensher [82]. Some other aspects have been also studied, as the economics of activity-travel decisions by Lawson [107], or the comparison of local spatial configurations of land use and transport to assess their impact on lifestyle and travel decisions by van Eck et al. [168].

Lawton [108] claims that there are four possible sources of information to build up or feed activity-based models, which may be used to identify parameters of an optimization model: household surveys (revealed preference) to study activities that influence travel demand, stated response survey to investigate activity-travel patterns, longitudinal panel surveys and retrospective surveys of activities to explore long term behaviour (e.g. household location decisions). The combination of these methods with ICT-supported new sources of information helps to identify personal parameters to establish a proper activity chain optimization model. Recent papers on traveler profile definition by Pronello and Camusso [137], on mobility styles and travel behaviour by Prillwitz and Barr [136], on target group segmentation for sustainable transport by Haustein and Hunecke [87], and on a study of innovative methods in activity-travel surveys by Auld et al. [9] paved the way for further research in this topic.

An important aspect of the UTRACS survey was the attempt to capture activity-travel planning attributes. The planning attributes were focused on timing and constraints of planning decisions. It was explored that user decision about transportation mode are mainly driven by routine, while the choice of start time of activities are more individual and impulsive.
When modeling the operation of the information system I have identified the classification and the optimization parameters, introduced the utility function using the defined parameters and the aggregated weights, as well as modeled the system architecture (Fig 37).

To represent detailed user requirements, a set of parameters were introduced. They were grouped into classification parameters (to classify users into user groups) and optimization parameters (to provide utilities to the optimization algorithm). Aggregated weights were assigned to the optimization parameters, which represent the individual preferences of the users. The aggregated weights are initially defined as average weights of the certain user group ($w_{\text{init}}$), but can be updated by personalized settings ($w_{\text{pers}}$). The utility function ($u(p,w)$) provides input for the optimization algorithm about preferences of the user.

The optimization algorithm has been included into the system architecture. The system architecture consists of the user interface, the server applications and external databases. The optimization parameters are provided through the external databases.

The classification parameters and the aggregated weights are determined by the user. The result of the optimization algorithm, namely the optimal activity chains, are presented to the users.

Fig 37 Operational model of the information system of the activity chain optimization
5.1 Definition of classification and optimization parameters

In order to model the complex requirements of users regarding an urban activity chain, the possible optimization parameters were identified. In the literature the following main optimization parameters are typical: time, cost and comfort. The quantification of the comfort parameters is not elaborated very detailed, therefore I defined, described and classified the most important comfort parameters. The parameter type, component type, possible values and data source were created for grouping the parameters.

**Parameter type:**

Two types of parameters have been introduced. The parameter type describes whether the parameter is a classification parameter or an optimization parameter. The detailed descriptions of the parameters follow the order of the parameter types, which are actually strongly linked to the component type.

- The **classification parameters** are not used directly in the optimization process, but are crucial input for the classification of users into user groups. The creation of user groups facilitates user decisions about setting the weights for the optimization parameters. The user groups possess predefined settings of the weights, which weights provide only an initial setting, and the values can be changed by personal preferences.

- The **optimization parameters** are used in the optimization process. Their two main groups are: the general optimization parameters without weights (with exception of time and cost), and the comfort optimization parameters with weights. Usually general optimization parameters are parameters with fix or predefined values, where weighting cannot be defined (e.g. opening times). Parameters present directly in the utility function are *italics*.

![Fig 38 Grouping of parameters by parameter type](image)
Component type:

Three types of component have been identified, the user, the trip and the location (Fig 39). Most parameters belong clearly to one component type, but some parameters influence more component types, therefore they were placed in the intersections.

- The **user** includes classification and optimization parameters, which are depending from the individual user.
- The **trip** contains optimization parameters and is divided into sub-types according to the transportation modes, as transportation modes have specific parameters.
- The **location** consists of those optimization parameters, which are connected to the location of the activity.

![Component type diagram](image)

**Fig 39 Identification of parameters by component type**

Possible values:

In case of optimization parameters with numeric values, the quantification and creation of categories is easy, as exact values can be assigned to the categories (e.g. prices). The quantification is also possible for optimization parameters with textual value sets by assigning artificially created value categories. In some cases, the optimization parameters can only be categorized by applying heuristic considerations or the exact values of the categories can be learned by collection of a large number of examples (e.g. crowding). In case of optimization parameters with weights, the parameter have the following possible values: low, medium, high. Low values represent “good” features, while high values represent “bad” features.

Data source:

The data source refers to the origin of the parameters, which can originate from the user (by setting the requested values), from the application (by collecting and evaluating usage statistics) or from external sources (by receiving data or datasets). The external sources can be represented by a transport operator, a municipality, a social media provider, a POI database or other databases.
5.1.1 Classification parameters

In the following chapters the parameters are grouped by the parameter type. The comfort optimization parameters are further divided by the component type.

The classification parameters and their attributes have been identified (Table 26). They are mainly connected to the user component type.

- **age, gender, occupation, income, car ownership, family status**: the basic socio-economic data, which are required to categorize users into user groups,
- **number of daily trips**: average number of trips during a day (e.g. users with family tend to have more daily trips, while pensioners probably have less daily trips),
- **flexibility**: average number of flexible activities during a day (e.g. users with flexible working hours and students tend to have more flexible activities),
- **number of changes**: average number of changes in daily activity plans (e.g. younger people tend to change their mind and have new unplanned events during the day).

<table>
<thead>
<tr>
<th>parameter type</th>
<th>component type</th>
<th>values</th>
<th>data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>classification</td>
<td>user</td>
<td>1 (0-20)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 (21-30)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 (31-40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 (40-60)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 (more than 60)</td>
</tr>
<tr>
<td>gender</td>
<td>classification</td>
<td>user</td>
<td>1 (male)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 (female)</td>
</tr>
<tr>
<td>occupation</td>
<td>classification</td>
<td>user</td>
<td>1 (student)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 (worker)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 (pensioner)</td>
</tr>
<tr>
<td>income</td>
<td>classification</td>
<td>user</td>
<td>1 (0-500)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 (500-1000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 (1000-2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 (2000-3000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 (more than 3000 Euro)</td>
</tr>
<tr>
<td>car ownership</td>
<td>classification</td>
<td>user / trip</td>
<td>1 (no car)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 (one car)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 (more cars)</td>
</tr>
<tr>
<td>family status</td>
<td>classification</td>
<td>user / trip / location</td>
<td>1 (single)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 (married)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 (has children)</td>
</tr>
<tr>
<td>number of daily trips</td>
<td>classification</td>
<td>user</td>
<td>1 (0-1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 (2-3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 (more than 3)</td>
</tr>
<tr>
<td>flexibility</td>
<td>classification</td>
<td>user</td>
<td>1 (0-1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 (2-3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 (more than 3)</td>
</tr>
<tr>
<td>number of changes</td>
<td>classification</td>
<td>user</td>
<td>1 (0-1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 (2-3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 (more than 3)</td>
</tr>
</tbody>
</table>

Table 26 Description of classification parameters
5.1.2 General optimization parameters

The general optimization parameters were identified. Most of these parameters are without weights, with the exception of time and cost. These parameters are mainly connected both to the user and to the location component type (Table 27).

- *time (p₁):* total travel time spent during a trip (e.g., driving time is calculated from the distance, public transport time is calculated from the timetable information),
- *cost (p₂):* price to be paid for the chosen trip for different transportation modes (e.g., driving cost is calculated from the distance, public transport cost is calculated from the number of needed tickets),
- *distance (p₃):* defines locations and distances, so that trip times can be calculated,
- *priority (p₄):* the priority of the actual activity that could be processed in the given location (e.g., fix or flexible),
- *opening hours (p₅):* time interval between the opening time and closing time of the given location,
- *processing time (p₆):* time spent by the user at the given location [minutes],
- *POI type (p₇):* the type of the location regarding the Point Of Interest typology (e.g., restaurant, post office),
- *specificity (p₈):* defines, whether one brand is searched by the user or more brands of the same POI type are possible solutions.

<table>
<thead>
<tr>
<th>Parameter Type</th>
<th>Component Type</th>
<th>Values</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>time</strong></td>
<td>with weight</td>
<td>trip</td>
<td>0-1000 minutes</td>
</tr>
<tr>
<td><strong>cost</strong></td>
<td>with weight</td>
<td>trip</td>
<td>0-1000 minutes</td>
</tr>
<tr>
<td><strong>distance</strong></td>
<td>without weight</td>
<td>trip</td>
<td>0-10,000 m</td>
</tr>
<tr>
<td><strong>priority</strong></td>
<td>without weight</td>
<td>user / location</td>
<td>1 (fix) 2 (temporally flexible) 3 (spatially flexible) 4 (totally flexible)</td>
</tr>
<tr>
<td><strong>opening hours</strong></td>
<td>without weight</td>
<td>location</td>
<td>0-24 hour</td>
</tr>
<tr>
<td><strong>processing time</strong></td>
<td>without weight</td>
<td>user / location</td>
<td>0-1000 minutes</td>
</tr>
<tr>
<td><strong>POI type</strong></td>
<td>without weight</td>
<td>user / location</td>
<td>several types (e.g., restaurant, post office)</td>
</tr>
<tr>
<td><strong>specificity</strong></td>
<td>without weight</td>
<td>user / location</td>
<td>special brands under POI type</td>
</tr>
</tbody>
</table>

Table 27 Description of general optimization parameters
5.1.3 Comfort optimization parameters

The comfort optimization parameters were described connected to the trip component type (Table 28).

- **weather** ($p_9$): measure for the actual daily average weather situation measured by the temperature and the humidity (e.g. rainy, windy),
- **eco-friendliness** ($p_{10}$): measure for environmental classification of different transportation modes measured by their average CO2 emission (e.g. biking is very eco-friendly, as it produces no emission),
- **transport comfort** ($p_{11}$): measure for the comfort level generally provided by transportation modes measured by social norms (e.g. cars are considered to be comfortable),

**bike:**

- **biking routes** ($p_{12}$): type of roads used by bikers, if during a trip more types are present, then their average value regarding the length has to be taken into account,

**car:**

- **road quality** ($p_{13}$): measure for the quality of roads,
- **traffic tolls** ($p_{14}$): average price to be paid on certain roads,
- **congestion** ($p_{15}$): average percentage of congestion on a certain trip during a day,
- **incidents index** ($p_{16}$): number of yearly incidents along a chosen trip per average hourly traffic volume,

**PT:**

- **number of transfers** ($p_{17}$): number of transfer between two consecutive activities,
- **vehicle modernity** ($p_{18}$): age of the used vehicle during a trip,
- **crowding** ($p_{19}$): percentage of passengers per capacity in the vehicle,

**walking:**

- **pavement quality** ($p_{20}$): measure for the quality of pavements,
- **street type** ($p_{21}$): type of streets (e.g. main road, park, alley),
- **slope and stairs** ($p_{22}$): measure for the steepness of the streets and existence of stairs during the trip (e.g. locations in hilly areas are harder to access for elderly people).
Finally comfort optimization parameters were identified connected to the location component type (Table 29).

