Manuela Rauch, BSc.

Dashboard Design for Monitoring, Visual Analysis and Decision Support in Industry Scenarios

Master's Thesis

to achieve the university degree of

Master of Science

Master's degree programme: Software Engineering and Management

submitted to

Graz University of Technology

Supervisor

Vedran Sabol, Dipl-Ing. Dr.techn.

Institute of Interactive Systems and Data Science

Graz, May 2019

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date

Signature

Acknowledgements

I want to thank Know-Center and especially my supervisor Vedran Sabol for giving me the opportunity to develop dashboards for many different industry-related scenarios and thus, receiving insights on data analysis performed in different domains. Additionally, I want to thank all colleagues from Know-Center who have been involved in the described projects. Conducting all these projects would not have been possible without the collaboration and support of many colleagues.

I want to thank my family, especially my parents, for supporting me throughout my studies. Without their support, I would not have had the possibility to study and write this thesis. Thus, I want to dedicate this thesis to them.

I want to thank my partner Martin for his constant support, encouraging me to finish my studies and patience throughout that time.

Besides may parents, I want to dedicate this thesis also to my prematurely deceased colleague Werner Klieber who put all his energy in conducting projects successfully.

Abstract

The amount of data available on the web, generated during production processes, stored on mass storage devices, in databases or in clouds, is enormous. Analysing and interpreting available data in the right way and thus, taking the right decision within a short period of time is not possible without further support. Visualisations have been identified as being suitable to support these processes. Therefore, this thesis introduces data dashboards suitable for monitoring, analysis and decision support by enabling analysts to access the required information in a visually appealing way.

Dashboards for seven different industry scenarios in four application domains, namely production plant monitoring and analysis, environmental data monitoring, project portfolio monitoring as well as competitor monitoring and analysis, have been carefully designed during several iterations to meet all requirements. Although the described dashboards have been implemented to handle problems in specific domains, many of them are general approaches, which can be easily applicable to other domains. The discussed dashboards have been designed following well-known dashboard design guidelines. Details about the usability of the dashboards related to these guidelines were obtained by user feedback and evaluations.

Evaluations of dashboards and collected user feedback revealed that the provided solutions are completely new approaches compared to previous analysis methods and are well suitable for their purposes. It is expected that they enable companies to obtain significant advantage over their competitors by enhancing the analysis and decision processes.

Keywords: dashboard design, design guidelines, monitoring, decision support, visualisation, visual analysis, evaluation, user feedback

Kurzfassung

Die Menge an Daten die im Internet verfügbar ist, die während verschiedener Produktionsschritte generiert werden, auf Massenspeichern, in Datenbanken oder in Cloud Applikationen gespeichert sind, ist enorm. Um diese Menge an Daten zu analysieren und die richtigen Schlüsse daraus zu ziehen, werden zusätzliche Hilfsmittel benötigt. Analysen haben gezeigt, dass Visualisierungen ein geeignetes Mittel sind um diese Prozesse zu unterstützen. Diese Arbeit beschreibt den iterative Design Prozess, die Implementierung und die Evaluierung von visuell ansprechenden Dashboards, welche für die Überwachung und Analyse von Daten als Entscheidungshilfe dienen.

Dashboards für sieben industrie-bezogene Szenarien in vier unterschiedlichen Anwendungsdomänen, nämlich Monitoring und Analyse von Produktionsdaten, Monitoring von Umgebungsdaten, Monitoring von Projekt Portfolios und Konkurrenzmonitoring und -analyse, wurden sorgfältig entworfen und in mehreren Iterationen verfeinert um alle Anforderungen zu entsprechen. Obwohl die beschriebenen Dashboards für spezielle Anwendungesfälle entworfen wurden, können viele auch in anderen Domänen Anwendung finden. Die beschriebenen Dashboards wurden unter Berücksichtigung bekannter Dashboard Design Richtlinien erstellt. Details hinsichtlich der Verwendbarkeit der besprochenen Dashboards in Bezug auf die Richtlinien wurden durch Benutzerfeedback und Evaluierungen erlangt.

Evaluierungsergebnisse und Benutzerfeedback haben gezeigt, dass die vorgestellten Dashboards im Vergleich zu vorherigen Analyse Methoden neue Ansätze in der Datenanalyse aufzeigen und ihre Zwecke gut erfüllen. Die Dashboards ermöglichen es Unternehmen ihre Analysen und Entscheidungsprozesse zu verbessern und damit einen wirtschaftlichen Vorteil gegenüber ihren Mitbewerbern zu erlagen.

Stichwörter: Dashboard Design, Design Richtlinien, Überwachung, Entscheidungshilfe, Visualisierung, visuelle Analyse, Evaluierung, Benutzerfeedback

Contents

List of Figures

List of Tables

Abbreviations

API	Application Programming Interface
CSV	Comma-Separated Values
CDTW	Constraint Dynamic Time Wrapping
DOM	Document Object Model
DTW	Dynamic Time Wrapping
D3	Data-Driven Documents
ERP	Enterprise Resource Planning
FDP	Force-Directed Placement
GWT	Google Web Toolkit
HTML	Hypertext Markup Language
HCP	Honeywell Connected Plant
IOT	Internet of Things
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
MES	Manufacturing Execution System
MIT	Massachusetts Institute of Technology
NLP	Natural Language Processing
PDF	Portable Document Format
PCA	Principle Component Analysis
PLM	Product Lifecycle Management
PLS	Partial Least Square
PNG	Portable Network Graphics
RDF	Resource Description Framework
RA	Rockwell Automation
RTS2	Regular Time Series File Format 2
SAX	Symbolic Aggregate Approximation
SOM	Self-Organising Map
SVG	Scalable Vector Graphics
SVM	Support Vector Machine
UI	User Interface
URL	Uniform Resource Locator
XML	Extensible Markup Language
WDC	Web Data Connector

Chapter 1

Introduction

This thesis discusses data dashboard designs for industry scenarios enabling users to monitor processes, visually analyse data and finally, take the right decisions based on the data shown within the dashboard. Thus, dashboards need to be designed carefully to meet all user requirements, which is usually done through several iterative steps. This includes collecting and considering feedback.

This thesis illustrates design processes for seven different industry scenarios. Their interaction and analysis possibilities are described and outlined in case studies. The evaluation of selected dashboards provides information on the usability of the proposed dashboard designs. Additionally, the proposed dashboards are investigated with regards to previously defined guidelines for dashboard design, and, to understand how well they fulfil the requirements.

In order to design dashboards, procedures are always similar but can slightly deviate depending on the requirements, the project volume and the main goal of the approach. In general, the first step always includes the identification of a specific problem and a lack of possibilities in handling this problem. To investigate possible solutions, the next step should always include the analysis of the current environment. Afterwards, requirements to overcome the specific problems need to be discussed in detail. This can be illustrated by describing specific use cases. After this, a concept to overcome the specific problem needs to be developed. This requires an investigation of available data but also the definition of application programming interfaces. Finally, the designs of the dashboard and server components are developed in order to serve all requirements. All these processes are required before the actual implementation can be performed. Finally, an evaluation supports dashboard designers in investigating whether the proposed dashboard serves all requirements.

Scenarios in this thesis do not outline the process of problem identification and the analysis of current processes since this would reveal details on production and analysis methods used in different companies. Nevertheless, general problems can be derived from the motivation and the requirements discussed for each scenario. Depending on available resources, use cases were described to gain further insights on all requirements. The process of data investigation is also not outlined in detail. This is again due to not revealing company secrets. Since this thesis concentrates on the design of the dashboard based on the motivation and requirements, the process of dashboard design and refinement is discussed in detail for each scenario. Besides the dashboard description, a case study provides further insights on how the dashboard can be used for data analysis. Finally, each dashboard is analysed with regard to pre-defined guidelines.

This thesis outlines the design process of data dashboards according to well-known design guidelines in different application domains, discusses lessons learnt obtained throughout the project but also through user feedback and evaluations, which were performed for three dashboards.

1.1 Motivation

The main motivation for this thesis is to develop dashboards for different industry related scenarios by outlining the design process supporting analysis and decision tasks. The proposed approaches were implemented for specific scenarios, but can be applied to different domains in most cases.

The demand for creating a dashboard is always derived from upcoming problems and missing opportunities to handle them using available tools and methods. Companies have many tools for monitoring production processes and data analysis. Often these tools provide just single numbers as a result, table representation or simple charts without further interaction possibilities and therefore, only provide limited analysis functionality.

Interactive visualisations have not been considered for monitoring, data analysis and decision support for a long time. This is for example due to missing digitalisation and standardisation of processes, but mainly due to a lack of expertise in visual data analysis. Handling interactive visualisations and drawing the right conclusions based on their content, of course requires a learning phase, additional expertise, while also consuming a lot of time. However, dashboards can provide an overview of the underlying data, can show data distributions and correlations, can support users in discovering and monitoring trends, can enable them in identifying anomalies and can support them in understanding cause effect relationships. Thus, dashboards enable users to react on unexpected events faster by drawing the right conclusions based on visualised data. Therefore, designing dashboards supporting these tasks while not overloading them with wrong or unimportant information often requires several iterations.

Dashboards for monitoring production processes and supporting analysis tasks enable users taking the right decisions faster and thus having significant advantages over companies not using such dashboards.

1.2 Structure of Thesis

This thesis is structured as follows.

Chapter 2 discusses related work for the four main application domains, which are handled within this thesis. Initially, dashboards and research in the area of project portfolio monitoring are discussed. Afterwards, dashboards for monitoring production plants are investigated. This ranges from tools, which are directly connected to the production plant and monitor data in real time or with some delay, tool used for analysing production data retrospectively and finally, tools used for analysing sensor data in particular. This is due to the fact that production plants usually generate many sensor measurements during the production process. Since search applications have been identified as being suitable for monitoring and analysis of competitors, commercial tools but also research in this area is discussed in detail. Finally, the ten most important guidelines, which should be considered when designing a dashboard for data analysis and monitoring, are discussed.

Chapters three to nine discuss the actual dashboards proposed in this thesis. These include seven dashboards in four application domains. These domains are (i) project portfolio monitoring, (ii) production plant monitoring and analysis, (iii) environmental data monitoring and (iv) competitor monitoring and analysis. Each scenario discussed in this thesis has a similar structure. The introduction provides a brief overview of the theme, the scope and the role of the author of this thesis. The motivation provides information about the purpose of providing a dashboard for the specific scenario. Afterwards, the requirements for the dashboard to perform different analysis steps and taking decisions are provided. This is followed by the description of the design process including single components and the overall dashboard design. If a mobile application was proposed, this is also described in detail. A case study provides an overview on how the dashboard can be used for data analysis. Furthermore, a short description about the implementation and technical details are provided. Each chapter concludes by discussing which design guidelines were considered during the design process and by providing a conclusion.

Chapter 3 discusses a dashboard for software project portfolio monitoring and is assigned to the first domain. In this scenario the main focus lies in the design of the project symbol and the representation in portfolio views. An evaluation was designed and conducted to receive feedback on the readability of the symbol and the usability of the provided portfolios, which is discussed in chapter 10. The author of this thesis contributed to the portfolio layout, fully implemented the dashboard, designed and performed the evaluation of the dashboard and published the results in the full paper *A Visual Approach to Project and Portfolio Monitoring* [98].

Chapter 4 discusses a dashboard for monitoring production plants on different levels of detail, enabling users to monitor current production and taking actions if required. This dashboard is assigned to the second domain. The author of this thesis was responsible for project management and for the conceptual design of the dashboard.

Chapter 5 discusses a dashboard for monitoring production of electronic components by animating the production process. This dashboard is also assigned to the second domain.

This allows users to observe the production flow from one station to another, and identifying bottlenecks within the production process. The author of this thesis was responsible for project management, mainly contributed to the conceptual design and fully implemented the dashboard and server component.

Chapter 6 discusses a dashboard for analysing large-scale time series data. In contrast to the previous two approaches, this dashboard is suitable for analysing sensor measurements afterwards rather than monitoring measurements in near real time. As the previous two, this dashboard is also assigned to the seconds domain. The author of this thesis was responsible for the conceptual design of the dashboard and mainly implemented the dashboard and server component.

Chapter 7 discusses a dashboard for analysing data from different production processes with different formats and thus is also assigned to the second domain. In contrast to all other scenarios, the framework itself, which enables users to select visualisations of interest, was not part of this thesis. The main focus in this project was the process of selecting the right visualisations and configuring a dashboard for enabling users to investigate data from different sources, while all data is processed on the client for security reasons. An evaluation was conducted to provide insights on the usability of the proposed approach, which is described in chapter 10. The author of this thesis was responsible for the conceptual design of the dashboard, implemented central visualisation components, designed and performed and evaluation and published the results in the poster *Evaluation of Visual Decision Support Systems used in Semiconductor Industry* [97] and contributed to the deliverable *Evaluation report on the effectiveness of the dashboard components* [26].

Chapter 8 discusses a dashboard for analysing environmental sensor data and is therefore assigned to the third domain. By providing a dashboard that allows to analyse geographical as well as sensor measurements, this approach enables users to identify wrong measurements not only by investigating plausibility of single values but also by comparing sensors from close locations. The author of this thesis was responsible for the conceptual design of the dashboard as well as fully implemented the dashboard.

Chapter 9 discusses a dashboard suitable for monitoring and analysis of competitors and is therefore assigned to the fourth domain. This is done by crawling interesting web pages and providing information to users by executing search queries. Thus, this dashboard is based on a search application allowing users to investigate search result but also connections between search results and extracted entities in detail. The author of this thesis was responsible for the conceptual design of the dashboard, developed major dashboard and visualisation components, designed and performed an evaluation and was the first author of two full papers, one describing the search dashboard *Knowminer Search - A Multi-visualisation Collaborative Approach to Search Result Analysis* [99], and one describing the graph approach discussed in section 9.5.5.3 *Semantic Blossom Graph: A New Approach for Visual Graph Exploration* [100].

Chapter 10 discusses the evaluation of selected dashboards. Evaluations were performed for dashboards discussed in chapter 3, 7 and 9. This chapter provides detailed information about the goals of the evaluation, the methodology and the results for each of the three scenarios separately. All these evaluations have been designed and performed by the author of this thesis. Results of each the evaluations have been published as described above.

Finally, chapter 11 provides a summary of this thesis, discusses lessons learnt during the design, implementation and evaluation phases of the discussed dashboards with regard to the design guidelines in different application domains and derives further recommendations based on gathered experience, as well as provides information about future work.

Chapter 2

Related Work

This chapter discusses related work focusing on the main application domains discussed in this thesis as well as well-known design guidelines for data dashboard design.

Stephen Few defined dashboards as follows [23]:

"A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance."

Dashboards should enable users to investigate KPIs (key performance indicators) relevant to a particular objective or business process [137]. Few classifies dashboards according to their role, which can be strategic, analytical or operational. While strategic dashboards mainly visualise data, which does not change often by using static visualisations, analytical dashboards are characterised by providing additional interaction possibilities within the visualisations such as drilling down. In contrast, operational dashboards enable users to monitor constantly changing data by highlighting arising events [23]. The dashboards discussed in this thesis can be categorised as analytical and operational. In contrast, visualisation systems such as Tableau [124] allow users to create custom dashboards. As the identification and selection of visualisation for data monitoring and analysis requires experience, dashboards based on these visualisation tools, should be designed by experienced dashboard designers.

The first section concentrates on dashboards for project portfolio monitoring. The second section handles dashboards for monitoring production plants as well as analysis of production results afterwards. Tools for analysing sensor data are investigated in more detail in this section. The third section focuses on competitor monitoring and analysis by using search applications. Finally, the fourth section discusses dashboard design guidelines.

2.1 Project Portfolio Monitoring

In order to conduct a project, a project manager is required to keep track of the whole process throughout the project duration. Different steps, depending on the project domain, are suggested to follow in order to overcome potential problems. As the dashboard proposed in this thesis refers to software projects, this section discusses how to support project managers in monitoring their software projects. There are numerous tools supporting project managers in performing their tasks. Some of them only allow monitoring and analysis of single projects, others allow comparing variables of different projects by providing a portfolio view. Therefore, related work is split up into single project monitoring and portfolio monitoring. A summary of related work in this domain has been published by Rauch et al. [98].

2.1.1 Project Monitoring

There are numerous tools available for monitoring different data fields of a project. They allow analysing projects in detail and support project managers when it comes to take decisions. Three examples are ZOHO Projects [140], ViCA [60] and PIVot [111], which will be discussed in the following.

ZOHO Projects [140] is a web-based software allowing to track different aspects of project management, including Gantt charts for showing milestones as well as tasks and their corresponding progresses using different colour codes. This allows project managers to analyse critical paths within the project and take actions accordingly. Furthermore, this tool allows users to manage human resources and monitor their workload. Colour encodings again support users in getting a brief overview. Different views allow tracking the number of open and finished tasks over time and therefore, monitor the progress over the whole project duration. They also provide a bug tracking system as well as a dashboard for planning and performing sprints. Additionally to the desktop application, a mobile application is available, which allows tracking the status all the time. ZOHO Projects provides a huge number of features and views to analyse and monitor all aspects of a project starting from the project set up to generating bills for sending to customers.

Kilponen et al. proposed a Java-based project-monitoring tool called ViCA [60]. Their tool contains several visualisations including Gantt charts, pie charts, bar charts, line charts and stream graphs showing different project variables. As the previous approach, this tool allows to analyse the variables only for one project and therefore, does not provide a portfolio view. A screenshot of the tool is shown in Figure 2.1.

A tool provided by Sharma et al. called PIVot [111] shows different insights compared to the previous tools. Analysis is done based on six categories. First of all, it allows analysing code quality, which is calculated using different variables. This allows the project manager to take actions if the quality decreases. Additionally, it provides information about the quality of component testing measuring the effectiveness of unit tests, the development efficiency and code activity. While the development efficiency is measured

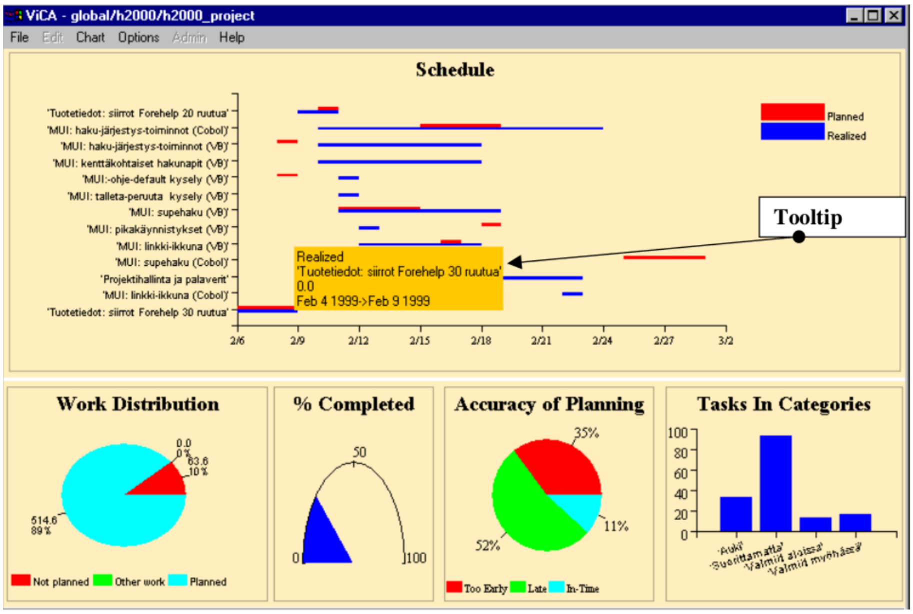


Figure 2.1: ViCA - A tool for visualising project statuses by Pekka Kilponen and Jari Vanhanen [60].

depending on the developers experience and salary, the code activity shows the number of code changes and improvements throughout the project. A team analysis feature allows the project manager to gain deeper insights on the task distribution among developers and makes the process of restructuring responsibilities easier and finally, a composite graph show the effect of code changes on the code quality.

2.1.2 Portfolio Monitoring

In contrast to the previous discussed approaches, tools discussed in the following allow to compare projects more easily as they provide a portfolio view.

An approach discussed by Auer et al. [7] allows a comparison of multiple projects by providing a multi-dimensional scaling method for displaying high-dimensional data in a two or three-dimensional visualisations. Their approach clusters projects according to their similarity in a high dimensional space and plots them on a two-dimensional chart. This approach mainly handles clustering of projects and therefore allows an identification of similar projects, but it does not provide any project details, which enable managers to compare project variables.

In contrast to the previous approach, Hackystat [56] provides a visual component allowing a comparison of variables from different projects. Different project variables are displayed in a table view containing simple visual representations including bar charts, see Figure 2.2. Different colour encodings are used for analysing the results more easily by indicating the health factor of the corresponding property. Properties can either by *healthy*, *unstable* or *critical*. Although the table view provides a quick overview of the status of the portfolio, the provided representation is difficult to read. Therefore, the comparison of variables is not easy.

Project	Members	Coverage	Complexity	Coupling	Churn	CodeIssue	Commit	Build	UnitTest	Size(LOC)	DevTime
hackystat-analysis-dailyprojectdata	2	85.0	2.4	10.9	185.0	N/A	22.0	2.0	160.0	12236.0	0.2
hackystat-analysis-telemetry	2	81.0	3.0	9.4	97.0	N/A	1.0	3.0	324.0	26812.0	0.2
hackystat-sensor-ant	2	4.0	2.9	6.6	N/A	N/A	N/A	2.0	56.0	32500.0	N/A
hackystat-sensor-eclipse	2	N/A	3.1	N/A	N/A	N/A	N/A	2.0	8.0	5587.0	N/A
hackystat-sensor-emacs	2	N/A	N/A	N/A	N/A	N/A	N/A	2.0	N/A	838.0	0.0
hackystat-sensor-example	2	N/A	N/A	N/A	N/A	N/A	N/A	2.0	N/A	2510.0	0.1
hackystat-sensor-shell	2	62.0	2.5	5.9	N/A	N/A	N/A	2.0	40.0	5442.0	N/A
hackystat-sensor-vim	2	N/A	N/A	3.0	N/A	N/A	N/A	2.0	N/A	699.0	N/A
hackystat-sensor-xmldata	2	50.0	2.3	7.2	N/A	N/A	N/A	2.0	88.0	5735.0	N/A
hackystat-sensorbase-postgres	2	90.0	N/A	N/A	N/A	0.0	N/A	N/A	N/A	35.0	0.2
hackystat-sensorbase-simdata	2	97.0	1.9	9.2	439.0	N/A	6.0	3.0	32.0	2536.0	N/A
hackystat-sensorbase-uh	2	85.0	3.0	9.9	995.0	0.0	23.0	3.0	440.0	20862.0	0.5
hackystat-ui-wicket	3	70.0	N/A	11.6	7.0	0.0	4.0	N/A	70.0	N/A	7.4
hackystat-utilities	2	70.0	1.9	5.4	N/A	N/A	N/A	1.0	152.0	5479.0	0.1

Figure 2.2: Hackystat table [56].

Another tool for analysing code quality is provided by SonarQube [122]. Their main focus lies in monitoring code quality and test coverage. The provided overview makes it possible to investigate whether a certain project successfully passed the code check, created warnings or failed. Furthermore, it provides information about the reliability, security vulnerabilities, maintainability, the percentage of code covered by tests, the project size and the used programming language. Depending on the corresponding status, the information is coloured green, yellow, orange or red. The information is arranged in a tabular form where each project property corresponds to a column and each project defines a single line. Filtering for certain properties simplifies the comparison of single projects. Other views enable managers to monitor all issues, allow a definition of rules for the code quality and other configurations. As this approach specifically handles the topic of code quality, it does not include statuses for time, human or financial resources and overall progresses. Connecting this information to a general tool for project monitoring would provide detailed insights, which are often not included.

Cable et al. [9] provide a tree map representation splitting the projects according to their project phase, see Figure 2.3. The project is divided into four phases: *Initiation*, *Planning*, *Execution* and *Close out*. The sizes of the corresponding rectangles refer to the project budget; the colour is used to show user defined project variables, for example performance with regards to the costs. Their approach allows managers to compare three different project variables with each other. While two are fixed, one can be configured by the user to investigate different aspects of the project. Clicking on one of the rectangles within the tree map provides further details to the project manager. The provided colour coding from green (okay) to red (critical) is intuitive and easy to read.