- **rating** ($p_{23}$): average value calculated from reviews of the location by other users,
- **price range** ($p_{24}$): average prices of typical services in the given location,
- **city area** ($p_{25}$): type of the far environment (e.g. district), where the place is located,
- **location area** ($p_{26}$): type of the close environment (e.g. main road, park, alley),
- **security** ($p_{27}$): number of yearly crimes in the city area, where the place is located,
- **accessibility** ($p_{28}$): measure for the easiness to access the location (with wheelchair),
- **parking fees** ($p_{29}$): average price of parking fees in the area of the location,
- **parking space** ($p_{30}$): number of free parking spaces in the area of the location.

<table>
<thead>
<tr>
<th>Table 28 Description of comfort optimization parameters of trip component type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>parameter type</strong></td>
</tr>
<tr>
<td><strong>weather</strong></td>
</tr>
<tr>
<td><strong>eco-friendliness</strong></td>
</tr>
<tr>
<td><strong>transport comfort</strong></td>
</tr>
<tr>
<td><strong>biking routes</strong></td>
</tr>
<tr>
<td><strong>road quality</strong></td>
</tr>
<tr>
<td><strong>traffic tolls</strong></td>
</tr>
<tr>
<td><strong>congestion</strong></td>
</tr>
<tr>
<td><strong>incident index</strong></td>
</tr>
<tr>
<td><strong>number of transfers</strong></td>
</tr>
<tr>
<td><strong>vehicle modernity</strong></td>
</tr>
<tr>
<td><strong>crowding</strong></td>
</tr>
<tr>
<td><strong>pavement quality</strong></td>
</tr>
<tr>
<td><strong>street type</strong></td>
</tr>
<tr>
<td><strong>slope and stairs</strong></td>
</tr>
</tbody>
</table>
5.2 Elaboration of the utility function

Utility functions have been introduced in order to combine the values of the optimization parameters and to measure the “goodness” of a chosen activity chain. The utility functions consist of optimization parameters and weights. Weights related to comfort optimization parameters are aggregated weights.

5.2.1 Aggregated weights

The aggregated weights were introduced to decrease the number of required settings by the users. The aggregated weights influence the relevance of more optimization parameters, thus the modeling of typical user requirements is present. The possible values of the aggregated weights can be between 1 and 5. These values are predefined by the user groups (average values), but can be changed by the user (Fig 40).

The utility functions \( u(p,w) \) regarding comfort optimization parameters were formalized creating the mathematical context of dependencies between optimization parameters and aggregated weights.

The following aggregated weights were defined:

- routine \( (w_r) \): measure of willingness to differ from well-known routes, this weight has a general effect on several parameters (e.g. willingness to make detours, if it is...
beneficial), it is a super aggregation with an effect on delay sensitivity, lifestyle, quality sensitivity, price sensitivity and area sensitivity,

- **delay sensitivity** ($w_1$): average delay tolerated by the user, which depends on the congestion and the incident probability of the chosen trip (e.g. users with high delay sensitivity should avoid congested routes),

  $$u_1(p, w) = (p_{15} * w_1 + p_{16} * w_1) * w_r$$  \hspace*{1cm} (30)

- **lifestyle** ($w_2$): measure for environmental consciousness and security features (e.g. rather using more eco-friendly transportation modes and avoiding dangerous areas),

  $$u_2(p, w) = (p_{10} * w_2 + p_{27} * w_2) * w_r$$  \hspace*{1cm} (31)

- **quality sensitivity** ($w_3$): measure for taking comfort features, price ranges and parking space into account (e.g. businessmen tend to use cars and visit places with higher prices),

  $$u_3(p, w) = (p_{11} * w_3 + p_{30} * w_3 + p_{24} * w_3) * w_r$$  \hspace*{1cm} (32)

- **price sensitivity** ($w_4$): willingness to pay for a certain trip, which includes traffic tolls and parking fees (e.g. workers may travel longer distances, where no traffic toll has to be paid),

  $$u_4(p, w) = (p_{14} * w_4 + p_{29} * w_4) * w_r$$  \hspace*{1cm} (33)

- **area sensitivity** ($w_5$): measure for taking features regarding ratings and the area of the location into account, as the city area and location area (e.g. users tend to visit restaurants in the city center, but a recreational activity rather close to a park),

  $$u_5(p, w) = (p_{26} * w_5 + p_{23} * w_5 + p_{25} * w_5) * w_r$$  \hspace*{1cm} (34)

- **biking preference** ($w_6$): measure of the willingness of using a bike during trips (e.g. students tend to bike more often), this weight has a general effect on biking related parameters,

- **biking habits** ($w_7$): requirements of the users regarding road quality, biking routes and weather (e.g. many users prefer built road and good weather),

  $$u_6(p, w) = (p_{12} * w_6 + p_{13} * w_6 + p_{9} * w_6) * w_b$$  \hspace*{1cm} (35)

- **car preference** ($w_8$): measure of the willingness of using a car during trips (e.g. businessman tend to use their own cars more often), this weight has a general effect on car related parameters,

- **car habits** ($w_9$): requirements of the users regarding road quality and weather (e.g. certain users do not use their cars in winter),
\[ u_7(p, w) = (p_{13} \times w_7 + p_9 \times w_7) \times w_c \] (36)

- **PT preference \( (w_7) \):** measure of the willingness of using PT during trips (e.g. younger people prefer public transportation, because they can utilize their time more efficiently with reading on the vehicles), this weight has a general effect on PT related parameters,

- **PT habits \( (w_8) \):** requirements of the users regarding number of transfers, crowding and vehicle types including cleanliness, comfortable seats, heating and air conditioning (e.g. users do not prefer old vehicles without air conditioning during the summer),

\[ u_8(p, w) = (p_{17} \times w_8 + p_{19} \times w_8 + p_{18} \times w_8) \times w_p \] (37)

- **Walking preference \( (w_8) \):** measure of the willingness of walking during trips (e.g. young people tend to walk more), this weight has a general effect on walking related parameters,

- **Walking habits \( (w_9) \):** requirements of users regarding pavement quality, street type and weather (e.g. certain users prefer nice road with trees and good weather).

\[ u_9(p, w) = (p_9 \times w_9 + p_{21} \times w_9 + p_{20} \times w_9) \times w_w \] (38)

- **Special needs \( (w_{10}) \):** the need for special services, as modern low floor vehicles, need of avoidance of stairs or slopes and accessibility of locations (e.g. users with wheelchairs do not like to visit places without ramps),

\[ u_{10}(p, w) = p_{22} \times w_{10} + p_{18} \times w_{10} + p_{28} \times w_{10} \] (39)

---

**Fig 40 Aggregated weights and related optimization parameters**
5.2.2 Utility function

The main utility function is defined as the sum of products of optimization parameters and weights (Fig 41). The optimization parameters are weighted, where weights represents the personal preferences of the users. In case of comfort optimization parameters the weights were grouped into aggregated weights, so that users can express their requirements. The optimization parameters are values retrieved from external data sources, whereas weights and aggregated weights are set by the user.

The value of time, the cost and the value of comfort are different among user groups. During the optimization the utility function is minimized. The minimization of time and cost is a well-known operation. In case of comfort parameters, the possible values were defined in such a way that low values represent ideal conditions and high values represent not preferred conditions.

\[
\min u(p, w) = p_{\text{time}} \times w_{\text{time}} + p_{\text{cost}} \times w_{\text{cost}} + \sum_{i=1}^{m} u_i(p, w)
\]

- \( p \) – optimization parameters,
- \( w \) – weights (including aggregated weights),
- \( u(p, w) \) – main utility function,
- \( p_{\text{time}} \) – value of time optimization parameter,
- \( w_{\text{time}} \) – weight of time optimization parameter,
- \( p_{\text{cost}} \) – value of cost optimization parameter,
- \( w_{\text{cost}} \) – weight of cost optimization parameter,
- \( u_i(p, w) \) – utility function regarding comfort optimization parameters, \( i = 1, \ldots, m \),
- \( m \) – number of utility functions regarding comfort optimization parameters.

Fig 41 Elements of the utility function
5.3 Modeling the system architecture

After the definition of the parameters and utility functions, the system architecture was modeled to provide a structure for the implementation of the elaborated optimization module (Fig 42). When planning online services, the three-tier system architecture [177] is most commonly used, where the main elements are:

- user interface,
- database server (part of server applications),
- application server (part of server applications),
- external databases.

The previously defined classification and optimization parameters and aggregated weights are included in the optimization module. The aim was to model the environment and connections of the optimization module. The idea behind this type of system architecture is the modularity with well-defined interfaces and the possibility of separate development of the elements. The modeled system architecture is the prerequisite and guideline for the implementation of a real activity chain optimization service.

Fig 42 Model of the system architecture
5.3.1 User interface

The user interface typically runs on a desktop or mobile phone as a web application or a mobile application. In our case the user interface is a touch optimized application for mobile devices. The mobile application communicates with the server applications through an online data connection provided by any type of mobile internet. The most data exchange is performed during planning of activity chains, however the connection would also be established during the update of traffic information or re-planning request.

The application provides an independent service, freely accessible to the users. The registration process requires basic socio-demographical data (user data) and preferences regarding location types and transportation modes (weights).

The user can import different types of calendars, which will serve as the basis of the planned daily activities. The basic information about the activities has to be manually completed with some information (e.g. priority, earliest start time and latest start time). New activities can be added by entering the data manually, by choosing from a POI list or by searching on a map (e.g. OSM or Google map).

The mobile application sends all data of the planned activities to the server, and is not calculating the request by itself. The generated activity chain is sent back by the server, and the mobile application is responsible for appropriate visualization of the data.