Another approach provided by Zheng [139] also addresses the problem of portfolio management. In comparison to the approach by Cable et al., Zheng tries to visualise more dimensions. Besides using Parallel Coordinates [53], Triple C [3] or Coviance Biplots [63], which have been identified as being suitable for visualising multiple dimensions, Zengh investigated radar charts for visualising different project properties within one visual com-

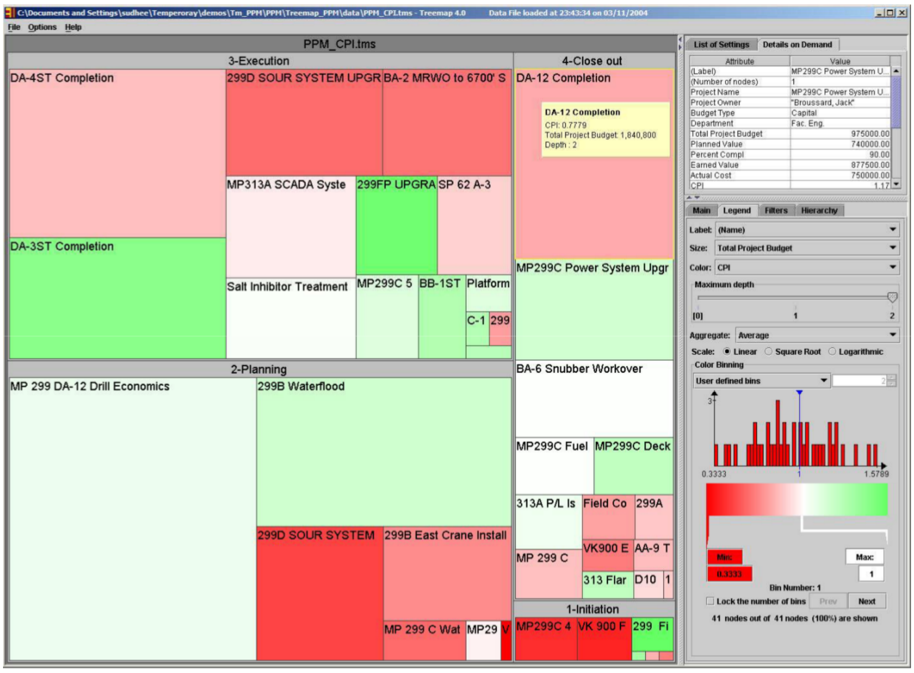


Figure 2.3: Tree maps introduced by Cable et al. [9]

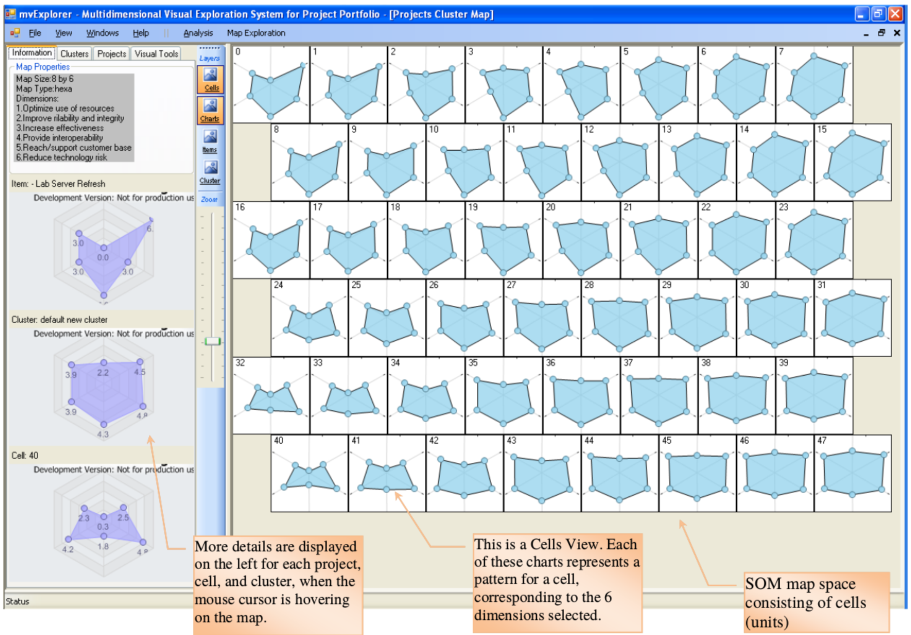


Figure 2.4: SOM layout of projects proposed by Zheng [139]

ponent. Additionally, Zheng grouped the projects according their properties using SOMs (self-organising maps) [62]. Projects containing similar properties are grouped together in a 2D similarity layout. An example for this is shown in Figure 2.4. As projects are positioned according to a grid layout, they can be compared more easily. Visualising too many variables may be confusing as different project variables are only distinguishable by the direction within the radar chart.

The proposed approach in this thesis is similar to Zheng, including several different data variables encoded within one project symbol. Appropriate colour coding, icons and visual components make it possible to easily distinguish between all the variables. Two portfolio views are suggested, one simple matrix layout positioning projects according to two selected data fields as well as a similarity layout, where similar projects are positioned close to each other, similar to the approach proposed by Zheng.

2.2 Monitoring Production Plants

This section discusses related work in the area of dashboards for monitoring production plants as well as for analysing production data. The demand for increasing productivity in highly automated production environments with high competition and costs has led to the need for monitoring and analysing production data and thus, avoiding all influencing factors, which decrease productivity. Influencing factors can causes rejects during production, arising bottlenecks and other problems. In order to enable managers, production workers and other decision makers to decrease these negative influencing factors, additional information is required. During the last years, dashboards connected to the production plants, have been implemented and considered as being suitable for supporting the decision processes. Although there are many general-purpose software tools available, which allow users to create a dashboard and display information, many manufacturers prefer having their own solution for handling their specific problems and demands.

During production, an important factor is to see negative impacts in the dashboard immediately after arising, to enable users to stop and adapt the production immediately or at least within a short period of time to decrease rejects. Therefore, providing a dashboard showing the current production, not only analysing results of production afterwards, is a primary requirement. This section is subdivided into dashboards, which allow analysing the production data as well as monitoring the production process itself. There are systems, which combine both approaches but often tools are designed for one use case while the other one is not supported sufficiently. Furthermore, dashboards allowing an analysis of sensor data are investigated more in-depth as production plants usually have many sensors installed generating huge amounts of data every day.

2.2.1 Tools for Analysing Production Data

Many companies analyse their production data using data analysis tools allowing them to perform different operations to gain insights. These tools often do not provide any visualisations to support analysis tasks. Companies often do not employ specific tools but use standard tools such as Excel [69] or simple programming languages as R [132] or Python [93] to analyse their data. Those companies, which analyse their data using visualisations, commonly use general purpose visualisation tools such as Tableau [124], Power

BI [71] or Qlik [95], which may involve significant work to configure or customise.

In the following some data analysis tools are discussed in more detail including those, which have received a high rating by Gartner in the field of data science and machine learning 2018 [29]. These are KNIME [61], Alteryx [2], SAS [109], RapidMiner [96] and H2O.ai [42]. Additionally, tools for visual analysis are investigated, which are considered by Gartner 2019 [28] to be the market leaders in analytics and business intelligence platforms. These include Power BI [71], Tableau [124], Qlik [95] and ThoughtSpot [133].

One of the most common tools used for analysing numerical data is Excel [69]. Excel provides spreadsheets allowing users to create dashboards based on tabular data. These dashboards can be reused later for displaying the latest production values. In order to simplify this, Excel provides the possibility to connect different data sources. These include simple files such as CSV (comma-separated value) files, databases such as Oracle database [89], cloud platforms such as Microsoft Azure [67], online services including Microsoft Exchange [70] or other sources including external web pages. Excel already provides several visualisations, such as bar charts, stacked bar charts, pie charts, bubble charts, line charts, geographical maps and others. Although the provided visualisations allow displaying different insights, the number of visual representations is limited to those provided by Microsoft Excel. Users have to arrange charts on the sheets themselves and connect visualisations to data manually, which requires skills and extra time. Furthermore, different requirements lead to the generation of different dashboards, which can increase the number of Excel sheets tremendously and may lead to a loss of context when analysing multiple sheets at the same time. Although Excel also provides the possibility to generate mobile sheets, each device requires a licence for Microsoft Excel. Excel allows users to analyse the data afterwards and is not suitable for monitoring current production. Furthermore, it does not provide any interaction possibilities within the visualisations, which is often required for data analysis.

Another method often used for analysing production data is the statistical language R [132]. In contrast to Excel, generating charts using R requires programming knowledge, as it does not provide a user interface generating visualisations by mouse interactions. R supports a wide variety of visualisations for statistical data due to the large number of libraries available for R. This includes bar charts, pie charts, box plots, radar charts, density plots, stacked area charts, correlation charts and many others for analysing the provided dataset in detail. Visualisations generated with R are per default not interactive but dashboards containing several visualisations can be generated using for example Shiny [104] and ggplot2 [43] libraries. Thus, further effort is required to make them interactive. The disadvantage of R is, that dashboards have to be implemented or extended for each new data type separately.

In contrast to R, MATLAB [130] provides an environment, which allows users to generate dashboards using the provided user interface. Code is automatically generated and can be modified by its users. Equally to R [132], MATLAB allows to generate many different plots easily. Different sources can be connected ranging from simple files up to databases. Users can pre-process their data by filtering, smoothing, interpolation, convolution or Fast Fourier transformation. Generated dashboards or models, as they are called in MATLAB, allow re-use of previous configurations. Additionally, plots or even whole reports can be easily exchanged or exported. In order to provide an interactive dashboard, user interactions need to be implemented in the generated code to handle interactions accordingly.

KNIME [61] provides an interactive dashboard allowing users to select visual components and connect them via drag and drop. This allows users to generate complex multiple coordinated views for analysing their data. KNIME supports a set of representation possibilities including network visualisations, sunburst charts, line charts, scatter plots, box plots, parallel coordinates, geographical visualisations and other charts. It can be extended to display visualisations provided by JFreeChart [87], R [132], Python [93] and MATLAB [130]. Guided analytics in terms of providing additional information about configuration possibilities and steps for applying machine learning algorithms is included.

Another tool for designing and executing data processing pipelines is RapidMiner [96]. Single process steps can be added to the processing pipeline by drag and drop in the user interface. Process steps can range from simple calculations, for example filtering data, to the application of complex machine learning calculations. Recommendations of possible next steps are provided in a list supporting users to design the processing pipeline. RapidMiner allows extending the data processing using Python [93], R [132] or other scripts. Additionally, simple visualisations can be added to the processing pipeline to gain deeper insights after data processing. In order to enable users to investigate their data using visualisations, RapidMiner provides an API (application programming interface) connecting different visualisation tools including Qlik [95] and Tableau [124]. Making usage of web service-based integrations, a bi-directional communication between both tools is possible, meaning that data is initially sent from RapidMiner to the visualisation tool. Interactions within the visualisation tools trigger calculations within RapidMiner and return results to be visualised again.

Alteryx [2] also allows configuring process steps in order to analyse the data and identify hidden information. Similar to RapidMiner [96] many predefined algorithms can be applied to process data. The tool does not provide many visualisations for data analysis but can be connected to Tableau [124] in order to enable users to visually inspect their data using WDC's (web data connector).

In contrast to the previous two approaches, SAS [109] is more focused on visual analytics than on data processing. Its user interface provides a set of simple visualisations allowing users to investigate their data as well as generate reports. Dashboards can be created including several visualisations, for example geographical maps, bar charts, pie charts, line charts, bubble charts, tree maps and tag clouds. Furthermore, users can create their own visualisations and integrate them into the tool.

H2O.ai [42] provides a dashboard supporting automatically generated visualisations. Depending on the selected data content, the tool generates a set of visualisations. This kind of guidance supports the user in analysing their data. Further insights can be provided by highlighting anomalies within the visualisation. The auto-generated visualisations can only be investigated separately as there is no view coordination between different visualisations. Additionally, a dashboard can be created to show the results after applying different algorithms.

The described tools and approaches mentioned above are used for data analysis. Some of them also contain visualisations for inspection but it is not their main focus. The tools described in the following are more focused on visualisations than on statistical calculations. For each discussed tool, different features are investigated, such as: support of different data sources including databases, extensibility in terms of visualisations, extensibility of algorithmic calculations, availability of alerting and notification features, features for sharing insights and collaborative analysis as well as features providing guidance during dashboard generation and data analysis.

Power BI [71] allows users to connect many different sources including Excel [69], Salesforce [107], Google Analytics [33] or IOT (internet of things) devices. Tiles containing different visualisations or data fields allow users to configure a dashboard as well as a mobile view depending on their needs. Dashboards can be easily shared with other users enabling them to investigate the same data. There are different options to share insights, for example publishing the application by providing read access. Dashboards can be shared with groups or single individuals by integrating the dashboard in an I-frame, which requires authentication, by publishing it to the web, which requires no authentication and therefore provides no security, or in an embedded way having full control of all functionalities. Power BI provides an API allowing users to generate and add visualisations easily. Visualisations can be either implemented in TypeScript [74] using SVG (scalable vector graphics), canvas or other technologies or by implementing them in R. Alerting features allow users to define constraints for receiving alerts within the dashboard but also via e-mail on demand. Currently, there is no user guidance within the dashboard to provide suggestions on visualisations or interaction possibilities for data analysis. R [132] scripts can be included in Power BI for data processing before visualising. Data selection behaviour in one visualisation can be specified for other visualisation, for example pre-filtering of the selected category.

Tableau [124] can be used for visual analysis of data from an enormous number of different sources starting from simple files to databases, allowing continuous data updates over time. Additionally to the supported data sources, Tableau supports WDC's, which allow to connect to any existing web source in order to import data to Tableau. In terms of proposed visualisations, Tableau provides a large variety of different visualisations, but it is not easily extendible. Extending the provided dashboard by adding additional visualisations can be performed by publishing visualisations or dashboards on a server and including them in Tableau using a web page object. In order to make full use of the new visualisations, interactions need to be coordinated with existing visualisations. The Tableau JavaScript API allows users to integrate Tableau visualisations to external user interfaces. Additionally, Tableau allows connections to MATLAB [130], Python [93] and R [132] in order to apply algorithms on selected datasets, either before visualising or triggered by user interactions. Tableau's *Show Me* functionality activates only meaningful visualisations depending on the selection of available fields. In terms of further support on dashboard generation, Tableau does not provide any further guidance. Tableau allows users to configure alerts based on numerical values. This means that users can specify constraints, highlighting those data points, which meet the constraints. Viztalkjs [8] is a small JavaScript [138] library using TogetherJS [78], which can be included to Tableau's server main page and enables users to collaboratively analyse their data.

Qlik [95] also has its focus on visual data analysis. Like Tableau, it also supports a large number of data sources including files, databases and spreadsheets. A complete list of possible data sources is available in [94]. In order to connect different data sources for visualising, Qlik suggests potential fields suitable for connections, which can be reviewed and modified by users. Considering the visualisations available in Qlik, it already provides many visualisations and allows to further extend the set of visualisations by copying a provided template using D3 [75] (data-driven documents) and modifying it according to the users demands. In order to generate a multiple view coordinated dashboard, users can either first select the desired visualisation and add data fields afterwards or initially select the fields and then generate visualisations. The second approach automatically suggests visualisations depending on the selected data fields. Adding additional fields to a visualisation can change the visualisation, unless users switch off this feature enabling them to have full control over all visualisation properties. Additionally, users can take snapshots of single visualisations and embed them directly in the presentation mode included in Qlik. The responsive user interface allows analysing the data using different devices. Dashboards can be shared easily with other users by providing a dashboard link, where previous interactions are applied when opening the dashboard. Furthermore, users can also collaboratively analyse their data while seeing other users' interactions on their screens. Similar to other approaches, Qlik also provides the possibility to perform advanced analytics by supporting external compute servers. Examples therefore are R and Python, performing complex analytical operations connected to the dashboard. Qlik also supports alerting and notification management. While alert messages can be configured easily within the dashboard by defining simple constraints, while creating notification alerts for e-mails is more complex as conditions need to be formulated separately.

In comparison to previous approaches ThoughtSpot [133] follows a different approach. The number of data sources is smaller than for previously discussed tools but it still supports multiple sources. Visualisations are automatically generated depending on the text input query. Provided queries must match the underlying database field names to generate meaningful charts. As this approach is considered to be suitable for non-experts, the number of possible modifications possible within the dashboard is quite limited. Additionally, a speech recognition tool allows formulating queries by users. Created dashboards can be shared with other users allowing them to change the query of single visualisations. Additionally, users can rate visualisations. Based on this feedback, the tool tries to learn and suggests more accurate visualisations when typing in a query. The number of visualisations provided within ThoughtSpot is limited to those provided within the system, with no documentation available to provide any information about extensibility. ThoughtSpot allows advanced users to run their own self-configured R scripts [132] to execute additional data analysis algorithms for certain visualisations. Additionally, it contains a set of alerts within the dashboard, which raise attention if problems occur. These are related to the system health status but not to data constraints and cannot be extended or modified by users.

Visualisation tools discussed in this section represent powerful means for visually inspecting and analysing data. The provided user interfaces are configured for ease of use, where some provide automatic data manipulation before visualising data, for example, data aggregations. However, this lack of transparency when defining data manipulation, may lead to confusions during analysing the data.

2.2.2 Tools for Monitoring the Production Processes

Within manufacturing companies many tools are used for planning and monitoring, which includes ERP (enterprise resource planning) software but also MES (manufacturing execution system) software used for monitoring and controlling the manufacturing process. Companies use different software systems for monitoring their production, ranging from general approaches to specific MES software directly connected to the production plants. In contrast to MES software, generally applicable tools might not cover all aspects of production and might not outline problems immediately as they are not connected to the production process. Therefore, this section concentrates on MES tools. There are numerous MES tools available. According to Gartner [30] the MES systems described in the following are the leaders in 2018 and will be investigated in more detail. For some of these tools only little information about visualisations available for monitoring can be provided, which is due to a lack of information available on the web.

Honeywell Connected Plant (HCP) [45, 46], one of the leading manufactures of MES software, contains a dashboard with different visualisations, including bar charts, pie charts, line charts and others. Furthermore, for time series data HCP provides pattern search, value search, the combination of both approaches as well as filtering. HCP also provides statistical analysis such as PCA (principal component analysis), PLS (partial least square) regression, kernel regression and neuronal networks. Finally, HCP also provides big data analytics features like feature selection/extraction, random forests, SVM (support vector machine) and others for analysing their data.

Siemens PLM (product lifecycle management) Software [116] contains a huge variety of components including a MES tool. The MES tool consists of different components for production monitoring or planning and scheduling. A complete list of all features is provided by Siemens in [114]. The historian component allows users to investigate aggregated historical and live data using the optimal level of detail. Different calculations can be applied on the production data, while a dashboard shows the most important KPIs. Another component allows to manage, develop and configure product specifications. The line-monitoring tool allows live monitoring of the production process, the advanced planning and scheduling component allows users to plan production and ensure that components are available when required. Other components allow analysing electronic components, supporting the research and development process, supporting coordination of production across multiple production plants as well as supporting the data analysis process by providing visualisations.

SAP Manufacturing Execution [108], one of the world leading ERP software vendors, also includes a MES system for real time monitoring and controlling of production plants, tracking products during production, optimization of the production processes and quality improvement, and identification of defect process steps. Together with the SAP Manufacturing Integration and Intelligence System it allows synchronising information between different area of responsibility as well as optimising the production process by providing new insights of the production.

Dassault Systems [14] also provides a variety of software tools including a MES software called Delmia Apriso [15]. It allows to globally connect information for better coordinating operations, production processes, warehouse management, quality assurance, maintenance and real time analytics of the production flow. Additionally, the tool includes management features for process building, reporting and analytics features [16]. Visual components for analysing and controlling different processes are also available.

Rockwell Automation (RA) [103] provides a MES system but also an analytics platform enabling users to investigate data from various sources, use machine learning algorithms for data processing and tools for collaborative data analysis [102]. RA provides specific solutions for different industries including automotive, consumer packaged goods, life science and metal manufacturing. Additionally, RA provides different approaches for certain challenges including helping product management to manage and control orders, supporting performance management to reduce downtime, unproductive time and rejects in production processes, or quality management by ensuring that products are available in time with a specified quality level.

Most of these market leaders for MES software are well-known global companies with a large variety of software tools which enable customers to connect different tools and therefore benefit from analysing data from different sources during the whole planning and production process. Visual components support these processes in order to enable employees to take decisions. As the information on the discussed products is limited without having access and trying their functionality, detailed information on the proposed MES systems cannot be provided within this thesis.

The approaches proposed in this thesis describe specific approaches for distinct companies, which could be adapted for a more general usage. While the proposed dashboards cannot match large software vendors and their products in sheer function coverage, they are suitable for the specific companies to analyse their data and reduce complexity compared to using products from large software vendors.

2.2.3 Tools for Analysing Sensor Data

In order to monitor production processes, usually many sensors are installed in production plants, producing an enormous amount of data every day. These measurements have always been important in supporting the decision process, but are often not considered as the amount of data is comparable large and therefore requires extensive further processing before being visualised. Importance of sensor data even more increases due to the rise of IOT. Since then, the amount of sensor data for further processing and analysis has considerably grown and thereby the need for tools to process and visualise them. Thus, there are several frameworks for analysing sensor data, with some providing an integrated general-purpose solution for storing and displaying sensor data, while most concentrate on very specific functionality. Providing a complete solution including storing, processing and analysis of sensor data appears quite rare.

Available frameworks are investigated according to the features described in the following. The most important one is whether the tool allows displaying large-scale time series data and enables users to view full details when requesting for smaller time ranges. Only if this criterion is fulfilled, other features were investigated. First, the generation of user-defined annotations is investigated, including whether the tool can propose annotations automatically e.g. based on trained machine learning models and enable users to review these automatically generated annotations. Another feature, which is being investigated is whether it is possible to search for annotations. Finally, filtering of time series data, alerting and attention management as well as signal search are investigated. There are many databases for storing and managing different types of data, some of them are even suitable for big data infrastructure. When it comes to time series data, one of the most promising databases is InfluxDB [51]. Often it is used in combination with Grafana [37] to gain insights from the stored data. Data can be stored hierarchically by using different data granularities. Depending on the requested time range, data can be retrieved using a lower or higher granularity.

A framework providing a dashboard for visually inspecting and analysing large-scale time series data is the open source framework Grafana [37]. Grafana supports several data sources including Elastic Search [22], InfluxDB [51], MySQL [88] and PostgresSQL [131]. A complete list containing additional information regarding the usage is provided by Grafana [36]. Grafana is implemented in Go [35] and provides an API for adding new plug-ins using TypeScript [74]. There are already many plug-ins provided by Grafana and the corresponding community including visualisations, for example line charts, heat maps, pie charts, bar charts, lists, tables or bubble charts [40]. The provided visualisations are completely configurable according to the users' needs. Grafana provides an alert feature allowing users to define conditions. As soon as a condition is met, users receive notifications depending on their configuration [38]. Furthermore, Grafana provides different features for dashboard sharing [41], allows data annotation and querying for keyword-based annotations [39]. Although it allows searching for annotations, searching for signals as well as filtering within the time-series data is not supported.

MindSphere [115] is an IOT platform from Siemens. It contains a couple of applications, including the Visual Explorer [119]. The Visual Explorer is based on Tableau [124] and allows users to connect different data sources for visualising. As data is usually aggregated before displaying in Tableau, it is not well suitable for inspecting large-scale time series data. Another application provided by Siemens is the Fleet Manager for Machine Tools [113] containing an add-on for analysing time series data called Visual Analyzer [117]. The tool allows users to drill down, search or filter data and apply root-cause analysis algorithms on previous as well as on real-time data. A detailed list of features including KPI calculation, anomaly detection, event analytics, signal calculation, signal validation and trend prediction is provided in [118]. This tool provides many features for monitoring production and enables users to take decisions very quickly.

IBM Watson [48] allows users to investigate their data including time series. Users can add visualisation libraries and analyse data using Jupyter Notebook [92]. Therefore, analysing data using the tool requires programming knowledge and some configuration for set-up. Watson Analytics [49] furthermore provides a set of visualisations to analyse sensor data, for example line charts, correlation charts and bar charts. Depending on the requests, visualisations are suggested. The limitations of data analysis and visualisation correlate with those of Jupyter Notebook and the underlying API.

IBM Streaming Analytics [47] allows processing and analysing sensor data. A pipeline containing different operations can be created, supporting result analysis. Visualisations can be created to display the results for all process steps. The provided dashboard is suitable for investigating different insights but not for analysing large-scale time series data. Therefore, further criteria are not investigated for this tool.

Finally, Microsoft provides the Azure Time Series Insights [73] tool. It is also suitable for IOT applications and allows analysing large-scale time series data. On selecting a time range, the system automatically returns a certain number of data points meaning that for large time ranges, the number of data points is reduced by interpolation. Zooming in allows retrieving a higher resolution of the underlying data. Selection of certain time ranges allows users to investigate the original data shown in a table. The dashboard only allows displaying one measurement at a time but data can be split by selected string fields. Multiple lines can be displayed together by using one axis, separately or together by using multiple axes, one for each line. Time series data can be filtered by value constraints. Further processing to reduce noise is also possible. Filters can be complex as the provided input field allows adding multiple filter criteria. Tooltips within the line charts show current values but also minimum, maximum and average values for a certain line. Additionally, a heat map representation is provided over time, allowing to investigate different measures. The tool does not provide any functionality for data annotation, annotation learning and searching or signal search.

Other business intelligence tools such as Tableau [124] or Power BI [71] also allow analysing sensor data, but have no specific features when it comes to large-scale time series data. Therefore, they are not discussed here. The proposed solutions in this section show, that there is a lack of tools when it comes to analyse large-scale time series data and performing different operations like user defined annotations, learning annotations and other features.

2.3 Competitor Monitoring and Analysis

As the number of companies offering products is constantly rising, the competition among those offering similar products is exacerbating. In order to not loose in these competitions, companies need to identify and observe their competitors and take decisions if required. Search engines have been identified as being suitable for identifying and monitoring competitors as the amount of available information on the web has tremendously grown over the last years and companies usually provide information for existing and potential new customers on their websites. Thus, finding and presenting relevant information in a visually appealing way has become a big challenge. This section discusses different search applications. Additionally, a lot of research has been performed in the area of visual supported search interfaces. Therefore, research in this field is also investigated. A summary of this section has been published by Rauch et al. [99].

In order to enable users to retrieve the results they are interested in, search queries need to be formulated in a way that search engines are able to return the required results. Marchionini [66] in general distinguishes between two types of search, information lookup and exploratory search. While the characteristics of information lookup is to simply provide a set of results depending on the query as standard search engines do, exploratory search requires more interactions. In general, exploratory search can be further divided into learning and investigation tasks. In contrast to information lookup, learning tasks require cognitive processing and interpretation of the data. Investigation tasks however require more interactions over a longer period of time as it includes analysis, synthesis and evaluation of data. While users performing information lookup are only interested in the best results, investigative users try to gain further insights. As identifying and monitoring competitors can be categorised as exploratory search, the provided dashboard needs to offer additional features for data analysis.

2.3.1 Research on Search Applications

Additionally to all products described in the next section, a lot of research is being done in search applications, which will be discussed in the following.

As visual representations are considered to be read and interpreted faster compared to text, Hearst [44] identified four different visualisation categories for search interfaces:

- Showing the result set by using Boolean syntax. This set can be represented by Venn diagrams.
- Visualising query terms within retrieval result. This includes highlighting the query within the snippet text but also providing additional visualisations such as tile bars indicating the frequency and position within the text.
- Visualising relationships among words and documents using simple node-link diagrams but also information landscape visualisations using a distance measure for displaying two results.
- Visualisations for text mining allow users to investigate, which words or letters are followed by others. Typical visualisations for this category are word trees and stream graphs.

Standard search engines often provide filters for narrowing down the number of results. The provided data for filtering is often presented by semantic information extracted from the search results. Usually this semantic information is what different search results have in common. A tool for visualising semantic information between different datasets is called SemaVis [82]. The provided approach does not concentrate on displaying search results but on the semantic information within different results and reveals existing relations using visual approaches such as node-link diagrams, line charts or geographical visualisations. Depending on users' profiles and previous interactions, different visualisations are provided to the user to match the user's preferences.

Another framework for investigating semantic relationships has been proposed by Ding et al. [19]. As the provided tool is restricted to RDF (resource description framework) [136] triples consisting of subject, predicate and object, it is not generally applicable. Users can select any semantic property within the system and investigate its relationships using a table view, a time line, a node-link diagram, a geographical map, a tile view or a simple list.

Foo et al. [24] proposes a desktop search engine extended by different visualisations. Besides a list of search results, their approach includes a tree representation showing all search results based on their folder structure, a node-link diagram allowing to analyse which results were retrieved by certain query terms, a bubble chart enabling users to identify search results with higher relevance, a tiles view enabling users e.g. to identify different file types and a word cloud showing most relevant tags within the search results. Each of them provides only limited interaction possibilities. Evaluation results showed that the provided visualisations support users in search refinement.

Arnaud et al. [5] provide a tool called CoViz for collaboratively organising and collecting datasets as well as searching within those datasets. They initially assign all documents according to predefined categories namely tags, people, collections, quality ratings, free text comments and document types used for search refinement. Documents are displayed using a node-link diagram where the connecting edges indicate the similarity of the documents. Additionally, documents can be automatically clustered by the system or interactively clustered by the user.

Clarkson et al. [10] provided a search interface including a tree map-based visualisation. The provided tree map is clustered by topics. Extracted meta data of search results are displayed as rectangles within the different topic areas, the corresponding colours indicate different document types. Documents can contain multiple meta data and therefore can be assigned to multiple topics. Each meta data of a result element corresponds to one rectangle within the visualisation. Mouse events highlight all related rectangles on inspection. The tree map visualisation provides the possibility to refine faceted meta data and therefore, should provide a better preview of the result set for a certain query. An evaluation did not reveal statistically significant differences between using the tree map representation compared to inspecting results without the visualisation. However, tendencies towards the visual approach were recognized. Figure 2.5 shows the tree map approach suggested by Clarkson et al.

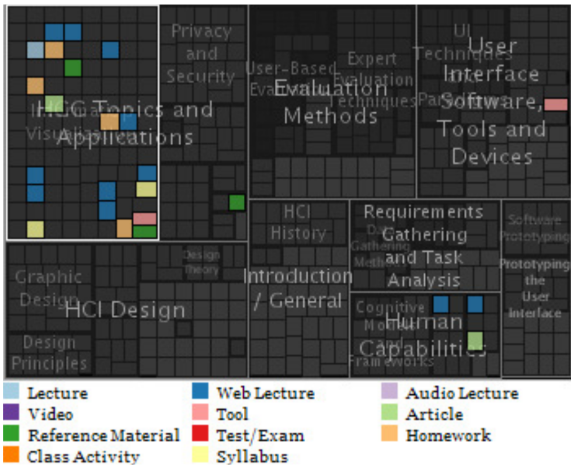


Figure 2.5: Topical tree map-based visualisation provided by Clarkson et al. [10].

Another approach for visualising search results has been provided by Nguyen et al. [85]. They use the solar system metaphor for visualising the result set of a certain search query, see Figure 2.6. The query is located in the centre of the solar system, search results are located around the query, whereby the distance to the centre indicates the relevance of each search result. Automatic clustering ensures that search results, which are close to each other, have a higher similarity than those, which are far away from each other. When defining a certain search result as the new centre, the system computes the similarity to other search results. Highlighting selected search results help users in observing the change of position if the centre of interest changes. The speed of elements movements when the centre changes, indicate the relevance to the selected centre.

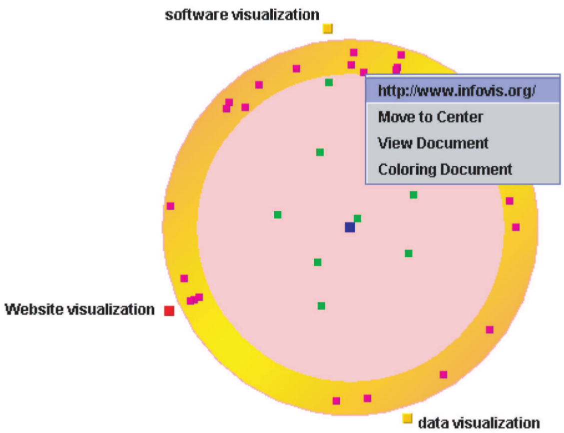


Figure 2.6: Solar system metaphor to indicate relevance of search results with respect to the provided query proposed by Nguyen et al. [85].

Due to the rising amount of sensor data available in our daily lives, Paparrizos et al. [90] proposed a search interface enabling user to search for sensor meta data. Different visualisations are provided depending on the retrieved meta data. This includes simple pie charts, bar charts and tag clouds but also more complex representations as geographical visualisations and graph approaches. They allow to investigate relationships and similarities in meta data but also showing locations which were identified in the result set. Figure 2.7 provides an example for location based meta data including a tag cloud as well as a graph representation.

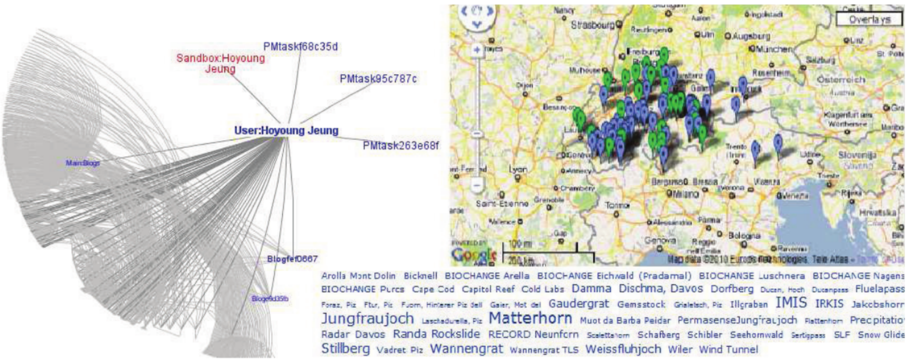


Figure 2.7: Signal meta data search result visualisation proposed by Paparrizos et al. [90].

Kienreich et al. [59] proposed ApaLabs, a web-based platform for retrieving and analysing news articles provided by the Austrian Press Agency. Their tool provides several visualisations supporting users in analysing their search results. A geographical map shows the distribution of search results in Austria. Users can refine their search by clicking on certain regions. A tag cloud provides further information on most frequent terms within the result set and allows to explore selected topics. A visualisation of the Austrian National Council enables users to identify, which delegates of the Austrian parliament have been mentioned within the search result. Each representative of the parliament is associated to a certain seat. The persons mentioned within the search result are coloured according to their affiliation of their political party. Depending on the number of occurrences, the representing figures are displayed smaller or larger. Mouse interactions provide further possibilities for search refinement. Additionally, a Brockhaus [83] lookup feature allows to gain further insights on relevant keywords automatically extracted from the search results. Figure 2.8 shows the visualisations for news article search.

Figure 2.8: News article search including a geographical map, a parliament seat visualisation and a tag cloud proposed by Kienreich et al. [59].

Sabol et al. [105] provide an earlier version of the Knowminer Search Interface focusing on visual analysis of meta data and their correlations to each other while the approach discussed in this thesis focuses on search results with the ability to analyse and refine the search by providing different interactive visualisations. The visualisations proposed by Sabol et al. include an information landscape, a stream graph and a tree representation of clustering results according to their topical affiliation (see Figure 2.9). Topical clustering has been performed by using highly scalable, incremental clustering approach for generating a hierarchical geometry used for visualisation [79]. This approach also enables users to analyse topical-temporal relationships via a stream graph. Additionally, the clustering visualisation provides further insights on meta data correlations with topics.

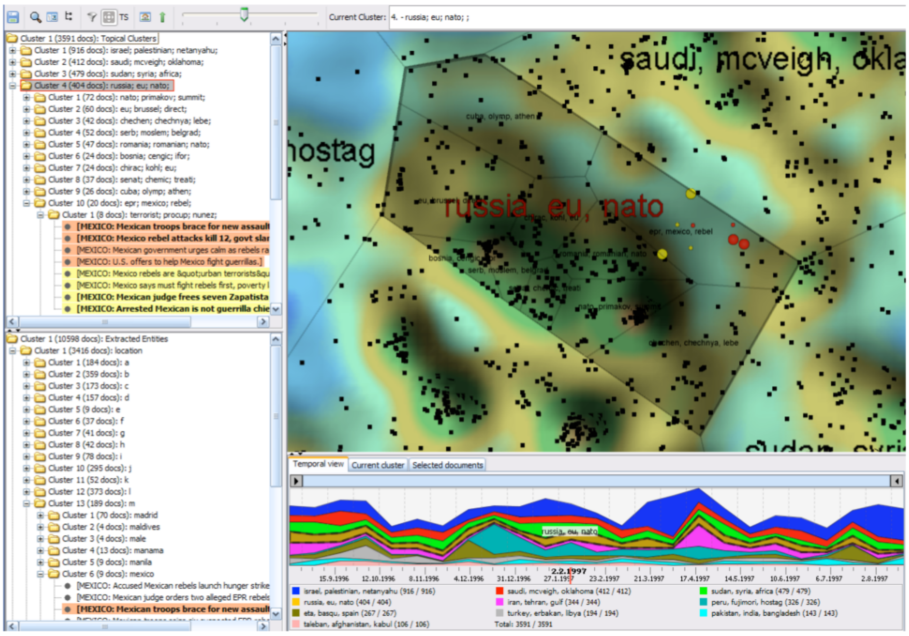


Figure 2.9: Visual analysis of topics meta data and trends based on text search results proposed by Sabol et al. [105].

2.3.2 Search Applications

According to Gartner's Magic Quadrant [27] the market leaders for insight engines in 2018 are Mindbreeze InSpire [76], Coveo [11], Lucidworks [65], Sinequa [120] and Microsoft SharePoint Search [72].

Mindbreeze InSpire [76] is a general applicable tool for searching, providing a faceted search interface [44]. Their responsive design is also suitable for mobile devices. Search results provide detailed information including snippet text as well as a preview image. Additionally, they provide a portfolio view for collecting documents of interest for further processing. The selection of displayed facet groups is based on configurations defined by each user. An editor allows to configure the whole dashboard in order to search depending on the users demands easily. Results can be displayed as list, as image gallery or geographical map. Additionally, filters can be displayed as pie charts, bar charts or line charts. The possibility to configure the dashboard based on an existing set of widgets with simple drag and drop interactions allows all users to defined their preferred search interface to access relevant information depending on their needs. Currently visual components and interaction possibilities are limited. Therefore, the possibility for performing investigation and learning tasks are limited but the provided solution is well suitable for information lookup. A management dashboard allows monitoring the search application itself including availability, search terms as well as the indexing of documents.

Coveo [11] also provides a general search application allowing users to customise their own search page including topics for search results as well as faceted filters. All fields within the search can be configured by the user to match their requirements. Coveo interface has three main areas, the search field, the filtering area and the search results. Each of these areas can be configured independently. Besides standard check boxes used for filter selection, they also support sliders for time and value ranges. The number of connectors to different data sources is quite diverse including databases, web pages, e-mails, social media and others. Additionally, they provide special solutions for integrating their product in Salesforce [107], Sitecore [121], Microsoft Dynamics [68] and ServiceNow [110]. Coveo does not support any visualisations for result analysis and refinement, therefore this tool is well suitable for information lookup but not for learning and investigation tasks. Coveo learns from user interactions. Depending on search queries and users' interactions the tool tries to learn which results are more relevant.

In contrast to previous approaches, Lucidworks [65] supports several visualisations for displaying search results [64]. This include bar charts, geographic maps showing clustering results in overlays or highlighting countries within the map, heat maps, histograms, time-series visualisations, scatter plots, sunbursts and tag clouds for analysing search results. The provided dashboard can be configured depending on user demands and can be used for monitoring selected topics. The system also learns from users' interactions and recommends search results which might be of interest. Additionally to the search results, the dashboard allows to reveal hidden trends and insights, spots relationships and correlations, as well to show outliers and anomalies within the dataset. Features including query suggestions, boosting of results as well as user ratings are also available in the dashboard. The latest version also provides connectors for Tableau [124], Power BI [71] and Apache Zeppelin [128] enabling the user to analyse their data using these powerful visualisation tools. Thus, in contrast to previous approaches, this search interface supports learning and investigation tasks.

Sinequa [120] also provides a search engine allowing users to analyse data including websites. Their user interface proposes query suggestions as well as different filter opportunities, which are positioned on the left and right side. Search results are placed in the centre of the web page. A tag cloud containing dominant topics depending on the result set, a cluster facet generated using natural language processing as well as simple pie charts for filtering other extracted entities are available in the dashboard. The cluster facet does not contain any labels but can be investigated using a heat map allowing different interactions. The provided user interface can be configured according to the users demands.

Sinequa allows, besides information lookup, performing investigation and learning tasks by using additional visual representations, such as line charts or pie charts.

In contrast to previous approaches, Microsoft SharePoint Search [72] is a tool for searching and retrieving company data located in SharePoint while web content is not included. The provided search interface is mainly suitable for fast information lookup including faceted filters. Only simple visual representations are included for providing an overview of the result set and selecting time ranges of document modification dates. Word documents as well as Power Point presentations can be modified within the preview page if an Office web application server is available. Learning and investigation tasks cannot be performed using this tool, as interaction possibilities are quite limited. As with previously described approaches, the provided user interface can also be configured according to users demands.

Netvibes by Dassault System [13] provides a dashboard for investigating data including web content. The dashboard can be configured on user demands for processing real-time data, including user-defined constraints, sentiment analysis and displaying information flows. Additionally, it provides automated reporting and device scalability. In order to display data and analyse web content, data is aggregated and displayed in several visualisations including line charts, bar charts, tag clouds, stream graphs and geographical maps to provide deeper insights to the analysed data. The dashboard allows performing learning and investigation tasks by using the different dashboard components whereas it is less suitable for simple information lookup.

IBM's Watson Explorer [50] is a tool for searching and analysing content. It provides several views and visualisations for analysing search results and allows performing learning and investigation tasks by refining search results. This tool is less suitable for information lookup since it does not provide result list. It can analyse structured as well as unstructured data. Watson can extract nouns, verbs, adjectives, adverbs and other word types and displays them in a facet view. Ontologies can be used for creating hierarchies. The dashboard view shows most frequent terms in tag clouds, trends over time using line charts and other charts, depending on the configuration, which can be used for search refinement. One part of the Watson Explorer is the Content Miner. It provides a user interface, which allows users to narrow down the result set by navigating along a graph supported by guidance features. Options for next steps are provided within the graph view. Suggestions for next steps can be analysing trends and anomalies or showing documents. Simple visualisations as line charts and bar charts support explorative analysis of extracted entities.

The lists of commercial tools can be continued further including Attivio [6] and others. As the above-mentioned tools are the market leaders outlining the direction of dashboards with integrated search functionalities quite well over the last few years, further tools are not described as they are expected to contain less or similar functionalities.

2.4 Guidelines for Dashboard Design

Dashboard functionalities can be described by defining a set of use cases. This supports all involved persons to create a common understanding on the underlying requirements. Before creating first dashboard mock-ups, all requirements need to be investigated in detail. Ranking all requirements according to their priority supports the design process by concentrating on most important features. Then, in order to design a dashboard several guidelines must be considered.

When designing a dashboard, which fulfils all requirements and enables users to easily perform their analysis task, different aspects need to be considered. Although there are many different approaches for defining guidelines for dashboard design, these are not discussed in this thesis. Instead, guidelines discussed in the following are considered to be the most important ones for the author of this thesis [17, 84, 125, 58]. These guidelines can mainly be derived from the common mistakes in dashboard design discussed by Few [23].

Guideline 1: Showing the needed information [23, 84, 58]

The first guideline requires ensuring that the information needed by users is displayed within a dashboard, enabling analysts to take decisions accordingly. In order to ensure this, a detailed requirement analysis has to be conducted, available data needs to be investigated ensuring that the data contains all relevant information for visualising.

Guideline 2: Choose suitable visualisations [23, 17, 84, 125, 58]

In order to outline most important data fields, the most suitable representation methods need to be selected. Recommendations, available in different sources such as [80], support this process. Often, it is also recommended to investigate which visualisations have been used before and users are therefore familiar with.

Guideline 3: Keep the dashboard simple [23, 17, 84, 58]

Overloading a dashboard containing many visualisations attempting to provide too many insights, does not support the analysis process as important data cannot be distinguished from irrelevant information. This complicates the analysis process and requires much more time and effort from users. Thus, designing a simple dashboard representing only really relevant data in suitable visualisations supports the analysis process.

Guideline 4: Reduce clutter within the visualisations [23, 84, 58]

Visualisations should only contain most important information, which are required for analysis. Important information requires to be outlined while hiding less important information and thus reducing clutter. Most important visual representations of data should adhere to the principles of readability and clarity, avoiding overdraw and redundant items.

Guideline 5: Providing suitable interactions with data [23, 125, 58]

Interactive visualisations in general support data analysis processes as long as the interaction possibilities are intuitive and all required interaction possibilities are available while not overwhelming users with irrelevant interaction possibilities or too many degrees of freedom when navigating. Therefore, dashboard designers have to find the right trade-off between providing powerful interaction possibilities and simplicity.

Guideline 6: Provide an overview, show details on demand [23, 17, 84, 125, 58]

Visualisations in general should provide an overview of the data including most relevant information, details on certain data fields can be investigated on demand. This guideline is part of the well-known InfoVis mantra by Shneiderman [112]: "Overview first, zoom and filter, then details-on-demand."

Guideline 7: Colour use within the dashboard [23, 17, 84, 58]

Colours used within a dashboard and especially within visualisations, need to be carefully selected. Dashboards, only highlighting most important data, enable users to focus on important measurements [23]. Using too many colours, can lead to several problems. First of all, they might complicate the process of identification of important measurements or even outline less important measurements. Secondly, additional problems arise when also considering the fact that many people have colour deficiencies and therefore difficulties in distinguishing colours [80]. In order to enable users to analyse their data, only few different colours should be used. The number of different colours always depends on the purpose but should not exceed human ability to distinguish colours easily, which is possible for about 12 different colours [80].

Guideline 8: Dashboard customisation [23, 125]

In order to design a dashboard for a specific use case, the design of the dashboard should reflect the corporate design of the project partner. This includes colours, icons, symbols and shapes. Furthermore, considering different user groups, the dashboard should meet all requirements. This can be supported by providing different views or by enable users to configure their dashboard depending on their demands.

Guideline 9: Design for device scalability [17]

In order to enable users to investigate their data using different devices, providing a responsive design or a separate dashboard for mobile devices may support users in performing their tasks. This is common for users who have to take decisions in production plants where no desktop computer is available. Complex visualisations might not be suitable for analysing in mobile devices, thus, providing a simplified version for mobile devices might be required.

Guideline 10: Dashboard evaluation [23, 84]

In order to investigate whether a dashboard provides all necessary pieces of information, is suitable for performing analysis tasks, and to discover opportunities for improvements, user feedback should be collected, and an evaluation should be conducted. User feedback can lead to adjustments within the dashboard but may also be relevant for designing new dashboards in the future.

All dashboards discussed in this thesis are investigated according to these guidelines.

Chapter 3

Software Project Portfolio Monitoring

3.1 Introduction

This chapter discusses a dashboard created for monitoring project portfolios.

The author of this thesis contributed to the portfolio layout, fully implemented the dashboard, designed and performed the evaluation of the dashboard and published the results in the full paper *A Visual Approach to Project and Portfolio Monitoring* [98].

The design and implementation of this dashboard has been a collaboration of several people. The project symbol has been proposed by Wolfgang Kienreich from Know-Center in close corporation with the project partner. The proposed layouts have been discussed by the author of this thesis, by Wolfgang Kienreich and Vedran Sabol, also from Know-Center. The dashboard has been implemented by the author of this thesis. An evaluation was designed, executed and evaluated mainly by the author of this thesis and was supported by Wolfgang Kienreich and Vedran Sabol. Furthermore, a summary of the project symbol, the two layouts and the evaluation results have been published by all involved people. The author of this thesis thanks all contributors for collaborating, especially for proposing the project symbol design and guidance throughout the implementation and evaluation phase.

3.2 Motivation

Software projects are usually complex and their complexity even increases by a higher project volume, by more people being involved as well as by other influencing factors. Therefore, handling only one project is already challenging, but managing several projects at the same time requires a lot of effort to coordinate. Thus, the main motivation of this project was to support managers in monitoring and taking decisions by considering single as well as whole project portfolios. Summarising most important project variables and displaying them in a portfolio layout, allows manager to analyse them fast and take decisions accordingly. Most of the time, decisions in one project affect other projects as they have shared resources. Therefore, taking the right decisions requires to consider a larger context including all projects and their variables.

3.3 Requirements

In order to allow managers to capture all project relevant information within a short period of time, a project symbol encoding most relevant information easily, is required. The project symbol should allow managers to receive an overview and take decision immediately while not overloading the symbol with less relevant information. Therefore, the project symbol requires simple visual components including icons and intuitive colour coding. Additionally, a portfolio layout that enabling users to capture all relevant data variables for monitoring multiple projects in parallel, is required. Positioning projects in a portfolio layout can be performed according to different criteria. This can range from two variables to more complex similarity computations considering many project variables. Using different criteria for positioning projects within the portfolio layout allows managers to receive additional insights, for example identifying outliers or critical projects. Using a similarity computation, projects are outlined by positioning similar projects close to each other whereas projects, which are not similar to each other, are positioned far away from each other. Finally, portfolio layouts should support managers in analysing the effects of decisions taken in one project on other projects.

3.4 Dashboard Design

The design of the dashboard was split into two steps. Initially, the design of a single project symbol reflecting most important variables was created. In the second step, projects were displayed in two portfolio layouts, one matrix and one similarity layout providing different insights on the project portfolio.

3.4.1 Project Symbol Design

In the first step most important project variables were identified to be encoded within the project symbol. Besides some general measurements specific data fields were selected by the project partner as well. The data fields encoded in the project symbol are listed in the following:

- project volume
- overall project progress

- existence of any risk
- ratio of already spent amount of money, time and resources to their allocated amount
- number of work packages, their volumes and progresses
- project stage

During several design phases, the proposed symbol was refined. In the following, the design process of the project symbol is discussed in detail.

One of the main parts of this project was the initial design of the project symbol. As the general corporate design of the project partner contains circular elements, designing a symbol conveying all project relevant information should also be mainly circular. The first

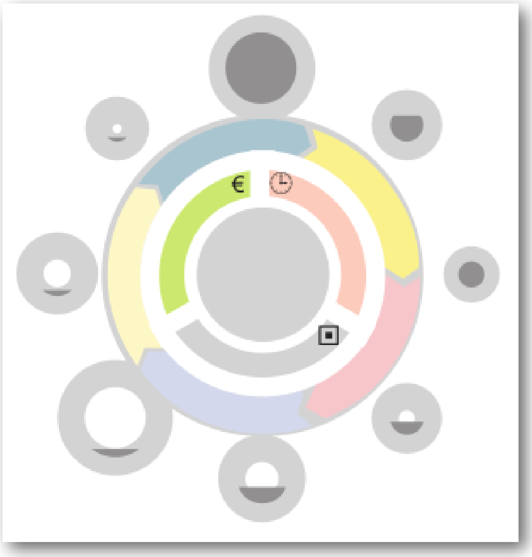


Figure 3.1: Initial proposal for a project symbol design.

design proposal shown in Figure 3.1, includes several circular elements. Most important data fields including overall progress are positioned in the centre. Around this circle, fields for money time and resources are positioned displayed as arcs. The third circle shows the five project stages defined by the project partner; additional circles could be easily added around this to provide further information. The circles in the surroundings represent the different work packages, whereby their sizes indicate their volumes and their fill statuses their corresponding progresses.

The amount of information within the first proposal was appreciated by the project partner but as a project can only be in one project stage and the number of nested circles was too high, further design improvements were proposed shown in Figure 3.2.

In the second phase, the third circle indicating the project stages was removed. The project stage was positioned at the top of the outermost circle connected to the work packages. Different project stages can be identified by colours proposed by the project partner. The new design reduces the amount of information provided by the symbol but still conveys all relevant information. The last (most right) layout in Figure 3.2 was considered to match the expectations best, therefore this design was selected for being implemented.

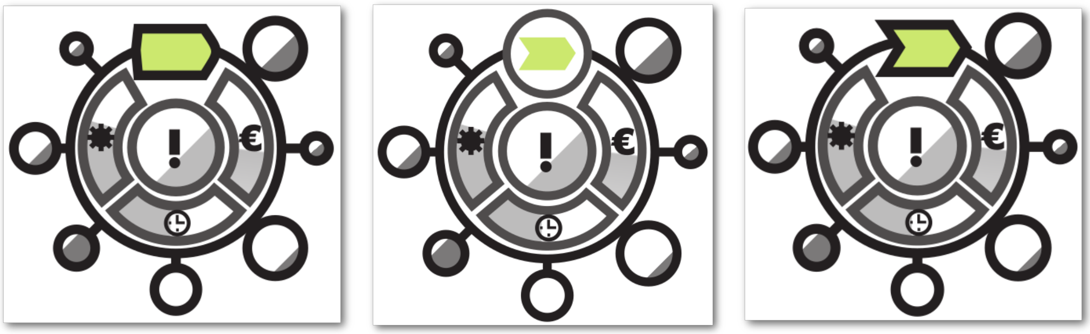


Figure 3.2: Second iteration of project symbol designs.