The overview of daily activity chain and details of particular trips are also available. Activities can be saved in the favourites, therefore regular activities can be easily added to the activity chain.

Alerts from transport operators are sent by the server, which may cause deviations in the planned activity chain. In this case a re-planning request can be sent by the user. Then the server calculates a new activity chain considering the changed traffic situation.

5.3.2 Server applications

The main functionalities of the background system are implemented in the server applications, where the application server is responsible for running the applications and the process logic and the database server contains data storage logic. The structure and connections of the server applications has been modeled.

**Database server:**

The function of the database server is the storage and update of data. Various types of data are stored in an SQL based database (central database) with access through predefined authentication levels. The data are sent event-driven, requested directly from the user or from the application server.

The individual user data, activities, planned activity chains and GPS logs require the most storage space in the central database, which can be only queried by the authenticated user or the system administrator. In order to keep the size of the database on a predefined level, not usable very detailed daily data are regularly deleted (e.g. GPS logs of the users). However aggregated data (calculated from the detailed daily data) are permanently stored.
The information about demand points (e.g. GPS data, opening time, closing time, description) are also stored in the central database, and are systematically updated from the external databases. The information about the transportation modes and locations is also contained by the central database.

**Application server:**

The application server is responsible for the data exchange between the database server, the mobile application and the external databases, which includes data processing and sending answer messages (services).

- The user service handles new registrations, log in processes and updates of user data initiated by the user.
- The traffic service provides actual traffic information for all transportation modes from different service providers (e.g. BKK info, Google maps).
- Weather information is collected through the weather service from meteorological providers.
- The public transport service handles not only the timetables of public transport vehicles (e.g. GTFS), but also more general data, as ages of the vehicles and crowding on the vehicles.
- The road network service provides information about road parameters, as traffic tolls and road quality.
- The POI service reads POI info of locations from external providers (e.g. OSM).
- The social media service requests data collected through social media providers, as ratings and price ranges (e.g. Foursquare).
- The statistics service requests statistical data about accidents and security, which are stored in the central database (e.g. national statistics office).
- The urban environment service handles data about streets and areas, which are usually provided by the municipality.
- The map service provides maps for the visualization (e.g. Google maps).

The optimization module processes activity planning requests, runs the optimization based on the elaborated algorithms and provides the optimized activity chains. The application server forwards the set of activities to the users and also the routes with the chosen transportation mode.

The planning of activity chains is executed by the application server individually for each user, however implementing a global optimum for all users is possible by the further extension of the optimization module. In case of an optimum for more users all requests have to be handled parallel and the optimization criteria will be changed from individual criteria to minimize the sum of all applied criteria regarding the requests.

5.3.3 External databases

The external databases represent storage tools for several data sources provided by external operators. Only those external databases were collected, which are required for the functioning of the background system. The data transfer is realized through standard interfaces, which are implemented in the particular services. Data from most
external systems are updated occasionally (e.g. road network, public transport, POI info, social media, statistics, urban environment information and maps), only a few data need to be updated regularly (e.g. weather, timetable), and other data have to be updated repeatedly during the day (e.g. traffic situation). The suggested update frequencies for the services are presented in Table 30, which also includes the possible data source in case of Budapest.

<table>
<thead>
<tr>
<th>external database</th>
<th>included parameter types</th>
<th>service</th>
<th>data source</th>
<th>update frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>traffic situation</td>
<td>congestion, parking space</td>
<td>Traffic service</td>
<td>BKKinfo, Google maps [traffic information]</td>
<td>quarterly, hourly</td>
</tr>
<tr>
<td>weather</td>
<td>weather</td>
<td>Weather service</td>
<td>Hungarian Meteorological Service – OMSZ (meteorology)</td>
<td>daily</td>
</tr>
<tr>
<td>public transport</td>
<td>timetable, number of transfers, vehicle</td>
<td>Public transport service</td>
<td>Centre for Budapest Transport – BKK (transport operator)</td>
<td>monthly (daily)</td>
</tr>
<tr>
<td></td>
<td>modernity, crowding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>road network</td>
<td>traffic tolls, road quality, parking fees</td>
<td>Road network service</td>
<td>Budapest Közút (road network operator)</td>
<td>monthly</td>
</tr>
<tr>
<td>POI info</td>
<td>opening hours, POI type, specificity,</td>
<td>POI service</td>
<td>Open Street Map – OSM (POI database)</td>
<td>monthly</td>
</tr>
<tr>
<td></td>
<td>accessibility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>social media</td>
<td>rating, price range</td>
<td>Social media service</td>
<td>Foursquare (social media)</td>
<td>monthly</td>
</tr>
<tr>
<td>statistics</td>
<td>incident index, security</td>
<td>Statistics service</td>
<td>Central Statistics Office – KSH (national office)</td>
<td>yearly</td>
</tr>
<tr>
<td>urban environment info</td>
<td>biking routes, street type, slope, city area,</td>
<td>Urban environment service</td>
<td>Municipality of Budapest (municipality)</td>
<td>yearly</td>
</tr>
<tr>
<td></td>
<td>location area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maps</td>
<td>distance</td>
<td>map service</td>
<td>Google maps (map provider)</td>
<td>yearly</td>
</tr>
</tbody>
</table>

Table 30 Update frequencies of external databases

The received data are strongly dependent on the organizational structure of operators of the external databases. In case of a reorganization of an operator the provided data structure and interfaces may also change. Therefore a regular update about the status of the operators is required.

5.3.4 Model of the menu structure of the mobile application

The menu structure of the mobile application was modeled, where the most important aspect is that it has to be kept as simple and intuitive as possible (Fig 43). After logging in it consists of 3 main menus: settings, input of activities and presentation of activity chains. On the top of all screens the most important menu items appear (e.g. schedule, map overview and activities).
Fig 43 Model of the menu structure of the mobile application

**Settings:**
- **User data:** basic information about the user (e.g. username, birth date, e-mail, password),
- **Personal preferences:** the values of the aggregated weights can be set by the user.

**Input of activities:**
- **History of activities:** list of previously searched activities,
- **Favourite activities:** list of saved activities, usually most typical activities,
- **New activities:** input of a new activity manually or from calendar,
- **Modify activities:** modification of attributes of existing activities (e.g. processing time, earliest start time, priority),
- **Activity planning:** list of chosen activities, possibility to send planning request.

**Presentation of activity chains:**
- **Schedule:** tabular view of the planned activity chain (e.g. name of activity, start time, end time, transportation mode, travel time),
- **Map overview:** map view of the planned activity chain (the activities and trips are clickable),
- **Activity properties:** detailed properties of selected activities (when clicking on activities in the schedule or map view), possibility to modify attributes of the chosen activity,
- **Trip properties:** detailed properties of selected routes (when clicking on transportation mode in the schedule view or on routes in the map view), possibility to view journey times, transfer locations, waiting times, etc.,
- **Alerts:** personalized text alerts regarding the planned activity chain, possibility to ask for a re-planning of the activity chain (e.g. traffic jam or disruptions at the locations of activities).
5.4 Development directions of information services

The progress of information services were also considered in order to outline future development directions regarding activity chain optimization (Fig 44). I identified the following main development directions:

- data collection (reliable real-time data and crowd sourcing data),
- data storage (standardization and functional integration),
- value-added services (personal preferences, multimodality, location based services and premium information).

5.4.1 Collection

One of the most debated issues is the reliability of the information. Real-time data is often originated from the operational control system of a transport operator (e.g. Deutsche Bahn) or from crowd-sourcing applications (e.g. Waze). Data collection for cars or passengers can be realized by FCD (Floating Car Data), by mobile cell or BlueTooth device information [159]. In order to produce reliable and real-time data for public transport vehicles [110], they should be equipped with a GPS receiver and telecommunication device (e.g. GPRS), moreover all the collected data can be integrated in a central database. In many countries some projects aim to establish such an infrastructure, but the realization requires in most cases rather long time.

A promising solution for collecting large amount of reliable data is crowd sourcing, where the source of the information is the user itself [1]. In order to filter incorrect information advanced algorithms are applied (e.g. only those messages are considered, where the location data of the user were also sent). Furthermore other users can rate the reliability of the provided disturbance information. The actuality of the information is also an essential aspect. Time delay between the occurrence of the incident and the appearance at the end users should be minimal. If this delay is too large, the information may be outdated.
In general data collection services aim to request less information handling operations directly from the user, but rather receiving information by learning habits and choices of the user.

5.4.2 Storage

The precondition for proper operation of an advanced system is the acquisition of all needed data (for route planning and information provision during the trip) through an integrated and updated database [111]. A European wide platform is not realized yet, but several data models exist for data storage (e.g. Transmodel [27]) and for data transfer (e.g. DATEX [46]). The realization is not only in the interest of passengers, but also of service providers, because they would be able to track trips of each registered passenger, store the result of their queries and create statistics. On the basis of the collected data typical travel behaviours and passenger demands can be determined. With that information the lines and timetables could be optimized, and demand responsive transport modes could be introduced, which would result in a more efficient operation.

5.4.3 Value-added services

In order to provide services for the users, data have to be collected, information has to be stored, and finally services have to be identified and receive the needed information to provide personalized services about several transportation modes.

To serve personal preferences, after the beginning of the trip the information system monitors and collects traffic data about the chosen route and sends alerts to the users [102]. Changing decisions of the users could be also supported (e.g. re-planning) [90]. Alternatives may be calculated based on personal preferences and the observed or revealed traveling behaviour [89]. These are initially given by the user and then the activity chain planner gets to know the traveling behaviour according to user decisions. In connection with it, estimation and planning of transport conditions based on dynamic data and other prognoses (e.g. weather forecast) could be provided [4], [164]. Thus it would be possible to adapt to the rapidly changing traffic situation by planning new alternative routes [183].

Special attention should be given to the support of multimodality and intermodality, because the destinations fall often outside of the service area of an operator. Accordingly an activity chain planner including more operators and transportation modes is more attractive. The railway, the urban transportation, the P+R, the B+R and the car-sharing systems should be integrated, thus realizing a comprehensive information service [154].

The activity chain planners could offer location-based services during the whole trip, especially at the transfer points. If the user wants to stop for a while at a transfer point, the activity chain planner can recalculate and refresh the further plan, and also help orientation at the bigger intermodal centres. During transfers and close to the destination the user may be supplied by some relevant, useful information (e.g. bank automats, shops, points of interest).
An ideal activity chain planner should offer premium information, which are defined as value-added information relevant for the users for what a business model could be built.