There have been small adjustments during the implementation phase to provide a clear and easy to read layout. Initially, the fill statuses were indicated to be shown diagonally instead of evenly, which changed to simplify the readability of the symbol. The initial design did not contain any colours, but the shapes were proposed to have borders. These borders were removed as they had a negative impact on the readability of the symbols, especially in the portfolio layout, where many projects are positioned close to each other. Each project symbol is displayed very small within the portfolio layout and thus, showing only one surrounding circle allowed to have a clean symbol design without further disturbance. Since the whole symbol is circular, the project risk symbol also received a rounded shape. Due to the fact that the general colour scheme of the partner company mainly includes grey scales, the overall design including fill statuses also became grey making it possible to highlight those variables, which convey important information within the project symbol. Colours for money, time and resources were carefully selected in cooperation with the project partner, to allow a clear differentiation between these variables indicating the ratio of spent to allocated values.

Figure 3.3 shows the proposed project symbol. The overall shapes of the symbol and containing elements are circular except the project stage at the top. The most important project variables are positioned in the centre, less relevant information is positioned in the surroundings. The innermost circle shows the overall progress of the project. The higher the fill status (dark grey) the further the progress. The risk symbol in the centre is only shown if any risks threaten the project. Around the inner circle, there are three arcs each showing the consumption status of one of the three main data variables: money, time and resources. Their fill statuses again indicate their consumption value, the higher the fill status, the more money, time or resources were already consumed. The arrow at top of the symbol indicates the project stage. There are five stages each project runs through, starting with the *idea*, followed by the *decision*, *planning* and *implementation* and concluding with the *completion* of the project. The corresponding colours are shown in Figure 3.4. Around the outer main circle, small circles are positioned representing the different work packages. The size of each circle shows their corresponding volume, the fill status again their progresses. Additionally to the encoded information, the overall size of the project symbol within the portfolio layout indicates the project volume.

Figure 3.3: Project symbol design.

Figure 3.4: Colour coding of the five project stages.

3.4.2 Portfolio Design

In order to allow manager to analyse a project portfolio, two different layouts, a matrix and a similarity layout, were proposed. This section concentrates on discussing these two approaches and outlines their advantages and drawbacks.

3.4.2.1 Matrix Layout

The matrix layout allows to position projects according to two selected properties, which can be selected depending on users' preferences. The projects are positioned in a 2D matrix were each selected property is assigned to one axis. In order to avoid occlusions due to similar values, a grid layout was used aligning the projects and therefore increasing readability. If two projects fall into the same grid position, there are two options in general: choosing a higher granularity of the grid or moving overlapping projects to neighbouring positions. Within this project, the second approach was selected for simplicity reasons.

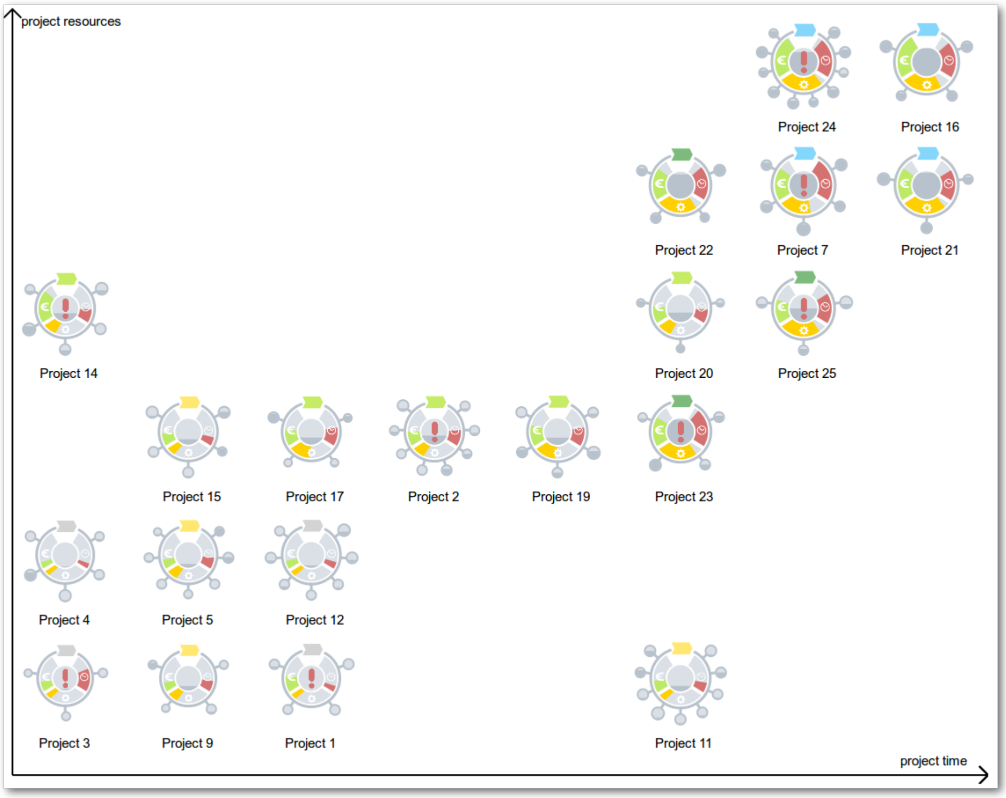


Figure 3.5: Matrix visualisation displaying allocated time on x-axis and allocated resources on y-axis.

Figure 3.5 shows the positioning of artificial projects according to the two data variables time and resources. Projects with small volume of time and resources are positioned towards the lower left, whereas projects with high volume of time and resources are positioned towards the upper right. The drawback of this portfolio visualisation is that projects can only be displayed according to two different properties. Therefore, outliers related to the two measurements can be identified easily. Identifying outliers in terms of money, requires the selection of different data variables for the matrix layout as well as further analysis.

3.4.2.2 Similarity Layout

In contrast to the matrix layout, all project variables are considered for positioning within the similarity layout. These are the overall progress, the existence of risks, the total amount of allocated as well as already spent money, time and resources, the number of work packages as well as the project stage. In order to enable displaying projects using complex similarity computations, an FDP (force-directed placement) algorithm was used [25] for positioning projects within a portfolio layout. Figure 3.6 shows an artificial portfolio layout using a similarity calculation. The similarity calculation positions projects with similar properties close to each other while projects with no or only little

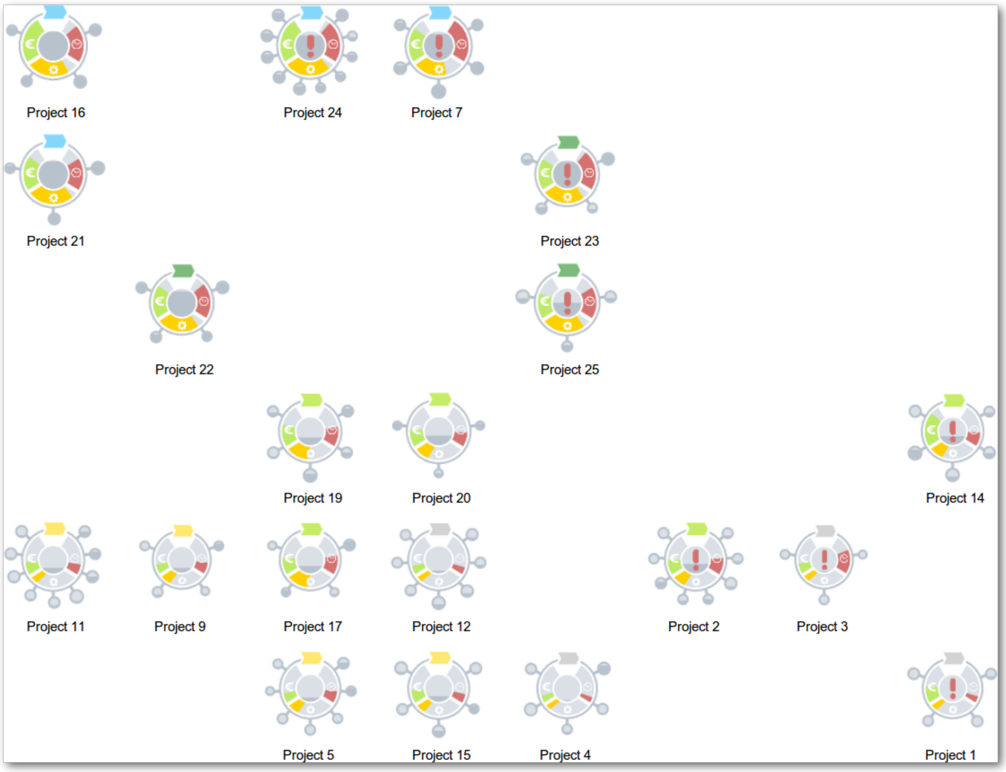


Figure 3.6: Similarity visualisation of the project portfolio.

common properties are positioned far away from each other. The projects are again positioned according to a grid with the same collision avoidance strategy as mentioned for the matrix layout to provide a good overview of the whole portfolio. The similarity computation and therefore the weighting of the data variables of a project can be defined by the user to emphasise those variables, which have a higher relevance.

In the proposed example the project risk gets a very high weight as critical projects are of special interest. Therefore, manager can observe that all projects containing a risk are positioned in the right. Furthermore, projects in a later stage and therefore consumed more money, time and resources are positioned in the upper area whereas projects which have just started are positioned in the lower area. The similarity computation proposed in this thesis is described Equation 3.1.

The similarity approach tries to overcome the drawbacks of the matrix layout by considering several variables for the similarity calculations. However, the proposed similarity calculation requires to be transparent to the project managers enabling them to draw the right conclusions from the proposed similarity layout. To support this, the dashboard allows managers to define their own sense-making similarity computation.

$$
\begin{aligned}
Similarity = (& \, stageSim + riskSim + resourceSim + moneySim + timeSim + \\
& overallProgressSim + workPackageSim + \\
& resourceConsumptionSim + moneyConsumptionSim + \\
& timeConsumptionSim \,) \,/\, 10
\end{aligned}
$$

$$stageSim: \; 1 \, if \, equal, \, otherwise \, \frac{3}{4}, \frac{1}{2}, \frac{1}{4}, 0$$

$$riskSim: \; 1 \, if \, equal, \, otherwise \, 0$$

$$
\begin{aligned}
workPackageSim: & \, scale \, numWorkPackages \, [0,1]; \\
& 1 - Math.abs(project1value - project2value)
\end{aligned}
$$

$$others: \; scale \, value[0,1]; \, 1 - Math.abs(project1value - project2value)$$

3.1: Equations for calculating similarity of project properties.

3.5 Case Study

The following description provides an example on how project managers can use the matrix and similarity layout for analysing their projects to overcome potential problems.

3.5.1 Matrix Layout

Investigating the matrix layout proposed in Figure 3.5, a project manager can immediately see that most of the projects are aligned along the diagonal indicating that project time and resources often correlate with each other. The matrix visualisation shows two exceptions, project 11 and 14. While project 11 has much time and only small resources, these data fields are inverted in project 14. Furthermore, the projects can be split up into the lower left area and the upper right area. While projects in the lower left area have only little time and resources and are mainly in earlier project stages, projects in the upper right area have a lot of time and resources and are in later stages consuming more money, time and resources. While outliers in terms of resources and time are only visible in this layout, the assignment of early project stages in the lower area compared to projects in their late stage in the upper right area is only related to the underlying artificial dataset. This information shows the manager that several larger projects will finish soon while no new project with larger volume is in the pipeline. This might encourage the manager to initiate new project with larger volumes. The portfolio layout shows that eight out of twenty projects have a risk indicated by the risk symbol in the centre. Investigating these projects separately, project three, positioned in the lower left area, shows that while the project is still in the first out of five stages, a lot of time has already been consumed within this project. The same can be observed for money consumption in project one. In project 14 however, almost all money has already been spent during the planning phase. Project two in the centre of the portfolio layout, also indicates a risk. This could be due to low overall and work package progresses. Projects 7, 23, 24 and 25 are all in later project stages and have already consumed most of the available resources while the overall and work package progresses are not at the expected level. The matrix visualisation enables the manager to identify groups with respect to resources and time allocated for projects.

3.5.2 Similarity Layout

Investigating the similarity layout proposed in Figure 3.6, a project manager can identify four different groups. Projects in the upper right area are those, which are in later project stages and containing a risk, while projects in the lower right area also contain a risk but are in earlier project stages. Both contain four projects respectively. In contrast to this, projects without any risk and in later stages are positioned in the upper left area and include tree projects. Projects in early stages without any risks include nine projects and are positioned in the lower left area. Since the similarity equation shown in Equation 3.1 puts a high weight on the risk, project containing a risk a strictly separated from those not containing a risk. Additionally, the similarity of project stages receives a high weight. Projects in the same or neighbouring stages are positioned closer to each other than projects in the first and in the last stage. Other data fields like overall progress, progress of money, time and resources and number of work packages have comparable smaller weights. The similarity layout allows project managers to identify similar projects more easily. Projects containing any risks are grouped together, which simplifies the analysis of problems for project managers.

3.6 Software Design and Technical Information

The proposed dashboard was implemented in Java as requested by the project partner. In order to display the project symbol and the proposed layouts in any web browser, the code was translated to JavaScript using GWT (Google Web Toolkit) [34]. An API has been defined to pass any project data for visualising.

In order to support displaying many projects within a portfolio while not running into performance problems, the visual components were implemented using Canvas. In contrast to SVG, it allows to render much more components, as elements are not directly attached to the DOM (document object model) tree. While the browser can simply render attribute changes in SVG, using Canvas, objects require to be redrawn if changes are applied.

The following API has been defined for dashboard configuration and displaying project data:

- This method allows to pass projects for visualising to the GWT component:

  public void setVisProjects(Set<Project>displayProjects);

- This method allows to define the size of a project symbol, if the size is not specified explicitly, 175 pixels are used as default:

  public void setSymbolSize(int size);

- The following two methods are related to mouse interactions and allow to identify the project as well as the containing elements for a certain mouse position:

  public ProjectData getProject(int x, int y);

  public ProjectData.ProjectElement getProjectElement(double x, double y);

- This method allows to set the canvas size, if the size is not set explicitly, the default size is 500x400 pixel:

  public void setSize(int width, int height);

- This method returns the preferred canvas size for a given number of projects for matrix and similarity layout (width equals height):

  public int getPreferredCanvasSize(int numProjects);

- This method returns the preferred symbol size in pixel for a given number of projects and a defined canvas size (width equals height):

  public int getPreferredSymbolSize(int numProjects, int canvasSize);

- This method allows to choose the type of visualizations, options are single project, matrix layout or similarity layout:

  public void setVisualizationType(DataProviderClient.VisualizationType type);

- The following two methods allow to define the variables for the x and y axes of the matrix layout, available are costs, resources, time and number of work packages:

  public void setMatrixVisXType(ProjectData.ProjectElement element);

  public void setMatrixVisYType(ProjectData.ProjectElement element);

- The following methods are related to the similarity layout, the first one defines whether the proposed standard similarity computation is used (true) or not (false), the second method allows to set similarity values for two selected projects, if nothing is provided, the similarity value is 0:

  public void setCalculateSimilarities(boolean calculate);

public void setSimilarityValue(int projectId1, int projectId2, double similarityValue);

- This method allows to defined whether to show or hide labels in the portfolio layout, if nothing is configured, they are show:

  public void setShowLabels(boolean showLabels);

Almost all variables can be configured within the settings allowing the project partner to configure the dashboard according to their demands. This includes colours and sizes of all project symbol components.

In order to provide a demonstrator of the component, a testing environment was prepared allowing to deploy the visual interface on a Jetty Runner [21]. This enables users to analyse artificial project data for evaluation and demonstration purpose.

3.7 Discussion

This section discusses whether and how the guidelines defined in section 2.4 were considered for this dashboard.

G1 (information): In the first phase of the project, most important data fields were identified for being visualised within the project symbol. Therefore, this guideline was fulfilled by identifying most relevant data fields.

G2 (visualisation) and G4 (clutter reduction): During several iterations, a project symbol was designed matching all requirements and expectations. Thus, the project symbol conveys most important data fields using simple visual components, while less relevant information is not included. Therefore, these two guidelines were fulfilled during the design phase.

G3 (UI): The two portfolio layouts position projects according to certain data fields. This can be done according to two data fields using a matrix visualisation or by using a more complex similarity computation. Positioning projects along a grid enables managers to identify groups sharing certain properties easily. Therefore, this guideline was fulfilled.

G4 (clutter reduction): During the implementation phase, the project symbol design was further refined to increase the readability of the symbol within the portfolio layout. Therefore, borders were removed except for the work packages. The same was true for diagonal fill statuses within the project symbol. By considering this guideline, the readability of the symbol increased dramatically.

G5 (interaction) and G6 (overview-detail): The two provided layouts included simple interaction possibilities, which can be handled by the project partner's code to investigate details about each project. Therefore, the two layouts provide an overview of the projects by showing most relevant data fields; details can be provided by simple click interactions. These guidelines were partially fulfilled as the dashboard allows user interactions and can provide details on demand; however, the was not handled within the project.

G7 (colour): The dashboard only contains few colours, the overall symbol is coloured grey, while only the project stages, the risk symbol and the three data fields time, money and resources contain other colours. This enables managers to concentrate on these data fields, while for example progresses of each work packages do not use colours and therefore do not attract attention while analysing. Therefore, this guideline was fulfilled during the design phase.

G8 (customisation): The project symbol reflects the project partners' corporate design, including shapes, colours and icons. The overall project symbol is circular and uses different shades of grey reflecting the project partners overall circular grey design. Colour coding for the project stages reflect the corresponding stages of the project partner. This is also true for the icons used for different data fields. Considering all these factors enables managers to easily investigate the project symbol as it uses the same design language as the project partner's application. Additionally, data fields for the two axes of the matrix layout or similarity measures can be configured for the similarity layout and thus, enabling managers to arrange projects according to their demands. Therefore, this guideline was fulfilled during the design phase.

G9 (devices): The layouts were created for desktop applications not considering mobile devices, since this was not a requirement within this project. However, single projects can be easily analysed on mobile devices. Therefore, this guideline was not fulfilled.

G10 (evaluation): An evaluation, see section 10.1, was conducted to investigate the readability of the project symbol and the usability of the two provided portfolio layouts. The evaluation revealed that the majority of participants appreciated having a symbol conveying most important project variables. Few participants stated that it still conveys too little information, for some others, the amount of information conveyed by the symbol was already too much. The two proposed portfolio layouts were appreciated by the participants. Some users preferred the matrix layout as the similarity calculation was not entirely clear to them. Results also showed that providing different portfolio layouts, where users can configure their axis or similarity computations, allow users to investigate different aspects as each layout provides different insights and therefore allows different conclusions. However, displaying many project symbols within a portfolio view decreases the project symbol size, thus, identifying different project volumes, indicated by their overall symbol size, is difficult to recognise. Therefore, this guideline was fulfilled and revealed further insights about the readability and usability of the symbol and the portfolio layouts.

3.8 Conclusion

In this project a new approach for visualising project portfolios has been introduced. Conveying all relevant project variables within one project symbol using simple visual components enables project managers to receive a quick overview of all projects. In order to compare projects and analyse them in a larger context, a matrix and a similarity layout were proposed, where projects are aligned along a grid according to selected properties. This enables project managers to identify outliers, group projects according to certain properties but also to easily identify projects containing risks. An evaluation was conducted to investigate whether the proposed project symbol is suitable for conveying all relevant information and whether the proposed portfolio layouts are suitable for analysing portfolios. Results showed that participants appreciated the project symbol, and the two portfolio layouts are suitable for analysis, whereby the matrix layout was preferred by more participants. The proposed symbol was integrated in the project partners monitoring application. Further ideas including changes over time or resources shared by several projects were proposed but not included to the portfolio layouts.

Chapter 4

Monitoring Production Plants

4.1 Introduction

This chapter discusses a dashboard created for monitoring the production of components in multiple production plants. The proposed dashboard is a general approach suitable for monitoring production in any domain. As production for this dashboard is fully automated, production data is aggregated on the fly, enabling users to investigate results immediately.

The author of this thesis was responsible for project management and for the conceptual design of the dashboard.

The design and implementation of this project has been a collaboration of several people. During the first phase of the project, the initial project proposal was discussed in cooperation with the project partner. During this process, five widgets were identified as being relevant for supporting the analysis process. The dashboard design was created based on this initial information. Visualisations for the five widgets were carefully selected and proposed by the author of this thesis. Design refinements of all dashboard components have been performed during the implementation phase. While the author of this thesis was responsible for the overall project management including the design and usability of the dashboard, the implementation has been a collaboration of several teams including the Service and Development team led by Werner Klieber as well as the Ubiquitous Personal Computing team led by Viktoria Pammer-Schindler. Therefore, the author thanks all contributor for developing the dashboard, including Werner Klieber and Werner Schmierdorfer for implementing the back-end and providing a test data generator as well as Patrick Lengauer, Jörg Simon, Peter Marton, Oliver Prentner and Alfred Wertner for working on the front-end and speech recognition. Furthermore, the author of this thesis thanks Vedran Sabol for supporting the project management and the dashboard design process.

4.2 Motivation

The main motivation for conducting this project was to enable management and production workers to monitor multiple production plants. This includes the overall productivity including rejects and components, which require rework. The proposed dashboard should allow managers to analyse the production down to station level, enabling them to identify, which stations cause problems during production. This enables responsible production workers to fix upcoming problems immediately. Problems can be diverse, ranging from long cycle times where stations become bottlenecks, over production stops due to breakdowns, up to stations producing many rejects. Therefore, providing a dashboard including widgets containing different insights enables managers and production workers to monitor production processes and take action immediately and therefore reducing rejects, reducing costs for rework and deliver produced units in time.

4.3 Requirements

The dashboard should be available for different user groups. On the one hand, managers should be able to monitor the overall productivity of production plants and receive details about where and when problems occur during production. On the other hand, the dashboard should allow production workers, to receive details on which problems arose during production and take actions accordingly. This includes details on disturbances, stations which have long cycle times and therefore need further actions or stations which produce many rejects or require rework of many produced units indicating that something is wrong at these stations. Therefore, having different user groups with different requirements, the dashboard should enable all users to monitor the information they require in order to perform their tasks. Production workers often do not analyse their data in front of a desktop computer. Therefore, a responsive dashboard design is required enabling any person within the factory to analyse current production data and thus, adjust changes in the production process immediately if required by using tablets or even mobile phones. Other than that, problems arose at any station within the production plant should be outlined within the dashboard by enabling the system to handle push notifications from the production plants. Additionally, the demand of reporting problems by management or production workers within the dashboard needs to be considered by providing a log view. This enables users to provide information or insights gained during data analysis or information about the status of certain problems.

4.4 Dashboard Design

Initially, five different widgets containing different insights were reported as a requirement. Based on this information, different proposals have been provided to the project partner for visualising information. This includes overview visualisations, time related visualisations and selection opportunities and visualisations allowing trend and correlation analysis.

As an entry point, an overview tab was proposed showing the production values of all production plants over the last week. The proposed overview contains pie charts and line charts as shown in Figure 4.1, allows users to identify those stations, which have high number of rejects during the last days. Since this entry point for receiving an overview was identified as being suitable, it was considered for implementation.

After choosing a production plant within the overview, the suggestion to provide a calendar view, shown in Figure 4.2, using different colours and thus, enabling users to identify days where production values were poor, was accepted. Therefore, this overview of production values for a single plant was implemented in the final dashboard.

As an alternative to the calendar view, it was suggested to display small pie charts for each day instead of using colours. This option was rejected as this would contains too many small pie charts not enabling users to easily identify days where production was lower.

For the main dashboard view containing the five widgets, the following visualisations were proposed. In order to show correlations between produced units and the corresponding groups and stations, an interactive sunburst visualisation was suggested for analysis. As an alternative to the sunburst visualisation, stacked bar charts allowing users to drill down to group and station level and collapsible tree representations were introduced. Since the sunburst visualisation was considered to be best suitable for analysis, the sunburst visualisation was implemented.

For visualising time related information, four different approaches were proposed. First, simple line charts containing multiple lines were proposed. The second proposed time-based visualisation was a time line, showing different events for example errors or log entries. Additionally, the colour of the time line can be used to encode the machine status. Furthermore, stream graphs and event drops were proposed showing the development for certain categories for example disturbances, cycle times or produced units. Both approaches allow identifying temporal correlations by encoding different categories using colours, while their corresponding thickness matches the value at a certain point in time. Two visualisations, the line chart as well as the stream graph (for machine runtime) were selected for the dashboard. As the event drop visualisation is similar to the stream graph with the drawback of identifying correlations more difficult, it was not selected. The time line showing events was not realised due to limited project resources.

Finally, a correlation chart was proposed allowing users to identify, which errors occurred at the same time and are therefore correlated to each other. As this visualisation requires more complex analysis capabilities, it was not considered for being implemented.

The notification and log view were not considered to contain visualisations.

4.4.1 Dashboard Description

The resulting dashboard provided in this project contains five main views. These include an overview of the productivity of all production plants over the last week, a calendar view visualising the productivity of one production plant over one year, a configurable production plant dashboard containing five widgets for detailed analysis, a notification view displaying messages directly generated by the production plant on unexpected events and finally a log view enabling users to create log messages to provide information on problems occurred during production. These five views provide different insights on the production. In the following the two terms *group* and *station* are used frequently. In general, a production plant can be divided into several groups. One group is a logical entity within the production plant performing for example the same or successive operations. One group is made up of several stations. A station can be a single or several machines performing an operation during the production process.

4.4.2 Overview of Production Plants

The entry point of the dashboard provides an overview of all configured production plants, showing the production values of the last week in terms of produced units, see Figure 4.1. Produced units are split up into those which were okay, those, which required rework in order to be okay and rejects, where rework is not possible. The pie chart shows the percentage of each category in relation to the total production within the last week. The line chart shows the production over time split up into the three categories okay, rework and rejects. These two charts easily enable managers and production workers to receive a fast overview of the production values within the last week. While the pie charts provide a fast overview, the line chart provides details on single days allowing to investigating whether the production was constant or varying much within the last week. Both proposed visualisations provide additional information using tooltips but do not allow any further interactions. If overall results of a production plant are comparable poor, users can select the corresponding line for further investigation.