- Some of this information is presented as a personalized, targeted recommendation based on personal settings and self-learning capabilities, but more importantly only necessary information should be presented in decision situations and during navigation. [129].
- The activity chain planner could also use available data of service provider, for example leading passengers to the less crowded vehicles.
- The activity chain planners can provide a wide range of comfort information services. In the course of the booking procedure the users are given some information about their seats, about the equipment of the vehicle (e.g. Wi-Fi connection) and about other additional services.
- Using transfer management methods the planned transfers could be guaranteed. The disturbance information is transmitted also for the concerned transportation providers, so that they can wait for the transferring passengers. If the expected delay is too long, the transfer vehicle leaves, and a replacement vehicle would be shipped for the transferring passengers or compensation has to be realized (e.g. free food ticket, hotel reservation). The realized mobility guarantee enhances the reliability and quality of the public transportation system.

The activity chain planner should be tested with several users. With the collected information, adaptation of personal weights of the optimization aspects can be performed. This means that the user starts with average weights of the user groups, then the algorithm learns from the choices of the users, and adapts the weights according to personal preferences. Through learning personal weights, the analysis of belongingness of users to user groups could be also run (e.g. the user likes more biking than the average of the group).

The most important development trends of information systems were discussed. Most of the presented features are already implemented in certain systems separately, but they do not operate together. The new generation solutions should handle different information sources in an integrated way and take into account the different expectations of user groups to enhance the satisfaction level of travelers.
6 Thesis summary

6.1 Thesis 1. Evaluation method development for multimodal journey planners

In order to evaluate multimodal journey planners framework of aspects was defined. Based on the aspects qualitative evaluation method was elaborated to compare and rank the journey planners. With the creation of user groups I have taken into account requirements of the users. The method was applied to international and Hungarian journey planners.

The evaluation and comparison of multimodal journey planners was previously performed only in a descriptive (not in a quantitative) way. The aim was to provide a quantitative evaluation method, to that a framework of aspects was defined. The most important aspects from the passenger point of view are route-planning services, booking and payment, handled data, comfort service information, supplementary information.

The method consists of two main steps. First the journey planners were evaluated based on the elaborated framework of aspects, which resulted in the general evaluation number (scoring). For the scoring step the Multi-Criteria Analysis (MCA) was adapted, because it produces clear and well-comparable results. Since the appreciation of the information service depends significantly on the personal characteristics of the passengers, user groups were created from the passengers by age, mobility features and motion abilities. As a second step taking preferences of the user groups into account the average evaluation number was calculated (weighting). Finally the multimodal journey planners were compared to each other.

Popular journey planners were selected, whose innovative features were highlighted. The developments mainly focus on the actuality of handled data and supplementary information, while no real novel solutions were available concerning booking and payment. The elaborated method helps the operators to rate and compare information services from the viewpoint of the passengers. The introduction of user groups did slightly affect the evaluation numbers, but it highlighted some aspects (e.g. route planning, handled data), which are more important for certain user groups.

After the survey of international journey planners, the Hungarian online travel information services were analyzed, primarily bus transport operators (Volán). According to the evaluation it can be stated that Volánbusz offers the highest level of information services, Kisalföld Volán, Vértes Volán and Kunság Volán also have outstanding features, as route planning functions and handling of dynamic data. Finally the most important attributes of an ideal journey planner was defined based on top features of current systems and development trends.

Related publications: [49], [53], [54], [48], [55], [57]
6.2 Thesis 2. Analysis and creation of user groups based on preferences and evaluation aspects

In order to improve the accuracy of the evaluation, an analysis method regarding user group preferences was elaborated. A survey was performed to define the importance of the aspects. Based on similarities of the answers of users, new user groups were created with application of Ward method. With these enhancements, the evaluation method provides more specific results for decision makers about directions of development based on real needs of the users. To make the survey more consistent, a new survey method was applied by customization of the AHP and the Fuzzy AHP method.

In order to describe the passenger needs in quantitative way, user groups were defined. The importance of the aspects was defined by a survey. Having the results statistical analysis was performed. The most important main aspects are route planning (33%) and handled data (31%), while booking and payment (16%), comfort service information (10%), supplementary information (10%) have lower relevance. The ranking of multimodal journey planners was updated with the real user preference values. Although, big differences were expected among the user groups, no significant differences could be detected concerning the main aspects.

The correspondences between single aspects were investigated using correlation analysis. In case of high correlations latent user needs can be discovered. Considering connections among more aspects I applied Ward clustering method and created such new user groups, whose answers are similar within the group, and different among the groups. The clustering algorithm was implemented in MatLab. As a result 5 new user groups were created with the following highlighted preferences: searching for alternative routes, need for visualization on the map, interested in dynamic information, rejecting mobile payment, not interested in WiFi. With the application of the further elaborated method the results of the evaluation are based on real user needs, thus they support the journey planner operators and decision makers in the definition of possible directions of development.

The Fuzzy AHP based survey method was elaborated for weight calculation regarding main aspects. The weights are different for the user groups. The results of AHP are the pairwise relative evaluations of the main aspects. With the AHP method the consistency of the pairwise comparison values can be checked, which reduces the possibility of contradictory results. I defined the original and Fuzzy AHP weights, which were compared. As a result it could be obtained that using the Fuzzy AHP method the survey can be simplified. The original survey method can be further developed by the pairwise relative questions.

Related publications: [62], [24], [50], [56], [52], [51]
A method with flexible demand points was elaborated to optimize the set of activities of travelers regarding travel time and the number of postponed activities. To reduce optimization processing time, a genetic algorithm was applied. The algorithm was tested on a real transportation network in Budapest using GTFS timetable data.

The initial assumption was that some activities performed by the users during a day are not necessarily fixed temporally and spatially, therefore they can be carried out in different times or locations. The order of flexible demand points can be also changed. Before and after an activity a travel phase is realized by different transportation modes. By introducing flexible demand points, it is possible to find all combinations and to choose the optimal activity chain by implementing a solution for the TSP-TW problem. When establishing activity chains, it is assumed that the user already is aware of the activities of a certain day. With the introduction of flexible demand points the number of possible sets of activity chains grows exponentially, and the calculation needs more computational resources. Therefore a genetic algorithm was applied to reduce computation time by 90% enabling the application of many flexible demand points. The extension with POI search algorithm was also introduced.

The developed algorithm takes into consideration many constraints, as opening times of the shops or maximum waiting times before the planned arrival. During the implementation 3 different modes of transportation were determined: car, public transport and public transport with car-sharing opportunity. The optimization criterion was the minimum travel time, as the most important parameter. Also other parameters can be taken into account (e.g. comfort features), but these are generally hard to be quantified.

As an output of the optimization Pareto optimal results are presented, where the parameters are the number of postponed activities and the total travel times. The simulation of activity chain optimizations was performed on arbitrarily chosen test networks in Budapest using Matlab. In case of car usage about 8%, with public transport about 10% and with car-sharing opportunity about 14% decrease of the total travel time was realized. The elaborated method can be build in an advanced information service.

Related publications: [63], [64], [61], [59]
6.4 Thesis 4. Modeling the information system of activity chain optimization

I have modeled the information system containing the activity chain optimization method. The most important optimization parameters were identified, described and classified. The utility function was determined combining the optimization parameters and the aggregated weights, in this way the individual user preferences have been mapped. The information system architecture was modeled to map information about the environment and connections of the optimization algorithm. I have summarized future development directions of information services based on activity chain optimization.

In order to model the complex requirements of users regarding an urban activity chain, the possible parameters were identified. They were grouped into two main types: classification parameters (to create user groups) and optimization parameters (to calculate utilities regarding the optimization algorithm). In case of optimization parameters further grouping was performed, I have introduced general and comfort parameters. The possible values and data sources of the parameters were also identified. In case of comfort optimization parameters, the possible values were defined in such a way that low values represent ideal conditions and high values represent not preferred conditions.

Utility functions have been introduced in order to take into account the optimization parameters and to measure the goodness of activity chains. The utility functions were created from optimization parameters and weights. Weights related to comfort optimization parameters have been aggregated to decrease the number of required settings by the users.

I have modeled the information system architecture, which contains the optimization algorithm and its environment. The system architecture consists of the user interface, the server applications and external databases. The optimization parameters are provided through the external databases, where data sources and update frequencies of information services were identified. The model is the prerequisite and guideline for the implementation of the information system.

I have identified main development directions of information services based on comprehensive analysis, which are the following: data collection (reliable real-time data and crowd sourcing data), data storage (standardization and functional integration), value-added services (personal preferences, multimodality, location based services and premium information).

Related publications: [58], [60]
Acknowledgement

First of all I want to thank my supervisor, Dr. Csaba Csizár, who accepted me as his PhD student and provided me a very interesting research topic. I am very grateful for the continuous guidance and ideas of my supervisor, especially the very accurate corrections and improvement suggestions of the document.

I want to say thank you to the leaders of the department and the faculty, Dr. János Tóth and Dr. István Varga. Their support and encouragement helped me to go through the long way of writing the dissertation.

I am very grateful for the cooperation of my colleagues, the long hours of thinking about new ideas and writing common papers, especially Tamás Tettamanti, Bálint Caesar, Dénes Válóczi, Zoltán Rózsa, Zoltán Koppányi.

I also would like to thank my first undergrad supervisor, Dr. András Oláh, who set me on the track of the academic path and showed me the advantages of a research community.

Of course I am the most grateful for my family, my mother, my grandparents and relatives, who always supported me and encouraged me even in the hardest times.

Finally, I am very thankful for my friends, who supported my ideas and provided me with inspiration to continue with the work.
References


on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Budapest, Hungary, 03.06.2015-05.06.2015., ISBN: 978-963-313-142-8


[54] Esztergár-Kiss D., Csiszár Cs. (2012) Közforgalmú internetes utazástervező rendszerek multikritériumos értékelő elemzése, Transportation Science Reviews (Közlekedés és közlekedészet), Vol. LXXII. No. 6, pp. 21-31., ISSN 0023-4362 *


[56] Esztergár-Kiss D., Csiszár Cs. (2016) Utazástervező rendszerek értékelési szempontjaihoz tartozó súlyozott módszerrel, Transportation Science Reviews (Közlekedés és közlekedészet), ISSN 0023-4362 – accepted *


vizsgálatának módszere, ennek szempontja), Közlekedéstudományi Szemle, 2004, LIV. 7., pp. 253-263. *


*in Hungarian
Appendix

Appendix 1 Detailed description and evaluation of Hungarian online travel information services

In Hungary bus transportation is one of the most important transportation modes with a share of 2/3 within public transportation, but also with most differences in service quality among the operators [106]. In the last decades the companies worked with different operational conditions, internal structure and fleet sizes, which is a consequence of dissimilarities of the served areas. For example the average number of employees was ca 700, but the number of employees was 10x more for Volánbusz, than for Hatvani Volán. Concerning financing the income from the tickets covered only 25-30% of the total expenses. The difference had to be covered by governmental subsidies, therefore the development of modern passenger information systems was not possible without national or EU funds.