4.4.3 Calendar View of Production Plant

After selecting one of the production plants in the overview, the calendar view is displayed showing the relation of produced units compared to the number of rejects. Depending on the percentage of rejects on the overall productivity in a production plant, the corresponding day is coloured green, yellow, orange or red. If the percentage is very low, green is selected. If the percentage of rejects is comparable high the corresponding day is coloured red. The calendar view enables managers or production workers to select either a single day or a longer period of time for investigating the productivity in the configurable production dashboard. Figure 4.2 shows the production of the artificial production plant one. Days, which are coloured white, are those days, where the production plant was in idle

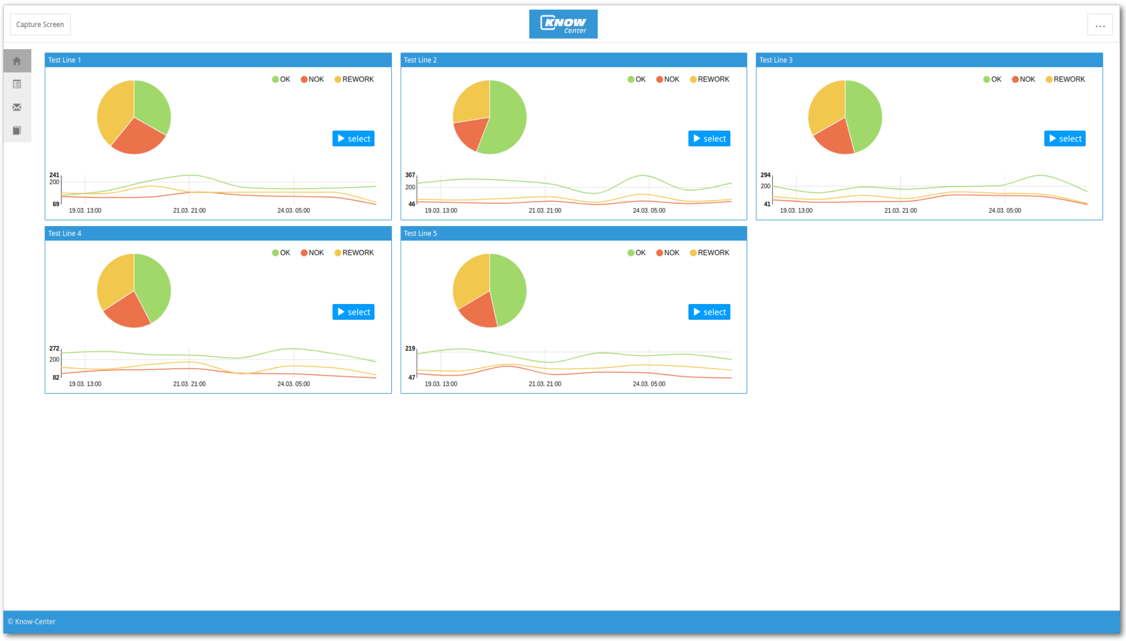


Figure 4.1: Overview of available production plants, showing produced units.

mode. After selecting one single day or a time range, the configurable production dashboard can be accessed.

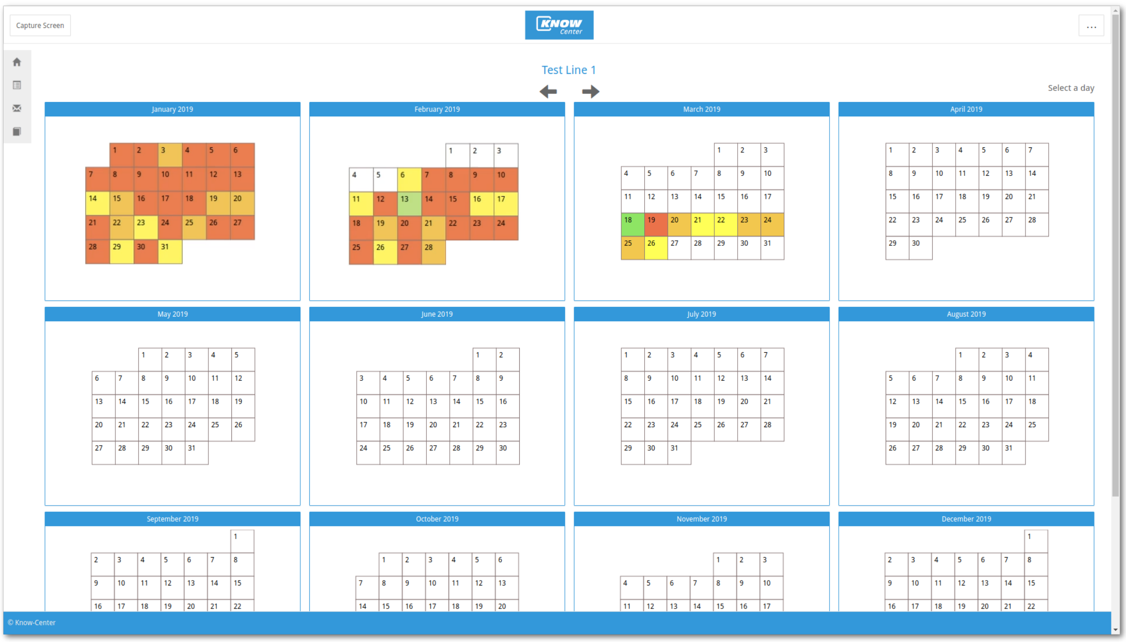


Figure 4.2: Calendar view showing the percentage of rejects compared to the total number of produced units on daily basis.

4.4.4 Configurable Production Dashboard

This view can either be entered via the calendar or by directly clicking on the second tab and choosing the corresponding production plant and date range for analysis. This view contains five configurable widgets providing different insights. Widgets can be added, removed resized and repositioned depending on user preferences. A screenshot feature enables users to take a snapshot of the current view for either creating a log entry or for forwarding the snapshot via e-mail. Figure 4.3 shows the dashboard containing the different widgets for the desktop view. All five widgets containing different insights are discussed in the following.

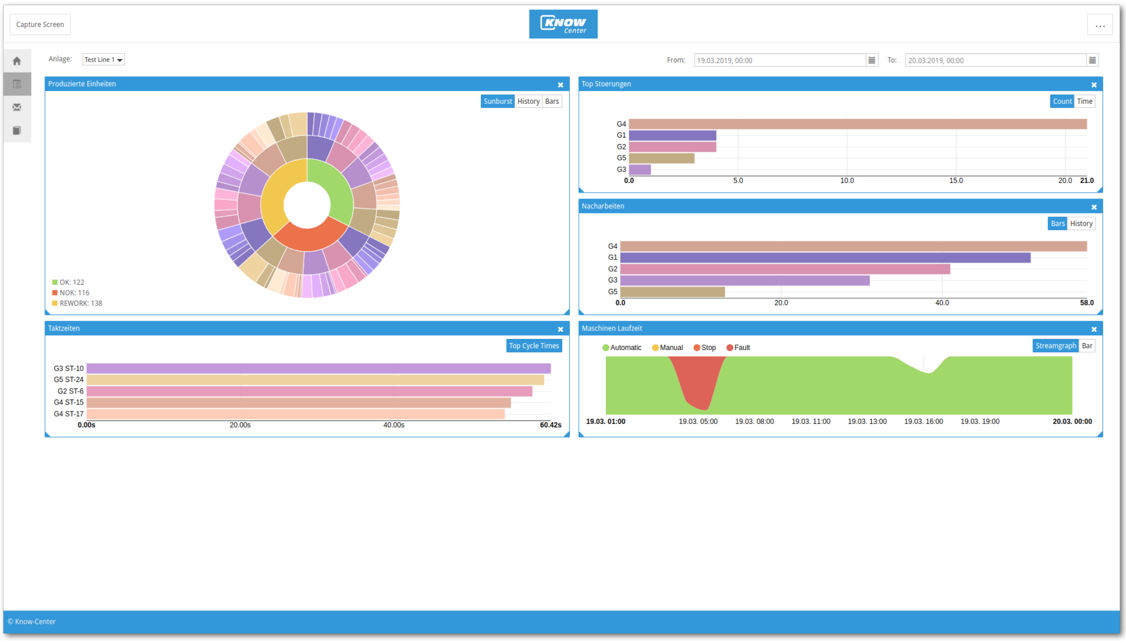


Figure 4.3: The dashboard contains five configurable widgets showing produced units, machine runtime, disturbances, rework and cycle times.

4.4.4.1 Produced Units

This widget contains three visualisations. The first one shows a sunburst visualisation. Each circle within the visualisation displays a different layer. The innermost circle contains three different colours, indicating the produced units: okay (green), rework (orange) and rejects (red). The sizes of the arcs match the percentage for each value compared to the total productivity, similar to the pie chart in the overview dashboard. The second layer shows the corresponding groups in relation to the innermost circle. This means that each group can occur multiple times within the second circle indicating that it produced units, which were okay, needed rework and were rejects. The same is true for the outermost circle, showing the corresponding stations producing these units. Again, the sizes correspond to the overall contribution of this station to the productivity. The sunburst visualisation allows drilling down to station level by selecting certain arcs for detailed analysis.

The second visualisation contains a line chart over the selected time range again showing the productivity. This visualisation does not provide further interactions. In order to allow displaying data over a larger period of time, data is aggregated on the server enabling the dashboard to display large time ranges. Tooltips provide information on the productivity within the line chart.

The third visualisation contains a stacked bar chart showing the productivity on group level. This visualisation allows identifying those groups, which caused highest rejects or required much rework. As the previous one, this visualisation does not provide any further interactions but again tooltips provide information about the productivity.

4.4.4.2 Disturbances

This widget contains two visualisations. The first one contains a bar chart showing the number of disturbances within the selected time range for each production group. The second widget contains a bar chart, which shows the duration of disturbances for each group allowing users to identify those groups within the production plant with long disturbance times. Both charts allow users to drill down and receive details. Clicking on one of the groups shows the distribution of disturbance classes within the group, again using the count or duration depending on the selected bar chart. Selecting one of the disturbance classes, the dashboard provides detailed information by showing disturbance messages including their start time and their duration.

4.4.4.3 Reworks

This widget contains two visualisations. The first visualisation shows the number of reworked units for each production plant group. Selecting one of the groups, allows users to drill down to station level and see which station required most rework within each group. The second visualisation shows the rework over time within a line chart for each group allowing to analyse those time periods where most rework was required. This visualisation does not support any further interactions but provides tooltips.

4.4.4.4 Cycle Times

This widget contains one visualisation. The bar chart shows stations with the longest cycle times and therefore being potential bottlenecks during the production process. In order to identify bottlenecks, cycle times need to be analysed in combination with rejects as well as reworks of products to receive further insights. This visualisation does not provide further possibilities for interacting; it just displays stations with long cycle times.

4.4.4.5 Machine Runtime

This widget contains two visualisations. The first visualisation contains a stream graph showing the percentage of different production plant statuses over the selected time period. A production plant can have the following statuses: *automatic*, *manual*, *stop* or *fault*. This visualisation allows analysing the different statuses over time to see how long the production plant was for example stopped due to a problem. The stream graph allows users to select any combination of statuses for detailed analysis. The second visualisation shows a stacked bar chart, displaying the overall distribution of the selected time range. This visualisation does not provide further interaction possibilities but only tooltips showing absolute values.

4.4.5 Notification View

This view displays notifications sent from machines of the production plant and therefore allows monitoring potential problems. Messages are sorted according to their date and can be filtered, and thus enabling users to search for distinct problems. A message contains a title, the date of occurrence, the corresponding production plant, the language, an expression indicating the affiliation of problems to predefined categories and the corresponding message. This information enables users to analyse which and how often problems occurred during the production process and whether this information can be assigned to a higher reject or rework number.

4.4.6 Log View

Furthermore, a view for log messages is provided in the dashboard. In contrast to automatically generated notifications, users generate messages in this view. This can range from fixing of arising problems, information on problems at a certain station or other comments. Furthermore, images or other files can be included to provide detailed information on certain problems.

4.4.7 Mobile Dashboard

In order to also allow displaying the dashboard using mobile devices, the implementation of the user interface was performed using Apache Cordova [126]. Thus, it requires to only have one common code base for the dashboard serving different devices. Details on the implementation are provided in section 4.6. Figure 4.4 shows the three main views using a mobile phone. The first image shows the production overview displaying the productivity of the last week. Only one plant can be shown at a time. Users can switch to other plants by wiping to the left and right. The second image shows the calendar view of a selected plant. Again, wiping to left and right allows switching to previous or next months. Other than that, the selection of days or time ranges is the same as in the desktop

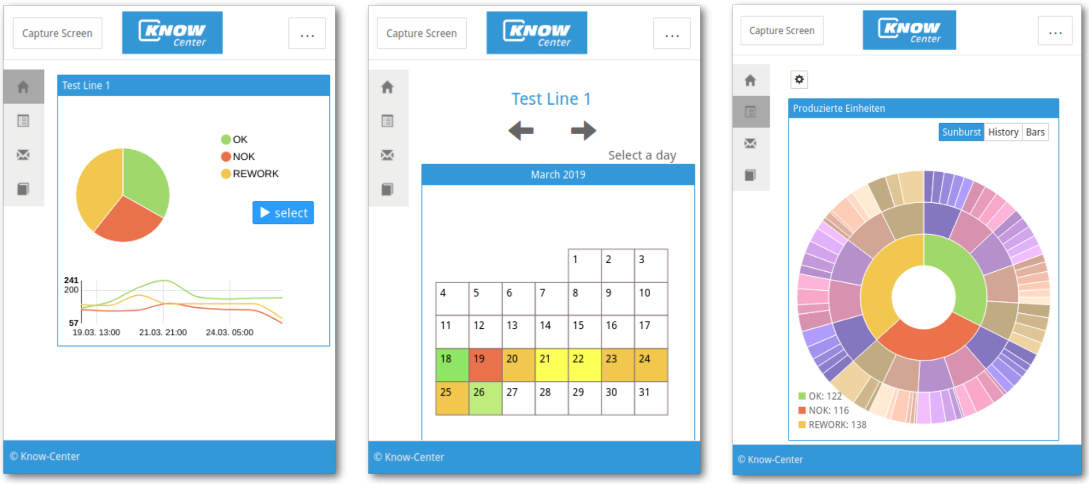


Figure 4.4: Overview, calendar view and configurable production view for mobile devices.

view. The configurable production view shows the five widgets, again one at a time. As the underlying code base is the same as for the desktop view, the information displayed in the mobile view is the same but only one widget can be displayed at a time. In terms of interaction possibilities, it supports the same interaction possibilities as the desktop view. Tooltips are also working in all views by keeping the element selected until information appears.

4.5 Case Study

The following description provides an example on how managers can use the dashboard in order to investigate the productivity of a production plant using the different views.

Assuming that a manager wants to investigate whether the overall productivity of all production plants is high, the manager first investigates the initial view showing the productivity of all plants within the last week, see Figure 4.1. The dashboard shows that the percentage of rejects and rework is comparable high for the first plant. Investigating the productivity for all plants show the following:

Line one has a total number of 2857 produced units during the last week, 955 out of them were okay (33.4%), 1107 required rework (38.7%) and 798 were rejects (27.9%). The high number and percentage of rejects requires further investigation but before continuing, the other plants are investigated.

For all other production plants, the percentage of rejects is lower, for plant two it has the lowest percentage of 16.5% followed by plant five with 19.8%, plant three has a reject rate of 20.9% and finally for plant four the reject rate lies at 23.5%. Investigating the overall productivity of all plants, the highest values were received by plant four with 3560, followed by plant two with 2920, plant one has 2857, plant three 2806 and finally plant five 2658. Although the percentage of rejects for plant four is lower than for plant one, the total number of rejects is higher than in plant one with a count of 837. Therefore, although producing comparable more units, which are okay, plant four also requires further investigation as it has the highest number of rejects.

First of all, the manager investigates plant one, as it has the highest percentage of rejects compared to the overall productivity, see Figure 4.5. The line chart of production plant one shows that the productivity was comparable low at the beginning of the displayed time, it was rising on 20th and 21st of March for all three values but comparable higher for units which required rework or were okay. After the 22nd of March the values were again falling and staying constant. The number of units, which were okay were again rising during the last day while the other two values were decreasing.

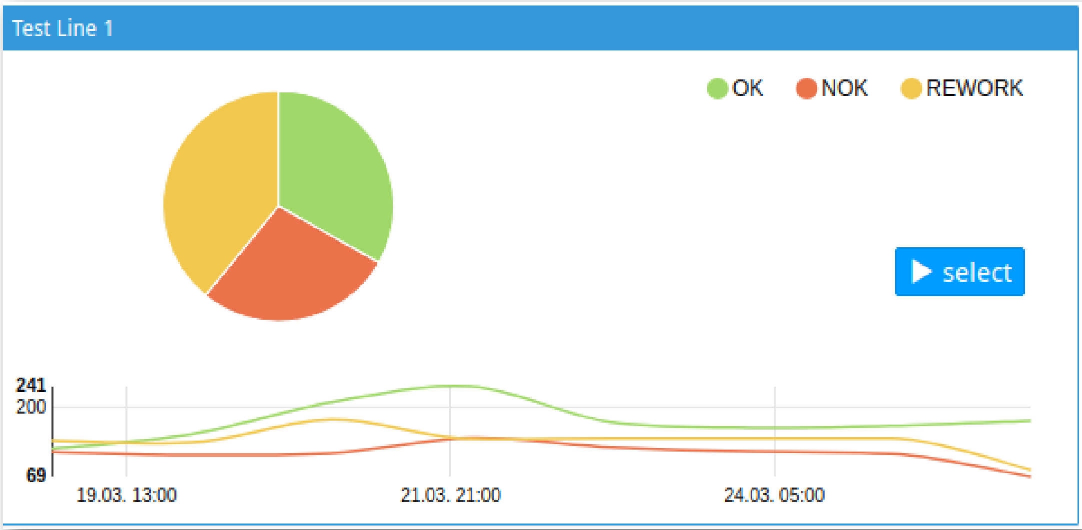


Figure 4.5: Investigating plant one within the overview tab.

After analysing the productivity within the last week, the manager selects the first plant and continues to investigate the calendar view, see Figure 4.6. The calendar view makes it possible to investigate the productivity for every day separately. Every day the production plant was running is colour encoded depending on the percentage of rejects on the overall productivity. While green indicate a low percentage of rejects, the number of rejects in red days is very high. White days in between indicate that the production was shut down for this plant. In the first month of the year, the overall productivity was really poor indicated by the red coloured days. Therefore, the production was stopped during the first five days in February to fix some problems. Afterwards the production plant was restarted. In February, the production values were again poor but slightly better compared to January. Because of this, the production was again stopped at the beginning of March for a total number of 17 days. After one week running again, the manager wants to investigate whether engineers could fix the problems and receive a higher productivity. In fact, the productivity during the last week from 18th of March to 26th of March was statistically better than in previous months. Only one day, namely the 19th of March, is coloured

red while all other days had better results. The manager investigates the 19th of March separately by hovering the mouse. For this day, the distribution of produced units was the following: 112 okay, 138 required rework and 116 were rejects. Therefore, the manager selects this single day for further investigation in the configurable production view.

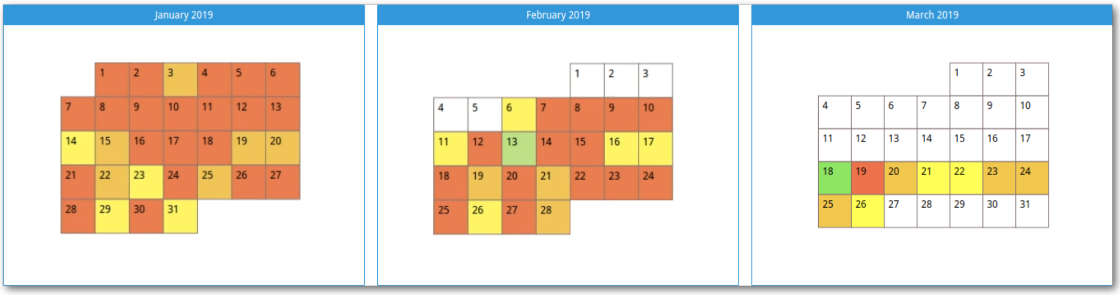


Figure 4.6: Analysing the first three months of production data by investigating the calendar few for plant one.

The production view, see Figure 4.3, shows the five widgets containing the production data for the 19th of March. The manager first investigates the produced units. As shown in Figure 4.7 and 4.8, there are three different visualisations within this widget. The sunburst visualisation shows the productivity split up into the three output categories, see left image of Figure 4.7. Furthermore, selecting one of the arcs, for example rejects, enlarge the arc and allows to investigate groups and stations which produced most rejects, see right image of Figure 4.7. The sunburst visualisation shows, that all four groups are affected. Group four has most rejects, namely 47. Group three has 23, group one has 20, group two 15 and group five 11 rejects. Investigating now the rejects in group four, there are six stations which are affected by rejects, two out of them have a higher number, namely station 19 has 17 rejects and station 17 has 15 rejects. These two stations produced highest rejects on the 19th of March.

Investigating the productivity over the whole day, shown in the left image of Figure 4.8, the overall productivity is not continuous, it fluctuates throughout the day. During the morning hours until about 7 am the number of rejects has three peeks whereas after 9 am there are also some peaks with a higher number of rejects but they are overall smaller. Finally, the stacked bar chart, displayed in the right image of Figure 4.8, shows the distribution for each group separately. It shows that the productivity of group one was the highest, including some rework but also some rejects, whereas the productivity of group five was the lowest but containing the lowest number of rework and rejects. Group four has the second lowest productivity with less than 300 units but has the highest number of reworked and rejected units.

Summarising results of the first widget, visualisations show that productivity was negatively influenced mainly by two stations within group four, which caused most rejects. Whereby the number of rejects was comparable high during the morning hours of the 19th of March.

The second widget, see Figure 4.9, shows the disturbances. Disturbances are displayed

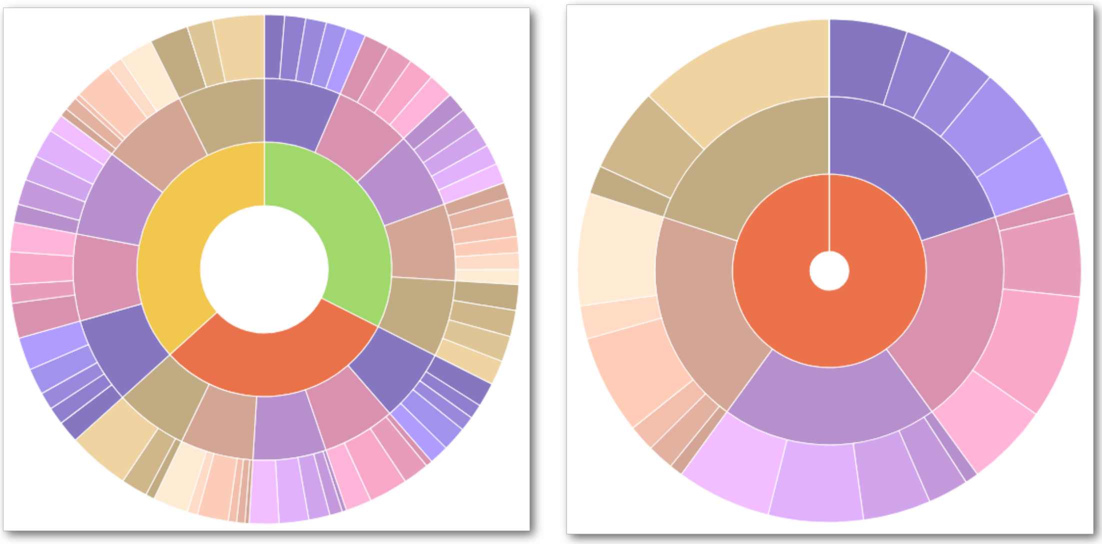


Figure 4.7: Sunburst visualisation showing the overall productivity of plant one (left), clicking on one of the arcs enlarges this area (right).

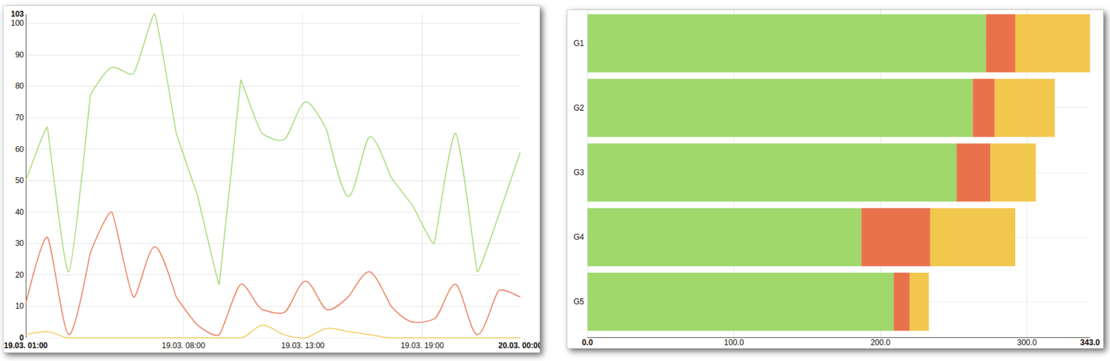


Figure 4.8: Line chart and stacked bar chart showing the productivity of plant one for a specific day.

by count (upper left image) or by duration (upper right image). The upper left image in Figure 4.9 shows that again group four has the largest number of disturbances, namely 21, whereas all others have four or even less. Selecting group four shows the corresponding message class, see lower left image of Figure 4.9. It outlines that message class three has the largest number of messages, namely six. Since the user wants to investigate the corresponding messages, the user clicks on the selected bar. Afterwards, details for each message are provided, see lower right image of Figure 4.9. Investigating the second bar chart for the disturbances, see upper right image of Figure 4.9, shows that disturbances in group four did not only have the highest number of disturbances but also took longer for being fixed compared to other groups where disturbances were significantly shorter. This leads to the conclusion that the high number of rejects in group four correlate with the disturbances occurred in group four on this day.