The new organizational structure of Volán operators was introduced in 2014, which aimed at the reduction of redundancies, more balanced number of employees, more efficient operation and more uniform service level for passengers. 7 regional operators were introduced by merging 2-6 previous operators (Fig 45), which are the following:

- ÉNYKK (Northern-Western Hungarian Transport Center),
- KNYKK (Middle-Western Hungarian Transport Center),
- KMKK (Middle-Eastern Hungarian Transport Center),
- ÉMKK (Northern Hungarian Transport Center),
- DDKK (Southern-Transdanubia Hungarian Transport Center),
- DAKK (Southern-Great Plain Hungarian Transport Center),
- Volánbusz (in Budapest and Pest county).

Fig 45 Spatial representation of the new organizational structure [106]
All bus related online travel information services in Hungary were studied, which are available to passengers on the internet and important for the domestic use. Only menetrendek.hu was not analyzed, as it was part of the international evaluation. Each operator is regional, uses only buses and is service provider dependent. The strengths and weaknesses of 23 online travel information services were also surveyed.

Considering features of the online travel information services the Kisalföld Volán possesses an extraordinary journey planner, which is the development of IQSYS and HC-Linear companies. The system can find the closest stop and transport service between two arbitrary chosen points on the map. The Vértes Volán and the Kunság Volán use the same planner as Kisalföld Volán, but its functionality and manageability lag behind the previous system. Nevertheless uniquely among the Volán operators some POIs can be shown, such as sightseeing places, health care facilities and shops. The Borsod Volán, the Hajdú Volán, the Jászkun Volán, the Balaton Volán, the Bakony Volán and the Vasi Volán do not provide journey planning service, only map visualization of stops and vehicles.

The Tisza Volán developed easy and well working seat reservation and online ticket purchase system for all lines with the option of selecting the exact seat. The Alba Volán and the Kunság Volán provide chip card pass for the passengers. The Borsod Volán, the Bakony Volán, the Somló Volán, the Vasi Volán and the Zala Volán use the seat reservation system of menetrendek.hu through a link.

Real-time information is handled only by the Kisalföld Volán and the Vasi Volán, the former gives alert in case of delay of the vehicle. The Agria Volán shows the delays using a very simple method (arrival times with colour code). In most of the cases the customer services are suitable, there is a forum for passengers developed by Kisalföld Volán and Tisza Volán. The Kisalföld Volán, the Vértes Volán, the Borsod Volán and the Tisza Volán show the low floor buses, while the Alba Volán established a homepage for disabled passengers.

The general evaluation numbers for the main aspects are presented in Table 31. The rows represent the aspects \( (i) \) and the columns represent the chosen journey planners \( (j) \). The general evaluation numbers of the online travel information services \( (u) \) are calculated by the summation of the evaluation matrix’s elements \( (p_{ij}) \) by columns. The maximum reachable points were 180, which includes values for single aspects added up. For each single aspect the maximum value was 10. The values were assigned to operators based on how much the chosen aspect was realized. For example Kisalföld Volán got 26 points for route-planning services, because starting points of a journey could only be searched by stops, but route-planning aspects were detailed and a good visualization of the routes was present.
The weighted evaluation of the online travel information services was also performed (Table 32), where the main aspects were weighted by preferences of the user groups. The values (in per cent) were calculated considering how close they are to the maximal given value of the certain aspect, which is therefore a relative value. In the last row of the table the average evaluation number ($a_{\text{avg}}$) is shown referred to all passengers.

Table 31 Evaluation of Hungarian online travel information services

<table>
<thead>
<tr>
<th>Main aspects</th>
<th>Name of operator after integration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ÉNVKK</td>
</tr>
<tr>
<td>1. Route-planning services</td>
<td>26</td>
</tr>
<tr>
<td>2. Booking and payment</td>
<td>7</td>
</tr>
<tr>
<td>3. Handled data, operational features</td>
<td>20</td>
</tr>
<tr>
<td>4. Comfort service information</td>
<td>1</td>
</tr>
<tr>
<td>5. Supplementary information</td>
<td>18</td>
</tr>
</tbody>
</table>

| General evaluation number        | 64    | 26    | 37   | 36    | 36   | 35    |
| %                                 | 36    | 14    | 21   | 18    | 20   | 18    |

Table 32 Evaluation of Hungarian online travel information services with user groups

<table>
<thead>
<tr>
<th>Main aspects</th>
<th>Name of operator after integration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ÉNVKK</td>
</tr>
<tr>
<td>Student</td>
<td>67</td>
</tr>
<tr>
<td>Worker</td>
<td>68</td>
</tr>
<tr>
<td>Tourist</td>
<td>58</td>
</tr>
<tr>
<td>Businessman</td>
<td>66</td>
</tr>
<tr>
<td>Pensioner</td>
<td>67</td>
</tr>
</tbody>
</table>

| Average evaluation number         | 66    | 26    | 37   | 31    | 36   | 33    | 57    | 20    | 16    | 19    | 16    | 25    | 27    | 38    | 23    |
| %                                 | 59    | 43    | 26   | 63    | 41   | 23    | 63    | 33    | 19    | 63    | 33    | 19    |

- 121 -
Appendix 2 Detailed description of the weighting of single aspects

1. Route-planning services:

The first main aspect contained two parts, data input and route planning service (Fig 46). Only those single aspects are shown in figures, which values are interesting or extraordinary. Concerning data input the participants were asked to provide information about the usage habits of journey planners (e.g. address, stop, GPS coordinates, POIs, pointing on the map). Using the Bartlett test for the address-based input, the standard deviation was rejected, because the critical value was 9,41 at 5% significant level, but the test value was 20,1. That means there is significant difference among the user groups. Using the t-test a difference among pensioners and all other groups was discovered.

In the case of name of the stops the t-test was performed, which has also confirmed the difference of businessman (4,81; 2,96) in contrast to other user groups (6,48; 2,76). Other data input methods, as pointing on the map (4,96; 2,89) or choosing POI (Point Of Interest) (4,95; 2,53) has an average importance for the user groups, which are not significantly different.

![Fig 46 Results of single aspects (route planning services)](image)

The route planning part contains information about travel duration, cost, waiting times, alternative routes and visualization on the map. The visualization on the map seems to be highly interesting for all user groups (8,91; 1,71), where the deviation is quite low compared to the other questions of this question group. In the case of pensioners (9,58; 0,7) the visualization plays a much more important role, than for other user groups, because it helps them much in the orientation. Concerning walking distance a difference is only to find in the ranked case between students and tourists.

2. Booking and payment:

Questions of booking and payment contained also two parts, pricing and payment. Concerning the pricing part questions were about fares, zones and reduced fares. Concerning the reduced fares students (increased from 8,05 to 8,1 after ranking) were more interested, than pensioners (from 5,58 to 5,64), because the pensioners may use the public transport services for a fixed price. The mean values were significantly different for almost all user groups, especially in the ranked case between students and workers-businessman. The least interested group was businessman (from 5,07 to 4,51).
The payment part referred to the payment opportunities, as cash, credit card or mobile payment (Fig 47). The most appreciated payment option was the credit card (after ranking from 5,96 to 5,56), then cash (from 4,99 to 4,65). But in both cases a high deviation was measured (from 3,55 to 3,33). For cash slight imbalances can be found concerning students-tourists and businessman. Mobile phone was barely used (from 2,1 to 1,96), because this payment method is not so wide-spread.

![Fig 47 Results of single aspects (booking and payment)](image)

3. Handled data:

Dealing with the main aspect of handled data (Fig 48), some operational features were investigated, as static (e.g. timetable, travel conditions), semi-dynamic (e.g. planned restrictions) and dynamic (e.g. traffic situation, estimated delay, alternative routes) data.

Information about timetable is far the most needed feature (from 9,15 to 10,77 after ranking), but concerning the deviation there are differences. In the case of students deviation (deviation after ranking: 1,1) is much lower, than among businessman and pensioners (deviation after ranking: 2,5). The Bartlett test rejected the equality of variances. This is verified by the t-tests, where between students and all other groups showed a significant difference.

The travel conditions – as expected – have less importance for all user groups (from 6,5 to 7,65) with a high variance. Both in the original and ranked case the actual traffic situation (8,62), the estimated delay (8,98), and the planned restrictions (8,86) reached very similar values. Alternative routes seem to be more important (9,39), because this information provides a solution, not only information about traffic.

![Fig 48 Results of single aspects (handled data)](image)
4. Comfort service information:

The next main aspect was comfort service information, which is about export features (e.g. printing or saving as PDF), favorites, Wi-Fi accessibility, luggage storage, electric supply or weather information. The answers for this question group resulted between 3 and 4, which is not significantly a very high value. Only the information about weather conditions was higher (5,18). Pensioners provided for these questions lower scores, but because of high variances this difference is not significant regarding the other user groups.

![Fig 49 Results of single aspects (comfort service information)](image)

5. Supplementary information:

The last main aspect was supplementary information (Fig 50), as environmental impacts, customer service, handicapped opportunities and low floor vehicles. Testing the answers for customer service information there was a significant difference between groups. Service information reached higher scores (4,42), than other features (3,2). The users are less concerned about environmental questions. The last question about low floor vehicles was originally more essential for students (6,41), than for workers (5,0), but after the ranking it became more important for workers (3,6), than for students (3,47).

Generally lower values were obtained, which means that these single aspects were not that important for the user groups. But it has to be considered that having high variances for this main aspect, the reliability of the answers is lower.

![Fig 50 Results of single aspects (supplementary information)](image)
Appendix 3 Detailed description of the results of the correlation analysis

Considering the correlation of the user groups, some differences of the preferred aspects have been arisen (Table 33). While in case of all user groups only 5 single aspect pairs were found with a high correlation (x > 0.7), then in the case of businessman 12, in case of workers 17, and the case of pensioners 22. This means that these 3 user groups provided more similar answers for some single aspects. Students and tourists did not reach a correlation level of 0.9 for any question pairs. This means that their opinion is not very consistent regarding the single aspects. Their diverse opinions could be explained with the higher number of users in these groups (29% and 25%).