Investigating the rework widget, showing those groups, which required most rework,

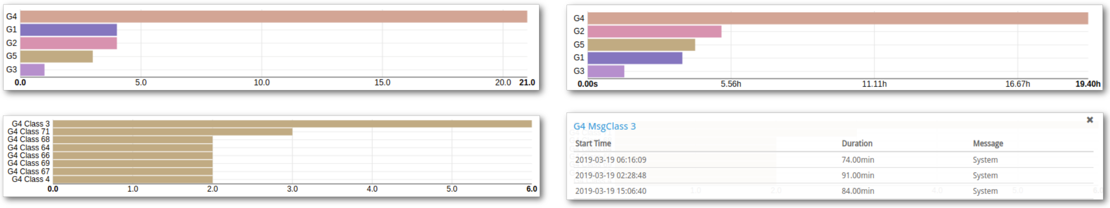


Figure 4.9: Disturbances of groups displayed by count (upper left) and by duration (upper right). Clicking on one of the groups shows the distribution of message classes (lower left), further selection shows the corresponding messages (lower right).

shows that group four again is listed first followed by group one and two. Selecting group four shows that station 17 and 19 had also most rework, see upper images of Figure 4.10. Selecting the line chart within the disturbance widget, it shows that rework was required at the beginning of a day as well as about 10 am until 4 pm, see lower image of Figure 4.10. Again, the high number of reworks at noon can be related to disturbances in group four.

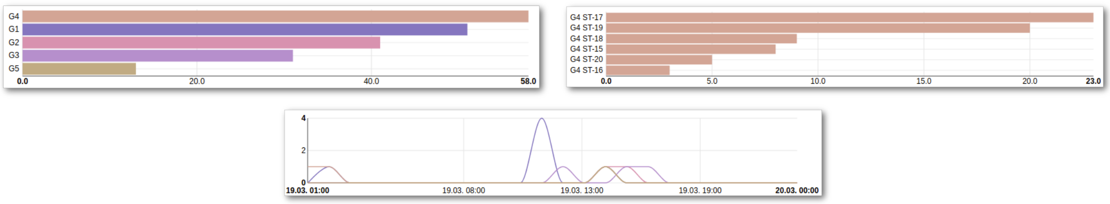


Figure 4.10: Rework of groups displayed by count (upper left) enables users to zoom in to station level (upper right), line charts show the production data of each group over time (lower centre).

The cycle times widget only provides a bar chart, showing bottlenecks in terms of processing durations, see Figure 4.11. The chart shows that station 10 in group four has the longest processing time followed by station 24 in group five. Station 17 in group four, one of the two main producers of rejects, is listed on the fifth position. These stations can be, besides stations producing many rejects or requiring much rework, potential bottlenecks during the production process.

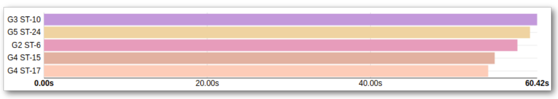


Figure 4.11: Visualisation within the widget *Cycle Time*.

Finally, the machine runtime widget shows that the production plant was in automatic mode most time of the day, see Figure 4.12. It was in fault mode only during the morning hours between about 4 am and 7 am. For some time in the afternoon at about 4 pm some measurements are missing, therefore this area is not colour coded. Removing the time where the production plant was in automatic mode, see upper right image of Figure 4.12

allows to investigate all other statuses in detail. This helps in investigating statuses which only occur for a short period of times. The corresponding stacked bar chart shown in the lower centre of Figure 4.12, again shows, the distribution between the different production plant statuses in absolute values.

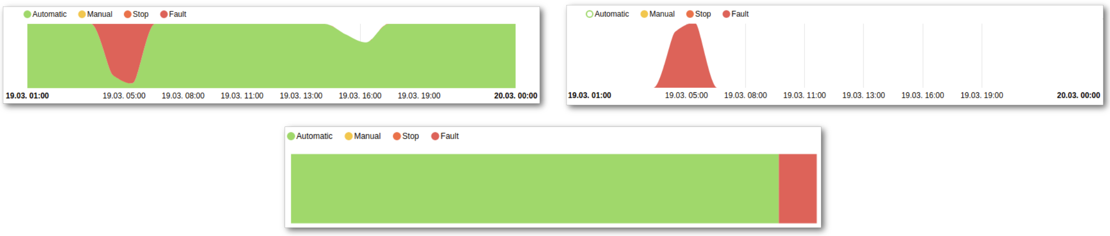


Figure 4.12: Visualisations within the widget *Machine Runtime*.

Concluding results from the configurable dashboard view, the manager now knows that group four mainly caused problems. Two stations mainly contributed to the high number of rejects, which require further investigation, for example by analysing the following days and prove whether production values in those days become better. Additionally, the manager needs to investigate production plant four, as the total number of rejects in absolute values is very high.

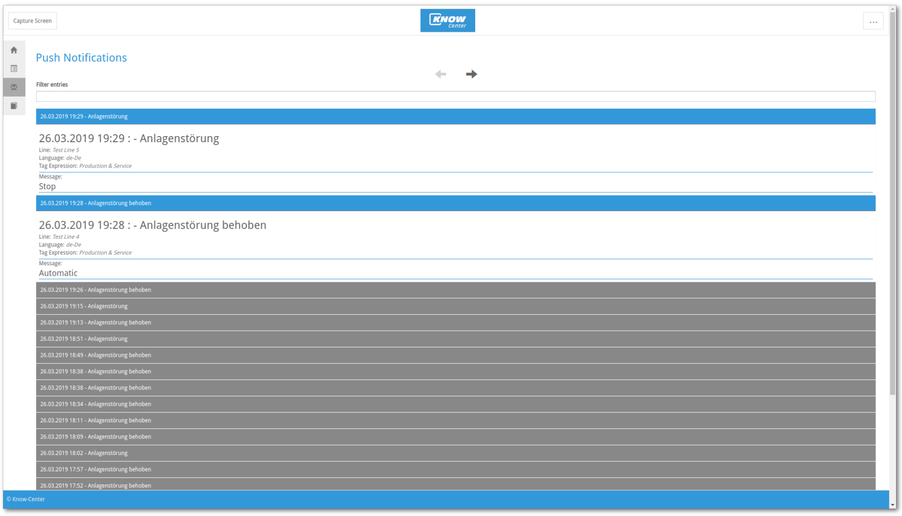


Figure 4.13: Notification view.

Investigating now the notification view, see Figure 4.13. The manager can see that there are several disturbances from different production plants reported. This information can help to gain further insights on what are the problems in different production plants for example the high number of rejects for station 19 and 17 in group four of the first production plant.

Finally, after analysing production plant one, the manager writes a log entry including a screenshot, using the screenshot feature proposed in the dashboard and provides infor-

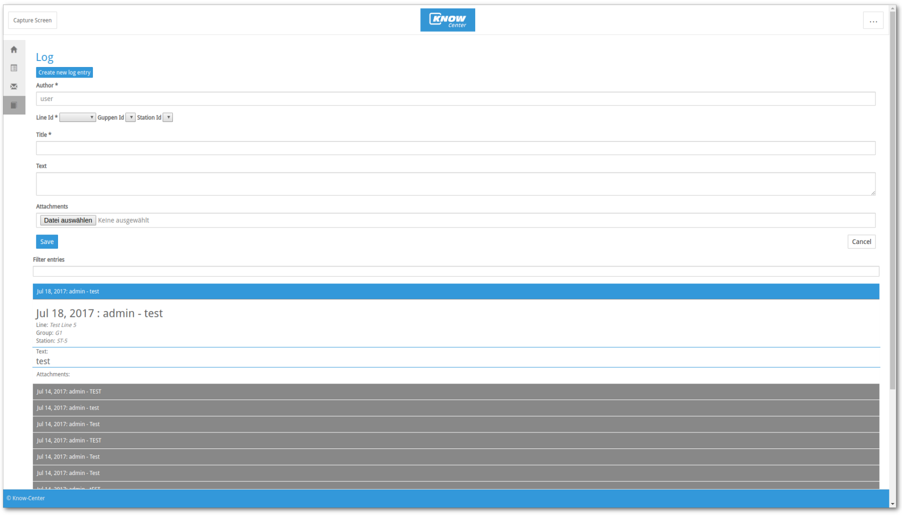


Figure 4.14: Log entry view.

mation that station 17 and 19 of group four requires to be investigated in detail for the following days, see Figure 4.14.

4.6 Software Design and Technical Information

The demand for providing a dashboard suitable for desktop but also for mobile devices enabling managers and production workers to always analyse their data raised during the project initiation phase. This led to the decision for using Apache Cordova [126] to have one uniform code base serving different devices at the same time. Therefore, Node.js [57] for client-side dependency management and development of the dashboard was selected. This enables simple dependency management as well as easy prototyping and development of the client dashboard. Additionally, AngularJS [32] was used to have a clear structure within the code base. In order to generate D3 [75] visualisations, a specific framework for AngularJS, called NVD3 [86] was used. The dashboard supports multiple languages as well as different URL (uniform resource locator) parameters allowing users to configure links for fast access. In order to provide a screenshot feature, different approaches have been investigated. Most of them do not draw the visualisations included properly. Therefore, a server-side solution called PhantomJS [91] was selected. This tool renders the provided HTML (hypertext markup language) page by itself and returns the base64 encoded image blob to the client. Furthermore, a feature for speech recognition was implemented enabling users to open different views of the dashboard without mouse interaction using pocketsphinx [1].

The back-end was implemented using Java programming language. In order to allow testing the framework without connecting it to an existing production plant, a data generation framework was proposed, generating artificial data. MongoDB [77] was selected for storing and retrieving data generated by the data generator and later by the production plant.

Loading data for a longer time range, requires data aggregation before transferring it to the dashboard. The aggregation level always depends on the selected time range and can be on minute, hourly, daily or weekly level.

Since the framework handles and shows sensitive data, providing a user management as well as encrypting information is a requirement for the final product. As this was not the primary goal of the prototype and implementing security features consume a lot of resources, only a simple user management was included for demonstration purpose. Further security features were not included within this prototype. When connecting this to existing production plants, the framework needs to include all necessary security features to ensure, that customers can only see their production plants and guarantee highest possible security level.

All three components including the client, the server and the data generator are bundled in a Docker container [20] enabling users to easily execute the proposed solution.

4.7 Discussion

This section discusses whether and how the guidelines defined in section 2.4 were considered for this dashboard.

G1 (information): Initially, major requirements containing different aspects of the production were identified. Based on this information, the dashboard was designed around five specialised widgets. Therefore, this guideline was fulfilled by identifying most relevant data fields.

G2 (visualisation): In order to identify suitable visualisations, different approaches were proposed to the project partner. This included an entry page for receiving an overview of production values of the last week, a calendar view showing production for single days and representation possibilities for the widgets within the configurable dashboard view. By providing different visualisations and select most suitable ones, this guideline was fulfilled.

G3 (UI): Considering again the proposed visualisations, only simple representation methods were considered as being suitable and thus implemented within the dashboard. This enables non-expert users to easily investigate the underlying data. Therefore, this guideline was also fulfilled during the design phase.

G4 (clutter reduction): In order to not overload the dashboard, only those data fields are shown within the charts, which have the highest values indicating bad production values. This is true for all contained visualisations. Therefore, this guideline was also considered for this dashboard.

G5 (interaction): In order to allow users to analyse production results in more detail, several interaction possibilities are provided. As visualisations in different widgets are not supposed to be coordinated with each other due to complex concepts, analysis possibilities remain limited. As this dashboard allows different interactions but not coordinating interactions between different widgets within the dashboard, which was not required, this guideline fulfilled. However, coordinating interactions between different widgets might support users while analysing their data.

G6 (overview-detail): All proposed visualisations contain only few data fields and thus most relevant information. Further details can be investigated by drilling down to the information of interest. Therefore, this guideline was fulfilled.

G7 (colour): In order to distinguish between products, which are okay (green), which required rework (orange) and rejects (red), colours were carefully selected to distinguish these categories. Colours for groups were selected by trying different colour schemes, which are different from the three selected values. Colours for stations were derived from their corresponding group colour using different opacity levels. Therefore, this guideline was also fulfilled.

G8 (customisation): The overall dashboard design including colours and icons was tweaked to match the project partner's corporate design. Additionally, the provided dashboard allows each user to customise the widgets according to their requirements and therefore enables different user groups to select the widgets they are interested in. Therefore, this guideline was also considered.

G9 (devices): In order to enable users to investigate current production data within the production plant without using a desktop computer, one of the main requirements was to provide a mobile application. Thus, the mobile application provides the same features as the desktop approach as both share a common code base. Since the dashboard does not contain complex interactions, which were also suitable for mobile devices, this guideline was also fulfilled.

G10 (evaluation): Finally, this dashboard was not evaluated with test users in a study. However, feedback from the project partner revealed that the provided dashboard was considered as being suitable for monitoring production. The visualisations implemented within the projects are simple charts enabling users to analyse their data and take decisions based on information visualised in the charts. Additionally, investigating production data revealed that providing pie charts within the initial view allows analysing the proportions within one production plant but does not enable users to compare production values between different plants. Other visualisations, for example bar charts, might enable users to compare production results from different production plants easier.

In order to enable users to visualise certain information, speech recognition was included to the dashboard using pocketsphinx [1]. Depending on the commands provided by users, the interface switches to the corresponding view. This feature was considered to support users to switch tabs without using their hands. As this feature requires a low degree of noise for recognising the commands, it turned out that it is not suitable for being used in a production plant. During the testing phase, it also turned out that commands used for enabling speech recognition requires a combination of words, single short words are not as reliable for initiating speech recognition. Therefore, this guideline was not fulfilled.

4.8 Conclusion

The proposed prototype was carefully designed to match the project partners' expectations and enable managers or production workers to analyse productivity and take decisions based on information visualised in the dashboard. Besides getting an overview of the production plants, the dashboard also allows to drill down to the desired information and identify the root causes of problems. The mobile dashboard enables production workers to analyse production values and arising problems directly in the production plant without using a desktop device. This enables employees to perform their tasks faster and more efficient, including immediate problem reporting, problem fixing, handling notifications from the production plant or taking actions based on log entries provided by the management. After introducing the prototype to the project partner, the proposed dashboard was considered to support the production process. Thus, they decided to create a product based on the prototype, which was done in-house for strategic reasons.

Chapter 5

Monitoring Production of Electronic Components

5.1 Introduction

This chapter discusses a dashboard created for monitoring the production of electronic components by animating production flows within the production plant. In contrast to the previous project, the production processes are not fully automated in this case. Thus, assembling of single components requires much longer. As a consequence, production data does not change much within a short period of time.

The author of this thesis was responsible for project management, mainly contributed to the conceptual design and fully implemented the dashboard and server component.

The design and implementation of this project has been a collaboration of several people. Especially the station design was refined during several iterations. Besides the author of this thesis and the project partner, Vedran Sabol, Patrick Lengauer and Wolfgang Kienreich, all from Know-Center, were involved in this process. The next step defined the creation of a layout reflecting the production plant and the animation process, which have been mainly designed by the author of this thesis, the project partner and Vedran Sabol. The dashboard has finally been implemented by the author of this thesis. During the second phase of the project, Sandra Feyertag designed a dashboard editor according to the instructions of the author of this thesis. The author of this thesis thanks everybody for collaborating, especially for contributing ideas to the station design.

5.2 Motivation

The main motivation of this project was to ensure that enough components required for assembling electronic components are available during production. As this is a critical factor in terms of production, the manufacturer can either buy all necessary components in advance or on demand. Both approaches of course have their advantages and disadvantages. On the one hand, buying all components in advance allows the manufacturer to install exactly the same hardware without further concerns but this requires planning of the whole production process in detail in advance, spending a huge amount of money in advance and providing enough storage for all components. On the other hand, ordering component for assembling just in time avoids all disadvantages of buying in advance but includes many risks starting from the production of components used for the final product, the delivery of components as well as further unexpected events which might have an effect on receiving the components in time and therefore may lead to a production stop. In order to find a trade-off between the two contradictory approaches the idea raised to implement a system, which allows monitoring the production and enabling people in responsibility to react on arising bottlenecks during production prematurely.

5.3 Requirements

Product processing time can range from few hours up to several days as most components are assembled by humans. Thus, single stations can become bottlenecks, which require early counteractions. In order to provide a dashboard for monitoring the production process supporting the identification of bottlenecks and other problems prematurely, visualising important resources available at each station as well as the production flow within the factory, is essential. Additionally to the latest production values, development of these values within the last hours is also relevant for analysis. If production values significantly deviate during a production period, managers should be able to go back to any point in time and observe the production back then. In order to allow monitoring the production, the layout of stations requires to be set up in a way which reflects the actual production flow. Stations and their statuses need to be outlined clearly by providing station names, the corresponding icons as well as simple visual components indicating the status of important data fields. Animations are required to reflect production flows allowing to analyse temporal development. Besides the actual dashboard, managers want to receive the latest production data while travelling. Therefore, a mobile version of the dashboard showing most important data fields is required.

5.4 Dashboard Design

The design of the whole dashboard was split up into several parts. Initially, the design of a single station reflecting the status of most important data fields was created during several iterations. The second step included the alignment of all stations reflecting the production flow. The third step defined the design of a simplified dashboard for mobile devices and the last step included the design of the animation flow.

5.4.1 Station Design

One of the major parts of this project was the station design containing all relevant data fields. These data fields are:

- station name and icon
- the number components in terms of money and hours waiting at each station for being processed
- both values, money and hours, with respect to a predefined minimum and maximum and the corresponding statuses
- the development of both values over time
- ratio of produced to expected produced components

There have been several iterations proposing different designs containing all this information. Figure 5.1 shows the designs within the first phase. The two data fields money and hours are positioned at the top. The actual statuses are shown in two rectangles filled according to its minimum and maximum value. The number as well as the status is displayed below the rectangles. The development over the last hours is shown in a line chart next to the rectangles. The icon and the name of the station are positioned in the centre. Arrows for components entering and leaving the stations were initially indicated to be on the left and right side. Additionally, the ratio of produced components to expected components was positioned in the lower part including a more complex visual representation. All three parts occupy the same amount of space indicating that all of them have the same relevance.

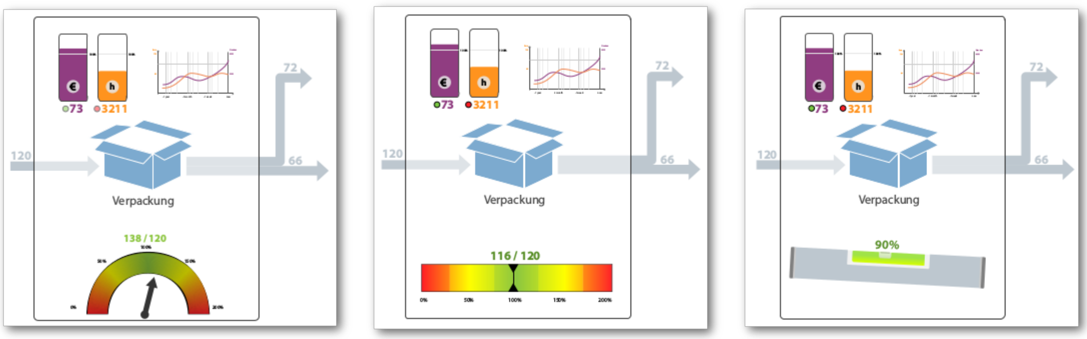


Figure 5.1: First iteration of proposed station designs.

During the next phase, designs have been adjusted due to the higher relevance of the two data fields money and hours compared to the icon and name and to the ratio of produced components, see Figure 5.2. Thus, the visual component at the bottom have been replaced by a simple traffic light symbol, which was moved to the upper right area next to the icon and station name, which are positioned in the upper left area. The two data fields money and hours are positioned in the centre indicting their status by using simple colour schemes. Due to the high relevance of the two data fields, the line chart is displayed larger to allow analysing the two values over a longer period of time.

Figure 5.2: Second iteration of proposed station designs.

After further discussions the station design has been refined to match the expectations of the customer, see Figure 5.3. The station name and icon are now in the lower part of the station. The productivity of components is still located in the upper right part or outside the actual station using a traffic light symbol. The two main data fields are displayed as rectangles or squares containing their status and the actual values. Each of the two data fields using their own colours, reused within the line chart, allows users to easily identify both data fields. In this phase also the idea arose that arrows entering and leaving the stations are positioned at the top and elements fall down to the rectangles indicating the movement of components.

Figure 5.3: Third iteration of proposed station designs.

After several iterations a layout similar to the last proposed one was selected. The traffic signal for the overall productivity was removed from the right and the ratio of produced components to expected components was moved to the lower right area of the station.

In order to support the design process, several mock-ups were generated including the status of several station embedded in a dashboard to illustrate the resulting view. An example therefore is shown in Figure 5.4. Embedding the stations in a dashboard supported the decision process as some visual components are displayed quite small and are hardly recognisable. Additionally, using many colours within the dashboard might lead to the fact that important measurements are difficult to recognise. Using red, yellow and green only in few places helps to identify critical values more easily.

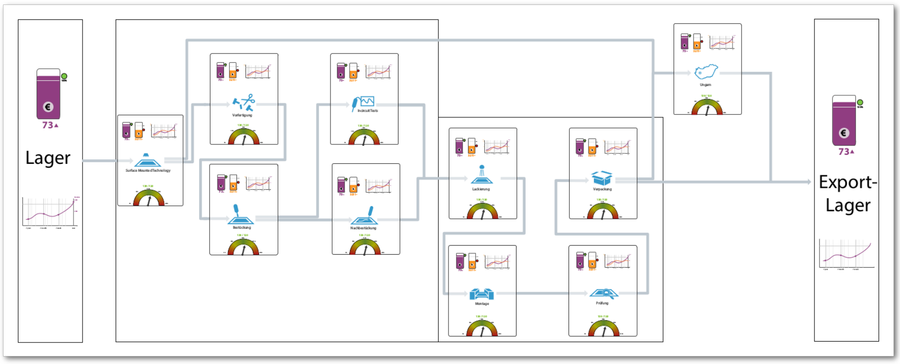


Figure 5.4: Dashboard view for one of the proposed designs.

5.4.2 Station Description

After several iterations of proposing different station designs, a simple design reflecting most important data fields, was selected, shown in Figure 5.5. The final design reflects the company's corporate design by using their colour scheme and simple shapes included in their logo. The upper two squares show the two most important data fields, with respect to their upper and lower limit. The blue one shows the number of components calculated in money, the grey one shows the number of components calculated in hours waiting at the station for being processed. Circles within the rectangles indicate the status of the data fields. The status can be okay (green), borderline (yellow) or critical (red). Displayed values within the squares provide their concrete values.

The combination of squares with circles inside was selected as it reflects the logo of the project partner. Below the two squares, a line chart is displayed showing the development of the two data fields over the last 72 hours. Clicking on the chart enlarges the line chart and shows the development over the last four weeks. In the lower part of the station symbol an icon and a name representing the station is provided. Furthermore, in the lower right part, the ratio of actual versus expected produced components during the last time period is shown. Additionally to the symbols used within the station, also the colour coding relates to the company's corporate design.

The proposed station design does only contain most important data fields, which enable managers to monitor the production process and identify bottlenecks in order to take actions accordingly. The station design involved several iterations to identify the optimal combination of visual components to provide a clear not overloaded design reflecting the company's corporate design.

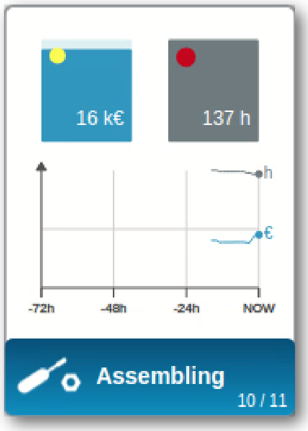


Figure 5.5: Station design for monitoring production flows.

5.4.3 Dashboard Layout

Two different layouts shown in Figure 5.6 have been generated reflecting the production flow. The original (left) one has been slightly adapted due to changes in the production plant. The stations are aligned along a grid providing a clear order of the production process. The layouts show, that the production is done in two locations, Austria and Hungary. Furthermore, connections between the stations visualising the production flow are displayed during animation. Detailed information about the animation flow will be discussed in section 5.4.5. The two layout views provide an overview of the production plant and allow managers to identify bottlenecks. This is ensured due to using simple shapes and few colours, which do not overload the dashboard but highlighting important data fields. For customers and guests entering the production plant, the animated dashboard provides an eye-catcher.

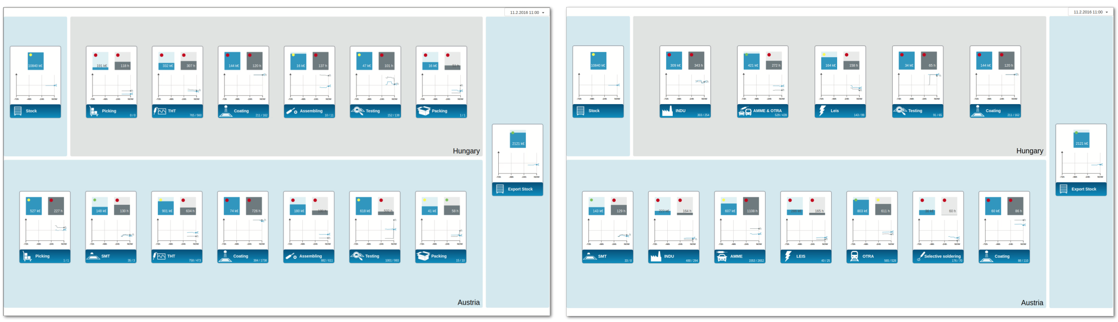


Figure 5.6: Two layouts reflecting the production flow.