<table>
<thead>
<tr>
<th>user group</th>
<th>1. single aspect</th>
<th>2. single aspect</th>
<th>correlation level</th>
<th>average values</th>
</tr>
</thead>
<tbody>
<tr>
<td>pensioners</td>
<td>luggage storage</td>
<td>information about shops</td>
<td>0.97</td>
<td>2.83/3.08</td>
</tr>
<tr>
<td>pensioners</td>
<td>WiFi at stations</td>
<td>luggage storage</td>
<td>0.89</td>
<td>2.55/2.63</td>
</tr>
<tr>
<td>pensioners</td>
<td>planned restrictions</td>
<td>information about current traffic situation</td>
<td>0.89</td>
<td>7.67/7.92</td>
</tr>
<tr>
<td>pensioners</td>
<td>WiFi at stations</td>
<td>WiFi in the vehicles</td>
<td>0.85</td>
<td>2.58/3.50</td>
</tr>
<tr>
<td>pensioners</td>
<td>information about current traffic situation</td>
<td>estimated delays</td>
<td>0.83</td>
<td>7.92/8.00</td>
</tr>
<tr>
<td>workers</td>
<td>WiFi at stations</td>
<td>WiFi in the vehicles</td>
<td>0.95</td>
<td>3.5/3.86</td>
</tr>
<tr>
<td>workers</td>
<td>address based data input</td>
<td>travel time and duration</td>
<td>0.94</td>
<td>9.32/9.32</td>
</tr>
<tr>
<td>workers</td>
<td>alternative routes during the planning phase</td>
<td>information about alternatives</td>
<td>0.94</td>
<td>7.82/7.64</td>
</tr>
<tr>
<td>workers</td>
<td>WiFi in the vehicles</td>
<td>electric supply in the vehicles</td>
<td>0.91</td>
<td>3.86/3.36</td>
</tr>
<tr>
<td>workers</td>
<td>WiFi at stations</td>
<td>electric supply in the vehicles</td>
<td>0.91</td>
<td>3.50/3.36</td>
</tr>
<tr>
<td>businessman</td>
<td>WiFi at stations</td>
<td>WiFi in the vehicles</td>
<td>0.87</td>
<td>3.37/3.63</td>
</tr>
<tr>
<td>businessman</td>
<td>luggage storage</td>
<td>dining opportunities</td>
<td>0.84</td>
<td>2.89/3.59</td>
</tr>
<tr>
<td>businessman</td>
<td>information about current traffic situation</td>
<td>estimated delays</td>
<td>0.81</td>
<td>7.67/7.41</td>
</tr>
<tr>
<td>businessman</td>
<td>visualization on the map</td>
<td>timetable</td>
<td>0.81</td>
<td>9.07/8.96</td>
</tr>
<tr>
<td>businessman</td>
<td>WiFi at stations</td>
<td>electric supply in the vehicles</td>
<td>0.8</td>
<td>3.37/3.33</td>
</tr>
<tr>
<td>students</td>
<td>WiFi at stations</td>
<td>WiFi in the vehicles</td>
<td>0.89</td>
<td>3.60/4.24</td>
</tr>
<tr>
<td>students</td>
<td>planned restrictions</td>
<td>information about current traffic situation</td>
<td>0.81</td>
<td>4.24/4.24</td>
</tr>
<tr>
<td>students</td>
<td>WiFi in the vehicles</td>
<td>electric supply in the vehicles</td>
<td>0.80</td>
<td>4.26/4.28</td>
</tr>
<tr>
<td>students</td>
<td>prices</td>
<td>information about discounts</td>
<td>0.75</td>
<td>7.67/7.85</td>
</tr>
<tr>
<td>tourists</td>
<td>WiFi at stations</td>
<td>WiFi in the vehicles</td>
<td>0.79</td>
<td>3.49/4.27</td>
</tr>
</tbody>
</table>

Table 33 Single aspect pairs of high correlation
The most interesting highest correlations were obtained for pensioners for questions about Wi-Fi at stations and luggage storage (correlation: 0,89; average values for single aspects: 2,83/3,08), where this groups showed less interest in both services. Some single aspect pairs appear for more user groups, as WiFi at stations and WiFi in the vehicles, information about current traffic situation and estimated delays.

Workers showed similarity in address based data input and travel time and duration (0,94; 9,32/9,32), alternative routes during the planning phase and information about alternatives (0,94; 7,82/7,64). They are mainly concerned about arrival to their workplaces on time, and if it would not be possible, they will try to find alternative routes. For businessman a highly ranked connection with the highest values (0,81; 9,07/8,96) is the visualization on the map and timetable, which is due to their working situation, where quick visual overview and punctuality is essential. Students found prices and information about discounts relevant (0,76; 7,67/8,25). The detailed results of the mentioned highly correlated single aspects are shown in Fig 51, where the distribution of the answers can be observed.

It has to be mentioned that negative correlation values were also higher, especially for businessman (-0,72) and pensioners (-0,72), while for students no single aspect pair reaches the correlation level of -0,4.
Appendix 4 Detailed description of the Fuzzy AHP analysis

From the fuzzy elements the m x m sized fuzzy AHP value matrix \( \tilde{A} \) can be generated.

\[
\tilde{A} = \begin{bmatrix}
(1,1,1) & \tilde{a}_{12} & \ldots & \tilde{a}_{1m} \\
\tilde{a}_{21} & (1,1,1) & \ldots & \tilde{a}_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{a}_{m1} & \tilde{a}_{12} & \ldots & (1,1,1)
\end{bmatrix}
\] (41)

The calculation of the weights is similar to the basic AHP method, however performing fuzzy operations is more complex. First aggregated fuzzy elements \( A_M \) are generated for each main aspect, where all three fuzzy elements are summed up separately.

\[
A_M = \sum_{k=1}^{m} (l_{M_k}, m_{M_k}, u_{M_k})
\] (42)

Then the degree of possibility of superiority \( \tilde{d}_{\text{sup}} \) between two aggregated fuzzy elements \( A_{M_2} \geq A_{M_1} \) is calculated, which values can be between 0 and 1. In case when the point with the highest membership function is higher for \( A_{M_2} \), then the degree of superiority is 1. In case when the lower limit of \( A_{M_1} \) is bigger than the upper limit of \( A_{M_2} \), then there is obviously no superiority possible. In any other cases the superiority is calculated based on the given equation.

\[
\tilde{d}_{\text{sup}}(A_{M_2} \geq A_{M_1}) = \begin{cases}
1, & m_2 \geq m_1 \\
0, & l_1 \geq u_2 \\
\frac{l_1-u_2}{(m_2-u_2)-(m_1-l_1)}, & \text{otherwise}
\end{cases}
\] (43)

Then the minimal degree of possibility of superiority \( \tilde{d}_{\text{supmin}} \) is searched for each aggregated fuzzy element over another aggregated fuzzy element. This gives an \( m \times (m-1) \) vector, which is then normalized. This final operation results in the Fuzzy weights \( \tilde{W}_j \).

Using the AHP and Fuzzy AHP method helps to define weights of main aspects for each user group more precisely (Table 34). As an input the survey results are used, which provide the proportions among two consecutive main aspects, which are route-planning services (1.), booking and payment (2.), handled data, operational features (3.), comfort service information (4.), supplementary information (5.). These proportions were defined for all user groups. With the calculation table the value matrix \( A \) was also generated separately for each user group, which contains exactly the AHP values.

Some results can be obtained from the proportions. Not surprisingly handled data (3.) is the most important main aspect, and route-planning services (1.) reached the second rank. This can be justified by the numbers in the table, as for each row of route planning service (3.) has values over 1. Furthermore this latter main aspect has the smallest difference among the user groups. Therefore it can be claimed that routing
options, alternatives and visualization of chosen routes are almost equally important for all users. Considering comfort service (4.) and supplementary information (5.) the user groups evaluated the former with higher points, especially students, businessman and pensioners.

<table>
<thead>
<tr>
<th>Proportions</th>
<th>AHP values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>2.</td>
</tr>
<tr>
<td>Student</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>1,00</td>
</tr>
<tr>
<td>2.</td>
<td>0,75</td>
</tr>
<tr>
<td>3.</td>
<td>1,16</td>
</tr>
<tr>
<td>4.</td>
<td>0,57</td>
</tr>
<tr>
<td>5.</td>
<td>0,75</td>
</tr>
<tr>
<td>Worker</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>1,00</td>
</tr>
<tr>
<td>2.</td>
<td>0,78</td>
</tr>
<tr>
<td>3.</td>
<td>1,08</td>
</tr>
<tr>
<td>4.</td>
<td>0,59</td>
</tr>
<tr>
<td>5.</td>
<td>0,70</td>
</tr>
<tr>
<td>Tourist</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>1,00</td>
</tr>
<tr>
<td>2.</td>
<td>0,80</td>
</tr>
<tr>
<td>3.</td>
<td>1,15</td>
</tr>
<tr>
<td>4.</td>
<td>0,57</td>
</tr>
<tr>
<td>5.</td>
<td>0,60</td>
</tr>
<tr>
<td>Businessman</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>1,00</td>
</tr>
<tr>
<td>2.</td>
<td>0,72</td>
</tr>
<tr>
<td>3.</td>
<td>1,18</td>
</tr>
<tr>
<td>4.</td>
<td>0,59</td>
</tr>
<tr>
<td>5.</td>
<td>0,75</td>
</tr>
<tr>
<td>Pensioner</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>1,00</td>
</tr>
<tr>
<td>2.</td>
<td>0,69</td>
</tr>
<tr>
<td>3.</td>
<td>1,22</td>
</tr>
<tr>
<td>4.</td>
<td>0,52</td>
</tr>
<tr>
<td>5.</td>
<td>0,71</td>
</tr>
</tbody>
</table>

Table 34 Proportions and AHP values of main aspects for each user group

The user groups had quite similar results when comparing route-planning services (1.) and supplementary information (5.). For students booking and payment (2.) and supplementary information (5.) were of the same importance, which means that younger people tend to be interested not only in the price of a journey, but also in information on low floor vehicles and environmental impacts. The biggest difference was 2,33 between handled data (3.) and comfort service information (4.) for the pensioners. They are more than twice keen on timetable and dynamic information, than on WiFi option and electric supply on vehicles. Tourists found booking and payment (2.) and route-planning services (1.) more equally important, than all other user groups (1,24). This confirms the fact, that tourists are very interested in different payment methods, e.g. using mobile phone or bank cards.