5.4.4 Mobile Design

In addition to the desktop layout, two different mobile layouts have been proposed shown in Figure 5.7. In the mobile dashboard the amount of information is reduced to a minimum due to limited amount of display space. Each station is displayed in a line according to the processing order within the factory. A station can either have one or two data fields equal to the desktop layout and shows the development of the data fields over time. Colour encoding is again used for showing the status of the station's data fields. In the left layout, the width of each bar within the rectangles indicates the time span between two measurements; the height of a bar provides information about the actual value. Circles next to the name provide information about the current status, which is always equal to the colour of the last bar. The right layout shows the same information displaying the development using a line chart within the rectangle. Small circles indicate the measurement times. As the line chart did not outline potential problems enough, the left layout was selected.

The mobile layout allows managers to monitor production and potential bottlenecks when travelling. It provides all necessary information to identify critical situations and enable managers to take decisions.

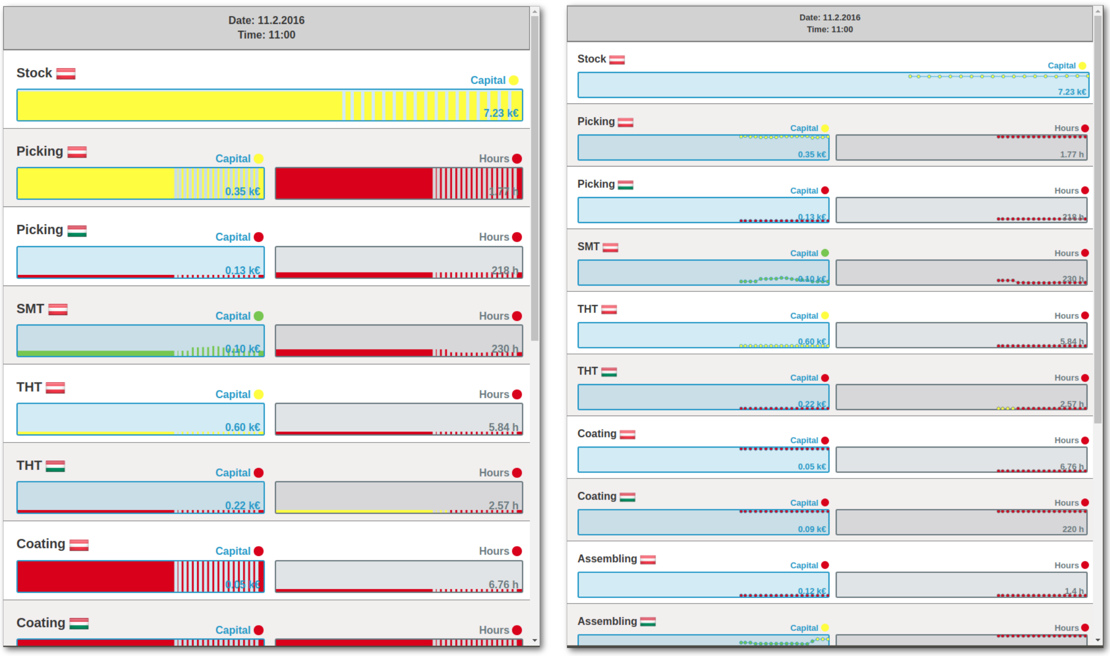


Figure 5.7: Two layouts for mobile devices showing most important data fields in production.

5.4.5 Animation Design

In order to enable managers observing the real production, the provided dashboard animates the production flow between two measurements. Per default the change during the last two measurements is animated but the dashboard allows the selection of previous time points.

The animation starts at the *Stock* where money in terms of components are first deposited and then obtained and distributed to different station. Thus, the amount of money available at the *Stock* first rises by adding components and then declines and animated towards stations which receive the components. Animation of input is indicated by moving rectangles from the upper left of the station, where incoming arrows appear, to the top of the squares where they drop in. Animation of output is shown by rectangles, which move up the squares and are animated towards the upper right of the station where the outgoing arrows origin, see in Figure 5.8. Before components actually leave the station, the station itself is updated in terms of money (and hours if available). The statuses *okay*, *borderline* or *critical* of the two data fields are updated as well as the line chart, containing the new values, shown in Figure 5.9. Afterwards, components are animated along the arrows indicating the production flow. The thickness of each arrow corresponds to the number of products in terms of money, which are transferred from one station to another. At the receiving stations the input is animated to the two squares indicating money and hours, see Figure 5.10. Depending on the number of products transferred from one station to the next, the number of moving rectangles deviates. This procedure is repeated for each station separately, stations with the same name are animated at the same time.

The production flow is only visible for the current station and hidden afterwards as otherwise too many connections would be displayed at the same time, which confuse the observer when components move from a source station to one or more target stations as the number of potential connections is defined by a fully connected network. Furthermore, the thickness of each arrow depends on the current production flow and changes over time. Connections popping-up allow clear a structure without overloading the dashboard with too much information and therefore keep it easy to understand.

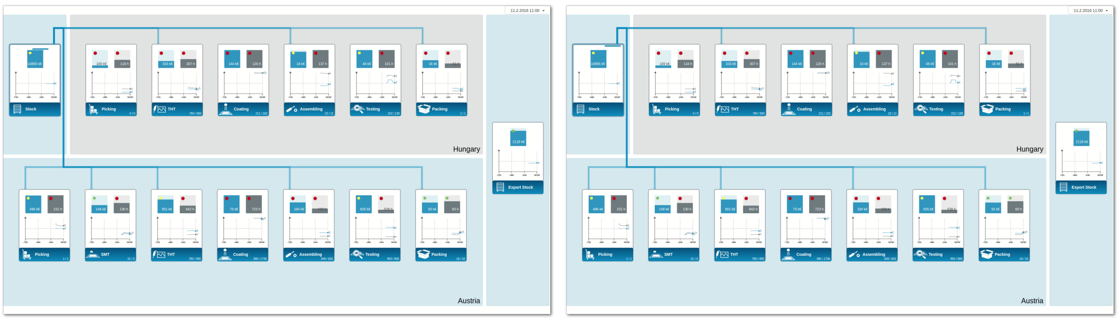


Figure 5.8: The two images show the animation of output from the *Stock*. Elements are animated from the square to the upper right where the outgoing arrows originate.

5.5 Case Study

The following description provides an example of how managers can use the dashboard in order to investigate the status of each station as well as the production flow.

Assuming that a manager wants to investigate whether everything worked out as expected or whether bottlenecks or other problems arose during the last production phase. There-

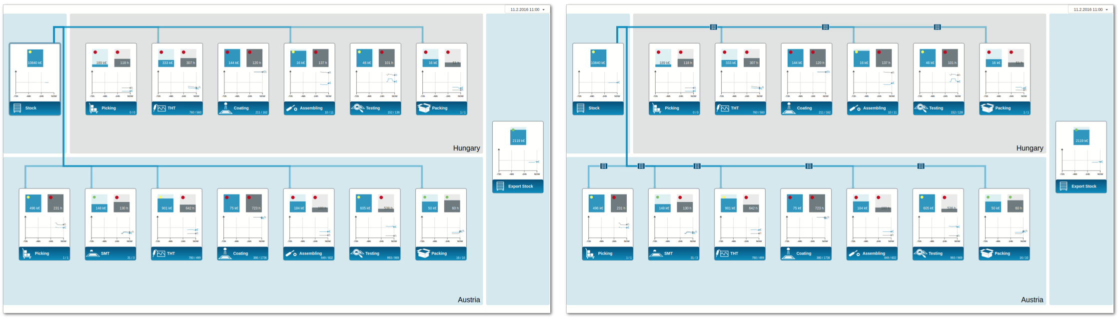


Figure 5.9: After animating the output of the station, the data fields within the station are updated including the overall production value, values for money and hours, their corresponding status and the line chart containing these values (left). Outgoing icons are animated along the paths indicating that components move from one station to others (right).

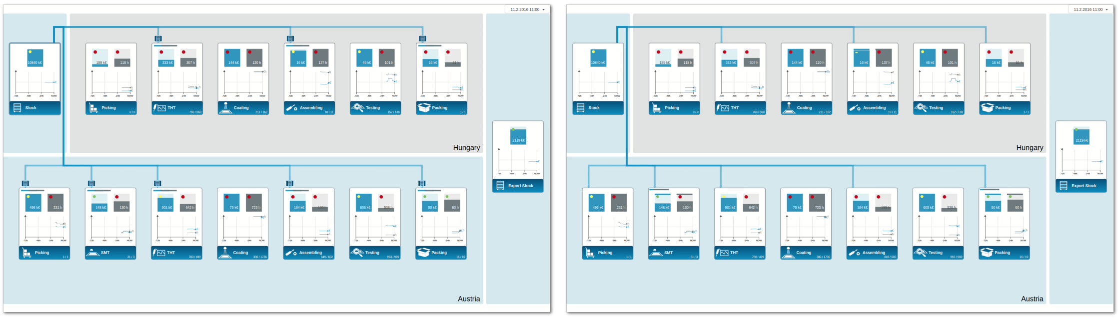


Figure 5.10: When the icons arrive at the target stations, rectangles are again animated from the top left to the corresponding squares. The number of rectangles dropping in the squares corresponds to the amount of money or hours arriving at this station accordingly.

fore, the manager opens the dashboard and investigates all stations before animated to observe the initial status. Figure 5.11 shows the status of all stations.

Investigating the *Stock*, the manager can see that the amount of money bound to this station is rising towards the upper limit of the station, as the fill status of the square is almost complete. The yellow circle additionally shows, that this value is already outside of the optimal range. The status indicates that there should no further assembling components be added to the *Stock*. The line chart indicates that the value has been constant over the last few hours.

Continuing with the *Picking* station in Austria, values for money and hours are at their upper limit and therefore a potential bottleneck. In contrast to this, for the same station in Hungary, both values are on the lower side of the optimal range. Therefore, both circles are displayed red. As the *Picking* station is only responsible for gathering components for being forwarded to other stations, values for this station at the lower limit are more critical than on the upper limit. Line charts again show, that the values were not changing much except the number of hours at the station in Austria was decreasing during the previous hours.

Figure 5.11: The initial status of all stations before animating the latest production phase.

The money value for the *SMT* station lies in the bottom half of the optimal range whereas the hours spend for processing is in the upper half of the optimal range close to the overall optimal average value. While the colour for money is displayed green, the one for hours is shown red. This might be due to sensitive definition of the colour coding for this station. The line charts show that the values rose significantly between two measurements within the last 24 hours and then again decreased slowly.

The next station, *THT*, is again available in both locations. Both values for the location in Hungary are comparable low and coloured red. Line charts show that the values were constantly decreasing during the last hours. Values for hours have about the same fill status considering the optimal range, whereas the money bound to the station in Austria is significantly higher than the one in Hungary.

Data values for both *Coating* stations are at their upper limit, indicating that they might be bottlenecks within the production process. In order to investigate whether this comes from downtime or from more components arriving at the station than being able to process the manager needs to investigate the ratio between components processed in the previous step versus expected number of processed components in the lower right of the station. The values for Hungary show that the performance was higher than expected (221/162) whereas the results in Austria show that the performance was comparable low at about 22 percent (380/1736). This indicates that there were some problems including downtime in Austria. The lower values in *SMT* and *THT* might come from the bottlenecks in *Coating* as the production was reduced in order to not bind too many hours and money to one station.

Assembling is the next station within the production process. Again, this station is available in both locations. While assembling data values in Austria are quite low, those in Hungary are towards the upper limit. Lower values in Austria might arise due to lower production values in the previous *Coating* station. Therefore, fewer components are forwarded to this station.

The *Testing* station in Hungary shows that both values are at their upper limit whereas they slightly decreased within the last few hours. Production values show that the performance was higher than expected. The line charts indicate that there were some problems before as there was a significant rise especially for the money several hours before but latest production values show that everything seems fine. In Austria the values are opposite. While the amount of money bound to the components for being processed reaches the upper limit, the number of hours for processing them moves towards the lower limit. This might also come from problems in the *Coating* process. Production values show that the number of results produced within the last period is higher than expected.

The *Packing* station shows that while values in Hungry are very low, decreasing during the last hours, values in Austria are within the expected range, rising during the last hours. Nevertheless, processed units are in the optimal range or even higher for both locations.

Finally, the *Export Stock*, again only located in Austria, shows that the amount of money bound to the final products is very high but still within the optimal range. This is indicated by the green circle.

After investigating the initial status, the animation flow and the thereof resulting data values can be investigated. Figure 5.12 shows the resulting values after animation. It shows, that hardly any data values changed compared to the initial situation. The amount of money for *Testing* in Hungary slightly decreased. Additionally, both values changed for *Coating*, *Assembling*, *Testing* and *Packing* as well as the amount of money in the *Export Stock* in Austria. As the time range between the two measurements of the production data was only one hour, there were only slightly but no significant changes.

Figure 5.12: The status of all stations after the animation was performed.

Additionally, the manager also wants to investigate the animation flows from each station to the others shown in Figure 5.13 to receive further insights. The process flow from the *Stock* to other stations seems to be appropriate considering that mainly *Coating*, *Testing*

Figure 5.13: The thickness of each connection shows the number of products in terms of money being forwarded to connected stations.

and *Export Stock* are not served. No products were passed forward from the *Picking* station in Hungary, whereas the one in Austria only forwarded components to *THT*. As the arrow is quite thick; this indicates that the amount of money bound to these products is comparable high. *SMT* only forwards components to the *Testing* station again binding a higher amount of money. In comparison to previous stations, *THT* passes fewer components to the *Testing* station in Hungary and the *Coating* station in Austria. The *Coating* station does not forward any components during the investigated time range. As a result of this, no arrows are displayed. Both *Assembling* stations forward components to the *Testing* station, whereas the one in Hungary binds a higher amount of money. The *Testing* station forwards its components to *Coating* and *Packing* in both locations and to the *Assembling* station in Austria. After *Packing* the components, they are again forwarded to *Testing*. This deviates from normal behaviour, as this station normally forwards its results to the *Export Stock*. The *Export Stock* neither receives nor forwards any components within the investigated time range.

If this information is not sufficient for the manager, selecting an arbitrary time in the past and performing the same analysis steps is supported by the dashboard.

5.6 Software Design and Technical Information

The proposed dashboard was implemented in HTML5 and JavaScript [138]. Consecutive simple animations are implemented using D3 [75] allowing to reconstruct the production process. A minor server for running the application and serving requests for the layout as well as for processing data is required. Therefore, Apache Tomcat [4] was selected. The client always requests the latest data files containing new data values from the server. The layout of the dashboard can be easily configured and adjusted according to changes within the factory by modifying the JSON (JavaScript Object Notation) file, no changes within the code are required. A sample JSON file containing the layout is proposed in Listening 5.1. In order to retrieve the latest production data, an external service has to calculate the corresponding values and push the values in a predefined structure to the server. An example for this is proposed in Listening 5.2. The dashboard implementation supports all common browsers including Chrome, Firefox and Internet Explorer. The mobile interface has been tested and optimized for running on Blackberry mobile phones.


```
{
  "screenGridX": 223,
  "screenGridY": 314,
  "maxDaysDisplayed": 2,
  "maxInputOutputAnimation": 5,
  "inoutduration1" : 800,
  "inoutduration2": 300,
  "inoutdelay": 400,
  "inoutanimationheight": 5,
  "pathanimation": 8000,
  "pathshowhide": 1000,
  "trafficanimation": 4000,
  "keepVisFinished": 5000,
  "displayedDates": 72,
  "displayedDatesDetail": 672,
  "animateImgWidth": 26,
  "addemptydata": 70,
  "fillUpWithEmptyData": false,
  "keepPrevValues": false,
  "localStorageLocation": "seidel",
  "departments":[
    {
      "departmentId": 1,
          "name": "Oesterreich",
          "label": "Austria",
      "position": {
                 "x": 0,
                 "y": 1.7
          },
          "grid": {
            "x": 7.55,
            "y": 1.65
          }
        },
    ...
```

```
  ],
  "workstations":[
    {
            "name": "Stock",
            "workstationId": 1,
            "departmentId": 1,
            "animateOrder": 1,
            "capital": true,
            "hours": false,
            "statusInfo": false,
            "location": "A",
            "locationIcon": "img/austria.png",
            "position": {
              "x": 0.1,
                  "y": 0.43
            },
            "size":{
              "x": 0.8,
                  "y": 0.8
            },
            "icon": "img/workstation/Lager.png",
            "moveIcon": "img/move3/Lager.png",
            "pathColor":"#0092ce"
    },
          ...
  ],
  "connections":[
    {
            "workstationIdStart": 1,
            "workstationIdEnd": 2,
            "positions":[
              {
                    "x": 0.8,
                    "y": 0.43
              },
                  {
                    "x": 0.8,
                    "y": 0.25
              },
                  {
                    "x": 0.95,
                    "y": 0.25
                  },
                  {
                    "x": 0.95,
```

```
                  "y": 1.8
                },
                {
                  "x": 1.39,
              "y": 1.8
            },
                {
                  "x": 1.39,
                  "y": 2.05
            }
      ]
        },
        ...
  ]
}
```

Listing 5.1: Example of a dashboard configuration JSON file allowing changes according to adjustments within the factory.

```
{
  "time":1455188419,
  "workstations": [
    {
      "workstationId": 1 ,
      "input": {
        "minHours": 0 ,
        "maxHours": 0 ,
        "curHours": 0 ,
        "statusHours": "OK" ,
        "minCapital": 5000 ,
        "maxCapital": 50000 ,
        "curCapital": 7000 ,
        "statusCapital": "MARGINAL"
      },
      "results": {
        "expectedResults": 20 ,
        "actualResults": 19 ,
        "statusResults": "OK"
      },
      "output": [
         {
           "workstationId":2,
           "number":0,
           "capital":200,
           "hours":0
         },
```

```
                {
                    "workstationId":3,
                    "number":0,
                    "capital":300,
                    "hours":0
                },
                ...
            ]
        },
        ...
    ]
}
```


Listing 5.2: Example of a production output JSON file used for animating the results for each station within the dashboard.

5.7 Discussion

This section discusses whether and how the guidelines defined in section 2.4 were considered for this dashboard.

G1 (information): During the first phase, most important data fields were identified for being visualised within the dashboard to enable users to monitor current production values. In order to monitor the production flow and allow interventions if bottlenecks or other unexpected events occur, initial plans included a data update of every ten minutes, enabling managers to receive the latest status. It was discovered that there are not many changes within the defined period of time as large parts of the production are not automated using machines assembling components but components are assembled by humans. Therefore, in order to find a trade-off between receiving the latest measurements and providing a meaningful time interval, the data was updated twice a day. Therefore, this guideline was fulfilled during the station design phase.

G2 (visualisation) and G4 (clutter reduction): During several iterations, a project symbol was designed matching all requirements and expectations. Thus, the project symbol conveys most important data fields using simple visual components, while less relevant information is not included. Connections between two stations are only shown during the animation process enabling users to investigate relevant information and thus reducing clutter. Both guidelines were considered for this dashboard.

G3 (UI): Stations within the dashboard are aligned along a grid. Furthermore, they are positioned according to their processing order enabling users to easily analyse process flows and important data values as the dashboard only contains most relevant measurements. The dashboard enables users to monitor production and analyse changes over time. Nevertheless, for analysis, it is sometimes helpful to freeze the dashboard or replay single stations. However, the provided dashboard does not offer these functionalities. It only offers the opportunity to restart the animation process from the beginning, which may complicate analysis processes. To investigate current values without animating changes from last data update, users can resort to the mobile application, which does not contain any animations. However, providing the possibility to stop the production animation and resume on demand, was not required, therefore, this guideline was fulfilled.

G5 (interaction): In terms of interactions, the dashboard provides a limited number of most important interaction possibilities. By clicking on the line chart of a certain station, the corresponding chart is enlarged and shows a longer period of time. Other than this, the dashboard only enables users to select a certain measurement time and animate the production flow. The dashboard does not enable users to pause the animation flow or go back to a certain point within the current animation, as this was not a requirement, however, it would simplify the analysis in some cases. Therefore, this guideline was fulfilled for the discussed dashboard.

G6 (overview-details): The dashboard provides all relevant information; only measurements within the line chart can be investigated for a longer period of time. Thus, the dashboard hardly supports users in drilling down to receive detailed information. Therefore, this guideline barely fulfilled, however, it was not a requirement on the dashboard.

G7 (colour): In general, colours within the dashboard are blue and grey, only the status of money and hours are displayed using traffic lights - green (okay), yellow (borderline) or red (critical) depending on their status. Thus, using only few colours enables users to identify stations, which are critical more easily. Using only few colours for highlighting most relevant information, this guideline was fulfilled for the dashboard design.

G8 (customisation): By proposing a dashboard containing simple geometric shapes, which reflect the corporate design of the project partner, enabled the employees to identify themselves with the final result. The dashboard created in this project was designed according to the specific requirements of the partner but can be adjusted to meet requirements for different use cases, by replacing the station symbol by other visual elements. Therefore, this guideline was fulfilled.

G9 (devices): As one of the main requirements within this project was to provide a simple mobile dashboard for monitoring production data, and the animated dashboard was not suitable for being displayed on small devices, a separate mobile layout was designed showing most important data values while not providing any interaction possibilities. Therefore, this guideline was also fulfilled for this dashboard.

G10 (evaluation): The provided dashboard was not evaluated in a case study to investigate whether it matches all requirements by performing a user study. However, feedback collected from the project partner during multiple design cycles revealed that they appreciated the design. Therefore, this guideline was not fulfilled.

5.8 Conclusion

The proposed dashboard was designed and developed during several iterations in close cooperation with the project partner. Therefore, creating a design, which reflects the corporate design of the company as well as projecting the production flow in the animation process, allowed them to identify themselves with the final result.

The proposed dashboard allows monitoring the production flow as well as the fixed capital and hours at each station, outlining potential bottlenecks. Thus, managers can immediately react on arising bottlenecks at certain stations. The mobile application allows remotely monitoring the production status without animation. Additionally, the dashboard is shown on large screens in the production plant providing an overview of their production process as well as on their current productivity to visitors. During the second phase of the project, a dashboard editor allowing to configure new layouts from scratch or modifying existing layouts without editing the JSON configuration file, was proposed. The mock-ups for the dashboard editor are added to appendix B.1.

Chapter 6

Large-Scale Time Series Data Analysis

6.1 Introduction

This chapter discusses a dashboard especially designed for analysing large-scale time series data in industry environments. Thus, a server and a client component were developed allowing users to load different data resolutions depending on the selected time range and zoom level. This enables users to receive an overview of the data when investigating long time ranges by displaying data in a low resolution, while analysing data in detail when zooming in or selecting a short time range in a high resolution.

The author of this thesis was responsible for the conceptual design of the dashboard and mainly implemented the dashboard and server component.

The design and implementation of this dashboard has been a collaboration of several people. The design of this dashboard has been proposed and mainly implemented by the author of this thesis. Peter Marton and Santokh Singh have implemented some features within the dashboard. The implementation of the back-end was mainly performed by the author of this thesis and Ralph Wozelka. Both, design and implementation phase, have been supported by Vedran Sabol. The underlying format for storing data has been proposed by Stefan Körner and implemented by Werner Schmierdorfer. The author of this thesis thanks all involved persons from Know-Center for collaborating in this project.

6.2 Motivation

In production plants, usually many sensors are employed to monitor production. As the number of sensors is constantly increasing in high technology environments, monitoring and analysing all results is not feasible and often not necessary. Investigating specific sensors is only required if results from the production plant do not have the expected quality. Assigning these low-quality products to specific production processes and therefore measurements is another challenge. This requires access to sensor measurements but also tools enabling users to investigate them. Therefore, the main motivation for this project was to provide a highly scalable tool for analysing any sensor data enabling users to cope with a huge number of measurements and identify those measurements, which lead to low quality products or rejects. By identifying those measurements, dashboards can be extended to raise awareness if certain measurement constraints are met and enable production workers to take actions immediately to avoid low quality products or rejects.

6.3 Requirements

As production plants have employed many sensors usually measuring values on a high frequency level, the number of measurements generated within a short period of time, is enormous. Therefore, the need for storing and accessing any measurement on demand is an essential requirement. Due to the large number of sensors, only most important sensors are monitored during production. Thus, many sensors are usually not investigated unless problems occur. Then, users need to investigate sensors connected to the production plant, which potentially cause the problems. Therefore, a tool allowing users to select a set of interesting sensors and visualise selected time ranges, is required. Additionally, users may want to investigate measurements over a longer period of time to identify and analyse problems occurred in the past. Therefore, a major demand on the dashboard is to enable users to analyse large time ranges. This can be performed by displaying sensor data in low resolution, whereas small time ranges can be displayed in higher resolution allowing users to analyse sensor measurements in detail. This simplifies the analysis process by enabling users to easily compare measurements over a longer period of time and thus, observing whether upcoming problems are already announced by deviating measurements before they occur.

6.4 Dashboard Design

This section discusses the design process for a dashboard enabling users to analyse large-scale time series data. The initial design considered three main views, the sensor selection, the sensor visualisation and the favourites. Furthermore, a view for providing a signal search was considered. Figure 6.1 shows two views, one for sensor selection (left) and one for managing favourites (right). The first one should enable users to select a set of sensors, the investigated time range as well as the corresponding visualisations. The favourite view shows dashboard configurations frequently used and therefore stored by the user.