The general ranking remained the same for all weighting methods, however the measures changed (Table 35). Comparing the original weights and AHP weights big
differences are obtainable between the main aspects (23.2% for all main aspects), especially for handled data (from 27.8% to 37.9%). AHP weights are not reflecting values of the original weights, therefore AHP weights are not advised to be used in the evaluation method. Comparing the original weights and Fuzzy weights, smaller differences were present (10.2% for all main aspects). Handled data (from 27.8% to 24.9%) and route-planning services (from 23.9% to 21.7%) lost points, while the other main aspects gained some percentages, i.e. booking and payment (from 17.9% to 18.9%), comfort service information (from 13.6% to 16.4%), supplementary information (from 16.8% to 18.1%). These differences are caused by the different calculation method of Fuzzy AHP, but they are small enough to be applied in the evaluation method.

<table>
<thead>
<tr>
<th></th>
<th>Route-planning services</th>
<th>Booking and payment</th>
<th>Handled data, operational features</th>
<th>Comfort service information</th>
<th>Supplementary information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student</strong></td>
<td>0.236</td>
<td>0.178</td>
<td>0.275</td>
<td>0.136</td>
<td>0.177</td>
</tr>
<tr>
<td><strong>Worker</strong></td>
<td>0.241</td>
<td>0.187</td>
<td>0.261</td>
<td>0.141</td>
<td>0.170</td>
</tr>
<tr>
<td><strong>Tourist</strong></td>
<td>0.242</td>
<td>0.195</td>
<td>0.279</td>
<td>0.138</td>
<td>0.148</td>
</tr>
<tr>
<td><strong>Businessman</strong></td>
<td>0.236</td>
<td>0.169</td>
<td>0.278</td>
<td>0.140</td>
<td>0.177</td>
</tr>
<tr>
<td><strong>Pensioner</strong></td>
<td>0.242</td>
<td>0.167</td>
<td>0.294</td>
<td>0.126</td>
<td>0.171</td>
</tr>
<tr>
<td><strong>Student</strong></td>
<td>0.242</td>
<td>0.190</td>
<td>0.398</td>
<td>0.079</td>
<td>0.122</td>
</tr>
<tr>
<td><strong>Worker</strong></td>
<td>0.252</td>
<td>0.177</td>
<td>0.359</td>
<td>0.067</td>
<td>0.125</td>
</tr>
<tr>
<td><strong>Tourist</strong></td>
<td>0.267</td>
<td>0.171</td>
<td>0.370</td>
<td>0.079</td>
<td>0.113</td>
</tr>
<tr>
<td><strong>Businessman</strong></td>
<td>0.287</td>
<td>0.171</td>
<td>0.370</td>
<td>0.079</td>
<td>0.113</td>
</tr>
<tr>
<td><strong>Pensioner</strong></td>
<td>0.242</td>
<td>0.122</td>
<td>0.398</td>
<td>0.079</td>
<td>0.160</td>
</tr>
<tr>
<td><strong>Student</strong></td>
<td>0.215</td>
<td>0.183</td>
<td>0.264</td>
<td>0.156</td>
<td>0.183</td>
</tr>
<tr>
<td><strong>Worker</strong></td>
<td>0.213</td>
<td>0.198</td>
<td>0.227</td>
<td>0.174</td>
<td>0.187</td>
</tr>
<tr>
<td><strong>Tourist</strong></td>
<td>0.229</td>
<td>0.197</td>
<td>0.244</td>
<td>0.160</td>
<td>0.170</td>
</tr>
<tr>
<td><strong>Businessman</strong></td>
<td>0.214</td>
<td>0.185</td>
<td>0.245</td>
<td>0.171</td>
<td>0.185</td>
</tr>
<tr>
<td><strong>Pensioner</strong></td>
<td>0.215</td>
<td>0.183</td>
<td>0.264</td>
<td>0.156</td>
<td>0.183</td>
</tr>
</tbody>
</table>

Table 35 Numerical comparison: original weights, AHP weights and Fuzzy weights
Appendix 5 Detailed description of travel chain analysis of Berlin dataset

The participants had to report the departure and destination location of their trips. In order to describe their mobility, it is essential to acquire the coordinates of these locations. The reported locations from the survey are represented with the ID of the districts and the sub-districts. Berlin is divided to 12 districts (Table 36), and the districts are further divided into 187 smaller areas (here we refer them as sub-districts).

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>ID</th>
<th>Name</th>
<th>ID</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Mitte</td>
<td>05</td>
<td>Spandau</td>
<td>09</td>
<td>Treptow-Köpenick</td>
</tr>
<tr>
<td>02</td>
<td>Friedrichshain-Kreuzberg</td>
<td>06</td>
<td>Steglitz-Zehlendorf</td>
<td>10</td>
<td>Marzahn-Hellersdorf</td>
</tr>
<tr>
<td>03</td>
<td>Pankow</td>
<td>07</td>
<td>Tempelhof-Schöneberg</td>
<td>11</td>
<td>Lichtenberg</td>
</tr>
<tr>
<td>04</td>
<td>Charlottenburg-Wilmersdorf</td>
<td>08</td>
<td>Neukölln</td>
<td>12</td>
<td>Reinickendorf</td>
</tr>
</tbody>
</table>

Table 36 District IDs and their names

The relationship between the name of the districts and their numbers is unambiguous. Knowing the names of the districts, their centroids (i.e. center point of the area) can be geocoded into WGS84 geographical coordinates. For this matching Google Geocoding API was used. Fig 52 shows the sub-districts' coordinates inside the border of Berlin. Note that these coordinates are only the estimation of the real district centers due to the geocoding error.

Fig 52 Centroids of sub-districts derived from Google Geocoding API

Spatial resolution is decreased by considering only trips between districts instead of sub-districts, thus transitions become more distinguishable for the visual comparison. Note that the dataset became also smaller, because only those trips have been investigated, where destination and departure districts are different. In order to filter unreliable data, persons reported less than 3 trips or filled the forms partially are omitted from the analysis. Furthermore, trips with departure or destination location outside Berlin are also filtered. The final dataset after filtering contains 4,462 persons and 27,888 reported trips that is approximately 1/5 of the total dataset.

- 130 -
The departures and destinations are available using geocoding, thus directed adjacency matrix can be built from the reported trips:

\[ A_{ij} = \begin{cases} n_{ij}, & \text{if there is connection between the } i\text{th and } j\text{th districts} \\ 0, & \text{otherwise} \end{cases} \]

where \( n_{ij} \) is the number of trips between the \( i \)th and \( j \)th districts.

Using adjacency matrix the traffic connection between districts can be visualized as a graph. The locations of the graph nodes are derived from the WGS84 coordinates of the districts. In order to show as much information as possible in a single figure, some other graphical tools are used: the size of the node represents the number of visits, and the width of the edges reflects the number of trips between two districts. The edges are represented as a curve and their direction can be determined by their curvature that follows clock-wise direction (Fig 53).

In the example the sizes of the nodes show that the number of visits is greater in Node 1 than in Node 2. The two curves (Edge 1 and Edge 2) between the nodes represent that there are trips in both direction. The direction can be determined by following the clock-wise directed curvature. The width of the edges represents the number of trips between the nodes. In this case Edge 2 contains more trips than Edge 1.

Fig 53 Graphical representation of the graph

Spatial resolution is decreased by considering only trips between districts instead of sub-districts, thus transitions become more distinguishable for the visual comparison. Note that the dataset became also smaller, because only those trips have been investigated, where destination and departure districts are different. In order to filter unreliable data, persons reported less than 3 trips or filled the forms partially are omitted from the analysis. Furthermore, trips with departure or destination location outside Berlin are also filtered. The final dataset after filtering contains 4,462 persons and 27,888 reported trips that is approximately 1/5 of the total dataset.

Fig 54 shows the reported trips between districts grouped by time intervals of the day and the purpose of the trips. Two selected purposes are presented: work and entertainment (such as watching a movie, or going to theater). The visit frequencies of the locations are normalized by 30 visits. The width of the edges also represents the visit frequencies, but only between two districts, and is also normalized. During the
normalization process the highest frequency of the sub-figures was set to a predefined value (in our case 1), and all the others in the same sub-figure were calculated proportionally according to this highest frequency; this value is displayed in the legend of the sub-figures. Note that the size of the nodes can be comparable through all the sub-figures due to the same normalization constant (here 30), but the edges can be only comparable within the same sub-figure, not between two sub-figures due to different constants.

Fig 54 Connection between the districts sorted by time and purpose

The sub-figures of the reported trips show that the central districts are playing a key role in the life of Berlin. District 1 (Mitte) is the central target area for working; in the morning passengers commute generally from outside districts to District 1. In addition, other major working areas are District 2, 4 and 7 (Friedrichshain-Kreuzberg, Charlottenburg-Wilmersdorf, Tempelhof-Schöneberg), which are located around District 1. The afternoon period is very surprising, because it shows more work trips; it is hard to say that this result comes from the unreliability or incompleteness of the dataset or this is the real situation. Anyway, it is worth to analyze this problem using different datasets and methods, because it can provide beneficial information to transportation authorities.
The number of entertainment purpose trips is slightly growing from morning to midnight. The frequently visited districts are distributed more uniformly than in other cases. The overall graph of the entertainment trips shows that the outer districts tend to be as important as central districts.

Similarly transportation modes were also examined in Fig 55 Connection regarding bicycle, car and public transportation. Cycling is popular in Berlin, the use of this transportation mode is located mainly around the central districts. People living in the outskirts prefer to use car to reach their workplaces or the centrum due to longer distances. Also note that cycling is more popular in the afternoon hours.

The dominance of the cars are not surprising, Berlin is a huge metropolis with well-developed road infrastructure. Comparing the sub-figures visits between districts show same frequencies through the day, which means that the mass points do not change in the 24 hours. The main destinations are the central districts. However, the rate of car travels is high compared to other transportation modes. Sub-figures of bicycle in the afternoon show similar frequencies as cars in the morning.

Concerning public transportation travelers use this transportation mode very frequently in order to reach their destinations, in the afternoon hours nearly as much, as cars. The importance of public transportation can be noticed especially in the morning hours.

Fig 55 Connection between the districts sorted by time and transportation mode
Annexes

Annex 1 Detailed calculations for the evaluation of multimodal journey planners

Route planning services:
- ways of data input: address (1), name of stop (2), institutions and service facilities (2) (e.g. museums, restaurants, swimming pool, cinemas, offices, schools), GPS coordinates (3), pointing out on the map (2);
- planning aspects: departure and arrival time (1), duration and costs (2), number of transfers and walking distance (1), other aspects (4) (e.g. preferred transportation mode, P&R, R, car-sharing);
- displayed data: travel duration and distance (2), transfer information and location plans (2), walking times (2), walking time and distance (2), alternative routes (2);

Payment and payment:
- fare information: zones (2), prices (2), reduced fees separately (2) (e.g. tabular view, fee of the entire travel chain (2), calculation method (2);
- method of booking and payment: way of data input (4), in how many steps (2), what kind of data is needed (2), possibility of choosing costs (2);
- payment options: types of bank cards (2), payment per mobile phones (2), transaction fees (2), types of vouchers (4) (e.g. SMS, code per e-mail, paper ticket printed at home/at the station/sent by post);

Handed data, operational features:
- static data timetable for a given route and/or date (6), travel conditions and rules (2) (e.g. animals, luggage), export features (2) (e.g. PDF, printing);
- semi-dynamic data: information about deviations (2), list of planned restrictions (4), visualization of planned restrictions (4);
- dynamic and estimated data: information about current and extraordinary traffic situations (4) (e.g. weather conditions, accidents), usage of crowd sourcing data (2), deviation from the timetable (2), calculation of the probable impacts of the extraordinary traffic situations (2) (e.g. alternative routes);

- personal data: creating a profile (2), setting personal preferences (4), saving searches and favourites (2), personalised offers (2);

Convenient service information:
- services at the stations/stops: Wi-Fi (2), luggage storage (2), other services (6) (e.g. newspaper, bakery, car-sharing);
- services on board: Wi-Fi (2), electrical supply (2), information about dining opportunities (3), other services (4);
- additional services: weather forecast (2), booking a room (2), car rental (2), sightseeing (2), opening time of the shops (2);

Supplementary information:
- environmental impact: degree of air pollution (6) (e.g. CO2, energy consumption (2), comparison of transport modes and travel chain (2);
- information in foreign language: only the planning module or the whole homepage is translated (6), number of foreign languages (2), automatic language choice based on IP address (2);
- customer service: requesting information via email and/or telephone (4), feedback opportunities (2), opinion about travel services (2), forum (2);
- information on equal opportunity: routes for disabled passengers (4), information about vehicles (2), webpages for visually impaired people (4);

| Town | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| Tandu | 4.98 | 4.10 | 4.50 | 3.50 | 1.03 | 3.60 | 4.50 | 3.31 | 2.53 | 3.38 | 1.64 | 6.33 | 4.97 | 3.62 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Ingel | 3.86 | 4.10 | 4.55 | 3.17 | 2.54 | 4.45 | 5.15 | 3.40 | 5.35 | 5.54 | 3.13 | 6.53 | 5.06 | 3.47 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Turku | 7.45 | 4.10 | 4.33 | 3.48 | 3.00 | 5.05 | 5.80 | 3.50 | 2.91 | 4.96 | 3.83 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Ylemeri | 4.85 | 7.32 | 4.23 | 3.41 | 1.30 | 5.83 | 4.33 | 3.23 | 2.93 | 4.59 | 2.68 | 7.18 | 4.51 | 3.73 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Tampere | 5.22 | 4.12 | 4.76 | 3.88 | 1.28 | 4.55 | 4.65 | 3.49 | 3.94 | 3.48 | 1.59 | 6.27 | 4.97 | 3.68 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

- 134 -
Annex 2 Survey of the user groups regarding aspects of journey planners

Hogyan használja az utazástervezőket?

A kérdőív célja az utazási szokások, különösen az internetes utazástervezőkkel kapcsolatos igények és tapasztalatok felmérése. Az utazástervezőket mindannyian ismerjük és használjuk, mint például Utvonaltervező, Elvira, Google Maps.

Nincs az egyes utazások érdekeltének miatt, hanem általános utazási szokásokat vizsgálnak. Tehát a kérdéseknek az általánosan érdekelőbb jelenlétű válaszra vagyunk kíváncsiak!

A felmérés teljes mértékben anonim, kizárólag kutatási-figyelemzési célokat szolgál. Kérjük, szája rá azt az 5-10 percet a különféle nagy segítséget jelent számunkra a válaszai

Közönséget:

BME Közlekedésipari és Közlekedéspolitikai Tanszék multimodális közlekedési rendszerek projekt team

1. Mennyire fontosak a következő tulajdonságok egy utazástervezőnek?
   (Tegye sorrendbe a válaszokat. minden esetben csak 1 választ jelöljön meg!)

<table>
<thead>
<tr>
<th>5. hely: a legkevésbé fontos tulajdonság</th>
<th>4. hely</th>
<th>3. hely</th>
<th>2. hely</th>
<th>1. hely: a legfontosabb tulajdonság</th>
</tr>
</thead>
<tbody>
<tr>
<td>utvonaltervezés</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>(adatbevétel, tervezési szempontok, megjelenítés)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>helyfoglalás, cijítetés</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>(tarifainformáció, mobil fizetési lehetőségek)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>információ az utazasokról</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>(tervezett lezárások, közlekedési dugók, késések előrejelzése)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>információ a szolgáltatásokról (WiFi a járművön, közeli üzletek listája)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>kiegészítő információ</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>(könyvek, szakmai hírek, idegen nyelvű iratok, információ meglásását)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

2. Milyen gyakran használja a következő adatbevételi lehetőségeket egy internetes utazástervezés során?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>clima</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>megálló neve</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>közintézmények (pl: híradó, iskolák)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>kulturális és sport létesítmények (pl: múzeum, színház, színpad)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>éttermek, szórakozóhelyek</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>GPS koordináták</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>raktár és egy posta a térképen</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
3. Mennyire fontosak a következő információk Önnek egy internetes utazástervezés során?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>utazási idő és távolság</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>költségek</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>átszállások</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>számla</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gyaloglási idő és távolság</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>variációi idő</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alternatív útvonalak</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>területi megjelenítés</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P+R és B+R lehetőségek</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Mennyire fontosak a következő tarifával kapcsolatos információk Önnek?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>áruk (pl. jegy, bérlő)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zónák területe és árai</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kedvezmények (pl. diák, nyugdíjas)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Milyen gyakran használja a következő fizetési lehetőségeket egy internetes utazástervezés során?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>készpénz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bankkártya</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fizetés mobil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>egyénileg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>egyéb (pl. utalvány, kupon)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. Mennyire érdeklik a következő utazással kapcsolatos információk?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>menetrend</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>utazási feltételek</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tervezett korlátogatások és lezárások</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>információ az aktuális forgalmi helyzetéről</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>becslési kézi értékek</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alternatív útvonalajánlás</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7. Milyen gyakran néz utána a következő szolgáltatásoknak egy internetes utazástervezés során?
1 : soha / 10 : mindig

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>exportálás lehetségek (pl. PDF, nyomatékalap)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>következók (pl. gyakran használt útvonalak mentése)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WiFi az állomáson</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WiFi a járművön</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>csomagmeghártyázás</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>elektromos áramellátás a járművön</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>elkezdi lehetőségek</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>időjárás előrejelzés</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>egyéb (pl. vasúti lehetőségek, autóerőforrás, szabadfoglalás)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8. Mennyire tartja fontosnak a következő szolgáltatásokat egy internetes utazástervezés során?
1 : egyáltalán nem fontos / 10 : kifejezetten fontos

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>könyezetlét hatások (levágózavaros, energiafelhasználás)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ügyfélszolgálat (telefonos, e-mall)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>felhasználók fórum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>honlap gyengén látóknak</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alcsony padlós járművön</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. Hány éves Ön?

10. Foglalkoztatottság alapján melyik kategorizálta tartozik?

11. Mozgásban vagy látásban korlátozott?

12. Milyen célból használja az utazástervezőket?
- munka (napilény, ügyintézés)
- szabadidő (sport, vasárnap, kultúra)
- tanulás (hosszú távú utak)
Annex 3 Statistical analysis of answers of user groups regarding single aspects
### Annex 4 Correlation results of the user groups

#### Correlation of all user groups

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 |
|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---
### Correlation of businesses

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

### Correlation of services

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

### Correlation of professors

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Annex 5 Matlab code for creation of Ward user groups

```matlab
clear all;
clol;
% Excel data input
exl = actxserver('excel.application');
exlWBk = exl.Workbooks;
exlFile = exlWBk.Open('C:\Users\[username]\Documents\[folder]\[file].xls');
exlWBk = exlFile;
exlSheet = exlWBk.Sheets.Item('[Sheet]');
sub = exlSheet.Columns.End(4); % Find the end of the column
numrow = sub row; % And determine what row it is
dat_range = '['sub.colstr(numrow)+']'; % Read to the last row
endrow = exlWBk.Range(dat_range);
exlData = endrow.Value;
for ii = 1:size(exlData,2)
    ExcelData(:,ii) = reshape(exlData(:,ii),size(exlData(:,ii)));
end

mData

% matrix row size [m], matrix column size [n]
matshape(1,2)

% A(k,n), b(k,n)
for j=1:n
    c=0
    for i=1:m
        c=c+ExcelData(i,j)
    end
    a(i)=c/n
end

% square VAR(k,0,1))
% normalized matrix (h)
for k=1:n
    for j=1:n
        ExcelData(k,j)=[ExcelData(k,j)-mean(k)]/std(k),
    end
end

% tavolag (d)
for k=1:n
    for j=1:n
        d(k,j)=ExcelData(k,j)-mean(j),,2+,
    end
end

% italakzi Ward inputt\ddot{a}

% Ward m\ddot{u}tesez, ill. Single linkage
V=
% Z\ddot{e}leke\ddot{e}, [\text{\textquoteleft}ward\textquoteright]\
V=\operatorname{cluster}(Z, \text{\textquoteleft}maxclust\textquoteright, R);
% exportform
T=\operatorname{cluster}(Z, \text{\textquoteleft}maxclust\textquoteright, T);
```
Annex 6 Finding similarities within and differences among the Ward user groups