Figure 6.2 show two mock-ups for visualising line charts as well as horizon charts. This view contains three main areas. On the left, users can select sensors, time ranges and

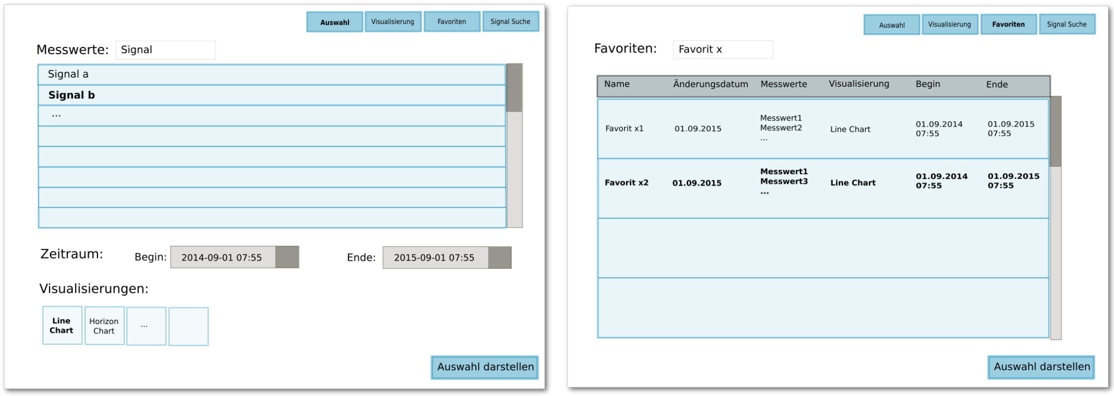


Figure 6.1: Mock-ups for a sensor data dashboard showing selection (left) and favourite (right) view.

visualisations similar to the selection view, on the right, favourites are displayed, and, in the centre, the selected charts are visualised. Both areas, the left and the right one, can be minimised on demand to provide more space for visualising data. The left image in Figure 6.2 show the proposed line chart. Several sensors can be displayed either in one chart using multiple axes or separately, each sensor in one chart. User defined annotations are highlighted within the line chart and can be assigned to certain problems or unexpected events. The right image in Figure 6.2 shows the same dashboard displaying sensor measurements using horizon charts. While saving space, they convey the same information as line charts do. In contrast to line charts, each horizon chart can only display one sensor. Additionally, values are not displayed for the y-axis. This is not possible, since the same space is assigned to multiple values depending on the colour at each point in time. Higher values are displayed by overlaying multiple layers, each of them darker than the previous one. Only tooltips can provide the actual values for each point in time.

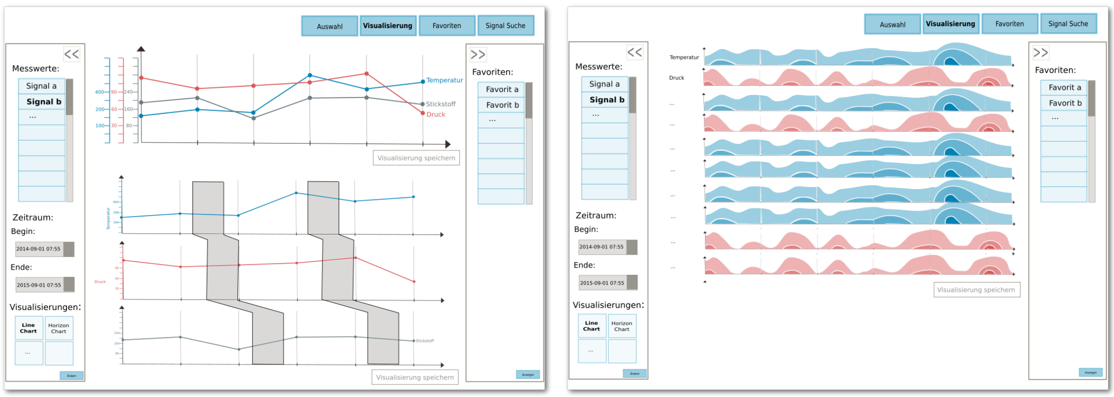


Figure 6.2: Mock-ups for a sensor data dashboard showing line charts (left) and horizon charts (right) [106].

As the visualisation view already contains an area for sensor selection and favourites in a suitable way, the selection and favourite views, proposed in Figure 6.1, were not required and thus discarded. Additionally, after prioritising all features, the signal search was not included in this dashboard but was implemented in a separate dashboard for another use case. Instead, a filtering feature was added to the dashboard enabling users to filter for selected value ranges.

Finally, the proposed design was considered to fulfil all requirements and therefore implemented as shown in Figure 6.2. All three areas, sensor selection, favourites and visualisation, were refined and implemented. The favourite area was moved from the right to the left enabling users to easily switch between the two tabs, sensor selection and favourites. Thus, additional display space for data visualisation is available.

As both, line charts and horizon charts, contain the same information but provide different representations, the line charts were chosen due to higher readability, for both, single sensors as well as multiple sensors within one chart. Additionally, correlation charts were investigated and considered as being suitable for analysing correlations between two sensors and were therefore integrated into the dashboard. The final dashboard is discussed in the following in detail, see Figure 6.3.

6.5 Dashboard Description

This section discusses the final implementation of the dashboard considering all visualisations and interaction possibilities.

6.5.1 Sensor Selection Area

The sensor selection area consists of several parts:

- Data source selection: a drop-down menu allows users to switch between different data sources, each containing a separate set of sensors. Different data sources can be used for different production plants.
- Add data columns: users can add new sensor by providing a name, selecting one or two existing sensors and applying an operation on their measurements. There are five operations available, *addition*, *subtraction*, *multiplication* and *division* of two sensors as well as *first derivation* of one sensor.
- Sensor search/filter: the search feature enables users to search for a sensor name in the list of all available sensors, the filtering feature only shows those sensors, which are currently selected.
- Sensor selection: allows selecting sensors for being visualised.
- Sensor value filtering: this feature allows users to restrict the displayed data for the selected value range of a specific sensor. Measurements, which are outside the selected value range, are displayed as missing values in the line chart. Multiple filters of different sensors are *AND* connected.

- Time range selection: in order to visualise sensors, a certain time range needs to be selected. Depending on the time range, different data resolutions are loaded for visualising.
- Visualisation selection: there are three visualisations available. Combined line charts show all selected sensors in one chart, each using a separate axis for its value range, and single line charts, where each sensor is displayed in a separate line chart. The third option are correlation charts. They are only available if two sensors are selected and show the correlations for common timestamps. Due to limited capacity, close measurements are combined to one data point, whereas the number of measurements is encoded by the size of the data point.

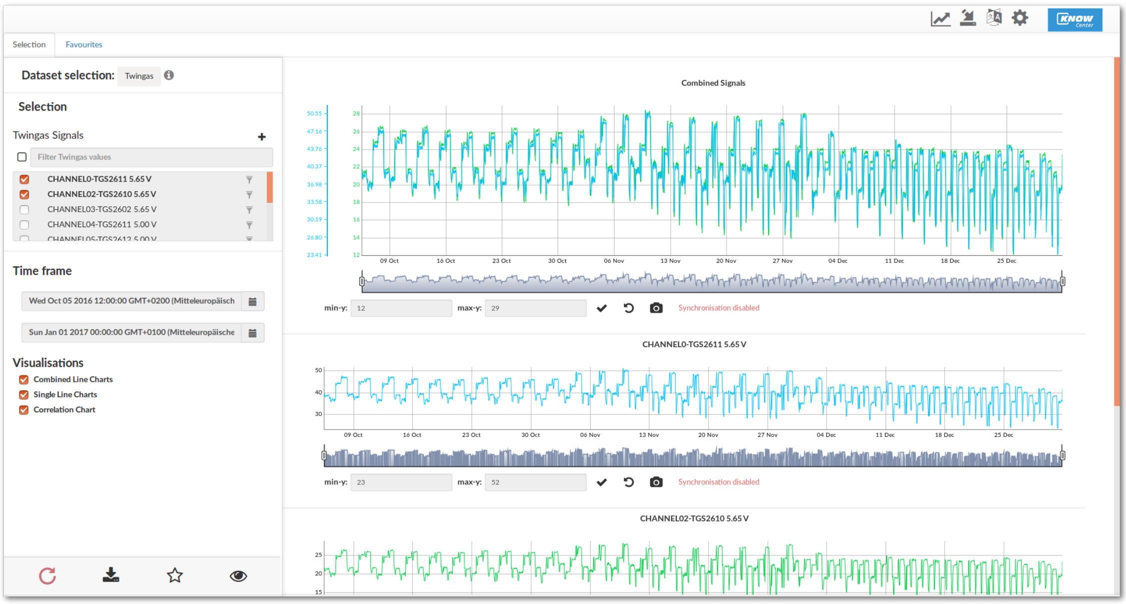


Figure 6.3: Dashboard for analysing large-scale time series data.

6.5.2 Visualisation Area

6.5.2.1 Line Chart

Line charts, see Figure 6.3, have been identified as being suitable for displaying and analysing time series data in general and sensor data in particular. In order to display sensor data, the well-designed canvas-based Dygraphs [12] library was chosen. It allows users to visualise large-scale sensor data. The library was published by Dan Vanderkam under the MIT (Massachusetts Institute of Technology) license [18] and provides several functionalities out of the box. Furthermore, functionalities can be overwritten and extended by developers depending on the requirements.

Since the number of available pixels, and therefore data points, is limited on a screen, it is not possible to show more data points than pixels are available. Furthermore, the performance dramatically decreases if too many data points are loaded. Therefore, a down-sampling approach was implemented to only retrieve the number of data points being able to display and therefore improving performance. The smaller the selected time range, the higher the data resolution, which is loaded from the server component. This allows users to get more details on demand, while also being able to get an overview if the selected time range is very long, see Figure 6.4. The algorithm used for this approach is called *Largest-Triangle-Three-Bucket* algorithm and was investigated by Steinarsson [123]. The characteristic of this algorithm is that it preserves the shape of a sensor while reducing the number of data points. Therefore, it is well suitable for visualising time series data. Other algorithms may calculate average values and can therefore dramatically decrease or even remove outliers when visualising. The line chart proposed in this dashboard allows the following interactions:

- Zooming can be performed by pushing the Shift key while selecting the area of interest within the line chart or by using the range selector.
- After zooming in, the original complete time range can be shown by performing a double-click or by using the range selector.
- Panning can be performed by dragging within the line chart or using the range selector.
- Annotation can be created in single line charts by pushing the Ctrl key while selecting the corresponding area. This opens a dialogue to provide further information for the annotation.

In order to include all required functionalities, the dashboard was further extended to support the following additional features:

- Multiple y-axes: for line charts containing several sensors, each sensor uses its own value range to make usage of the whole space of the charts' y-axis.
- Minimum/Maximum value range: for each sensor, users can specify the displayed y-axis value range. In general, this range automatically updates when zooming in or out. If users manually change these ranges, values outside the selected range are not displayed.
- Reset value range: after changing minimum and/or maximum value ranges, they can be reverted to the original value ranges. This enables users to see all sensor measurements within the time range.
- Export chart: this feature allows users to export the displayed area of the chart as PNG (portable network graphics) for presentations or for further analysis.
- Update y-axis: this feature enables users to define whether the y-axis range should be updated automatically when zooming in or panning. If this is disabled, value ranges of the y-axes do not adjust when performing interactions, which may lead to values not being displayed as they are outside the range when zooming out or panning.

- Coordinate charts: this feature allows users to define whether all or selected charts should be coordinated when navigating.

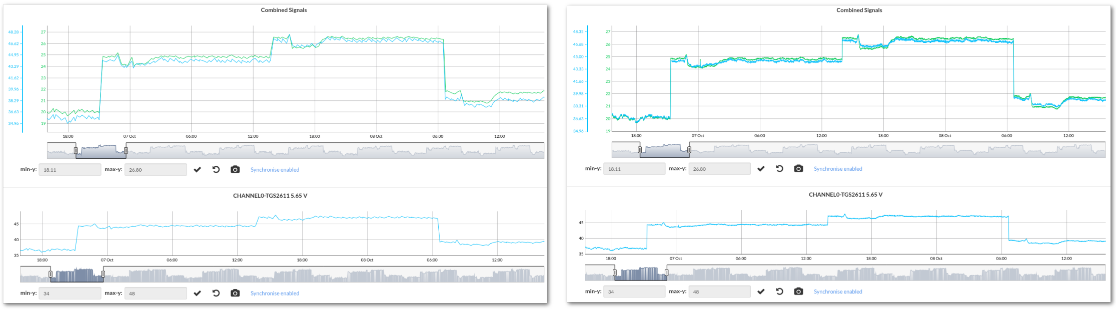


Figure 6.4: Zooming in updates the visualisations by loading more details from the server.

6.5.2.2 Correlation Chart

The correlation chart, see Figure 6.5, displays measurements of two selected sensors for common time points. The correlation chart is based on SVG using D3 [75] library. Compared to canvas-based visualisations, the capabilities of displaying elements in an SVG component are limited. Thus, for very large signals not all data points can be visualised. Therefore, the server-side calculation was extended, combining close measurements to summarised data points. The number of measurements represented by one point within the chart is encoded through the size of each point. In order to meet all requirements, the functionality of the correlation chart has been additionally extended to support the following features:

- Minimum/Maximum y-ranges: this feature can be used to zoom in/out by changing the y-axis ranges. Other than this, no further zooming possibilities are available for correlation charts.
- Reset value range: after changing minimum and/or maximum values, the value range can be reverted to the original range. This enables users to investigate all correlation points within the selected time range.
- Export chart: this feature allows users to export the displayed area of the chart as PNG for presentations or for further analysis.
- Switch axes: this feature allows users to switch x- and y-axis.
- Coordinate charts: this feature allows users to define whether all or selected charts should be coordinated when navigating.

6.5.3 Favourites Area

Besides the selection tab, a tab for managing favourites is available. Favourites can be easily stored by selecting all relevant data fields in the sensor selection area and finally

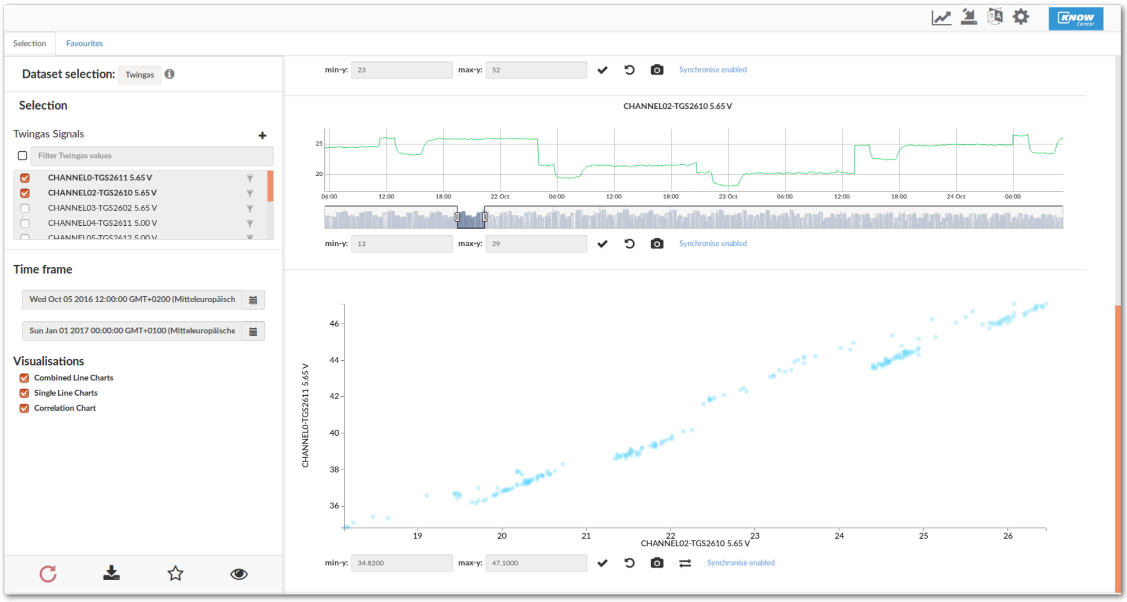


Figure 6.5: Data points within the correlation chart each representing one or multiple measurements.

clicking on the favourite icon to provide a name. Afterwards, the new favourite is shown in the favourite tab and can be easily select by one single click. Favourites can be selected either for visualising or for being enabled in the selection area to allow further modifications before visualising. Favourites can contain one or multiple sensors, and a value range filters as well as the visualisations used for analysis. Time ranges selected in the dashboard are not added to the favourites; therefore, users need to select the corresponding time range before visualising favourites.

To summarise, each favourite contains the following information:

- Name: needs to be defined when the favourite is created.
- Date: automatically generated when creating the favourite.
- Signals: contains a list of sensors selected when creating the favourite.
- Filters: contains a list of filters selected when creating the favourite.
- Visualisations: contains a list of visualisations selected when crating the favourite.

6.5.4 Additional Features

There are some additional features within the dashboard including multi-language support, different dashboard colour schemes and dataset upload. The upload view allows users to create a new dataset and then to upload data files. Figure 6.6 shows the corresponding view for uploading sensor data.

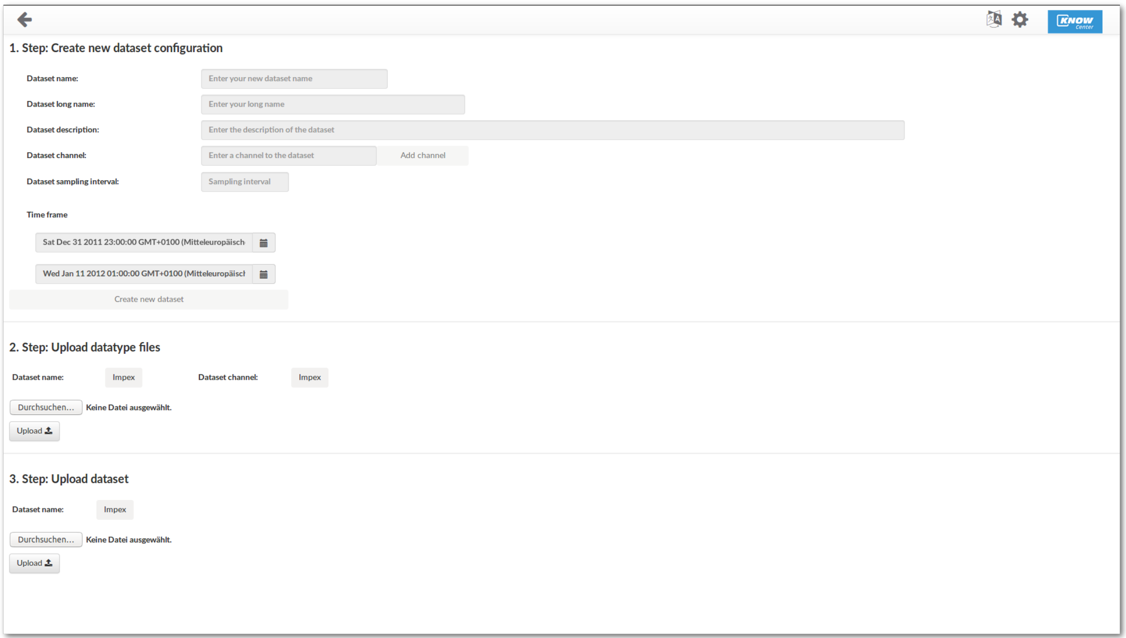


Figure 6.6: Dashboard for uploading new datasets.

6.6 Case Study

Let's assume that a user wants to investigate sensor measurements during production from October 1st 2016 to April 1st 2017. Thus, the user initially selects the corresponding dataset from the production plant containing eight different sensor channels. To receive an overview, the user selects all sensors, the maximum time range as well as the combined line chart, see Figure 6.7. The user can immediately observe that a certain pattern is recognisable. About every 50 days, the peak-to-peak amplitudes of some sensors are increasing slowly and suddenly decreasing again.

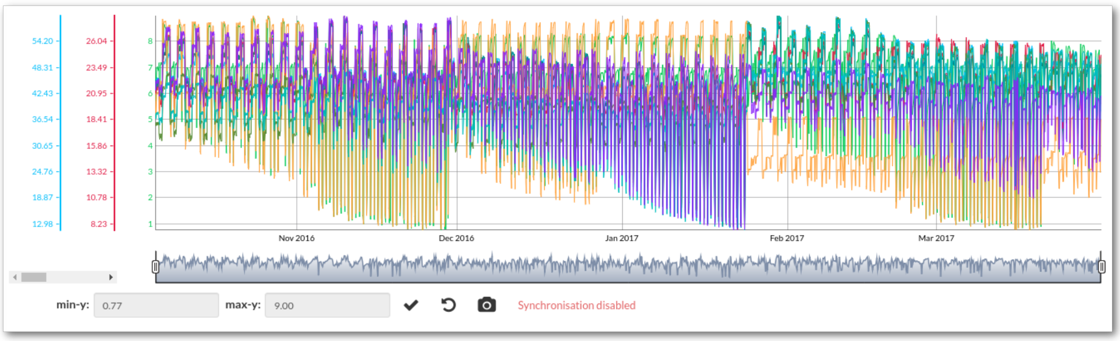


Figure 6.7: Combined line chart showing the development of eight sensors over six months.

The combined line chart is not suitable for analysing many sensors in detail at the same time since sensor measurements overlap each other causing clutter. Therefore, to investigate which sensors peak-to-peak amplitudes are increasing and decreasing, how often and which sensors have high peak-to-peak amplitude values at the same time, the user selects multiple single line charts instead of the combined one, see Figure 6.8. By investigating the displayed sensors, it turns out that slowly increasing measurements, which suddenly decrease and again increase, can be observed for all sensors. For most of them, this happens at about the same time. Only two sensors seem to follow the same behaviour but shifted in time. Additionally, for these two sensors, this pattern is repeated within a shorter period of time.

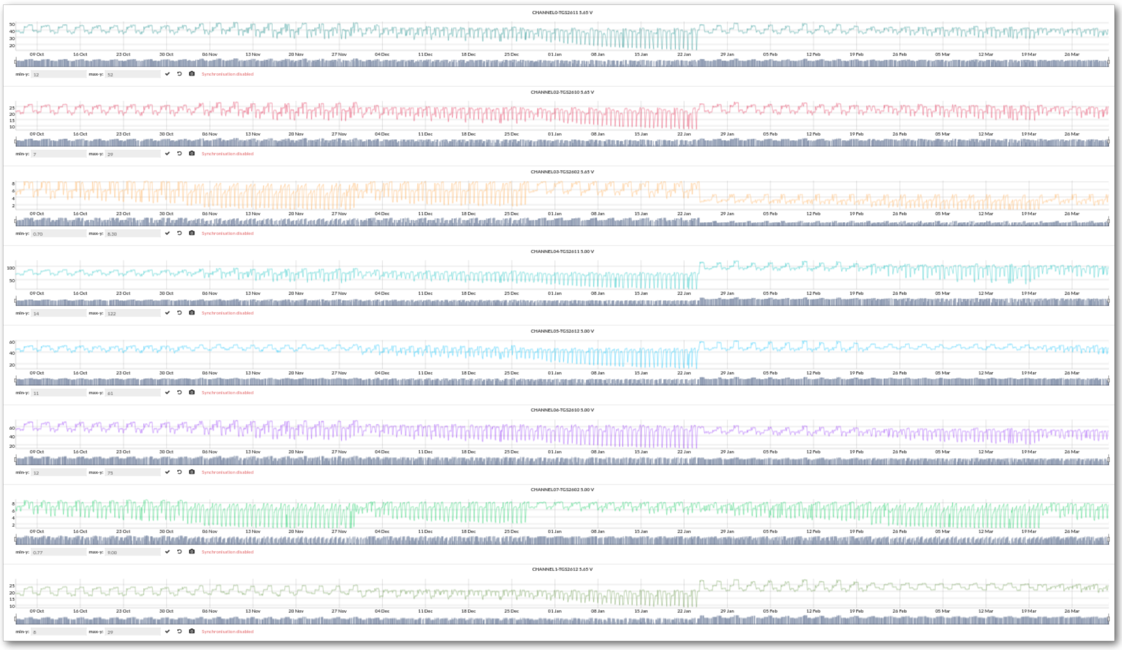


Figure 6.8: Single line charts showing the development of eight sensors over six months.

The user knows from experience that increasing peak-to-peak amplitude values are correlated to the need for replacing certain components in the production plant. As the measurements show, components are required to be replaced on regular basis. After replacing, production and therefore sensor measurements are again within an acceptable range for some time before the peak-to-peak amplitudes are increasing again. Thus, this slow development over time might not have a negative impact on the resulting product as long as components are exchanged on a regular basis.

Afterwards, the user investigates the first period within the six months, ranging from the beginning of October to the end of November 2016. The combined line chart in Figure 6.9 shows the corresponding time range. Here, the user can observe that in this time range mainly for two sensors the peak-to-peak amplitude is increasing dramatically, for all others, changes are not so significant. This can be also observed in the single line charts shown in Figure 6.10.

After zooming in (see Figure 6.10) the user can see, that for each sensor, certain patterns are repeated regularly. Investigating the third (orange) and seventh (light green) line,

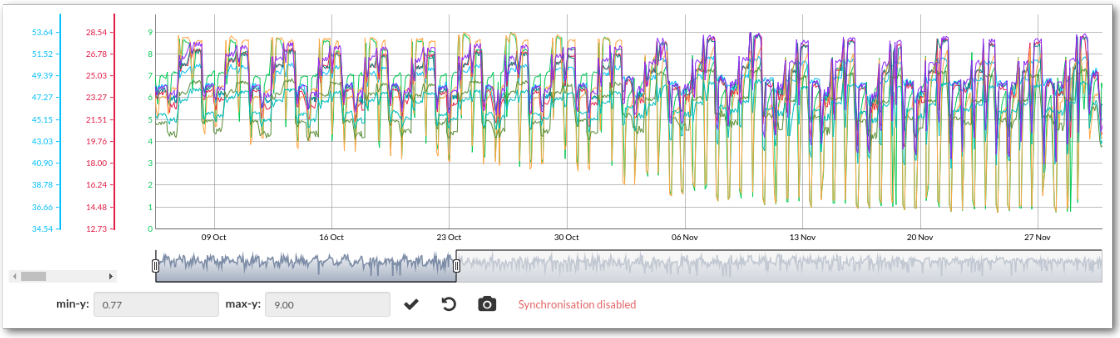


Figure 6.9: Combined line chart showing more details after zooming in.

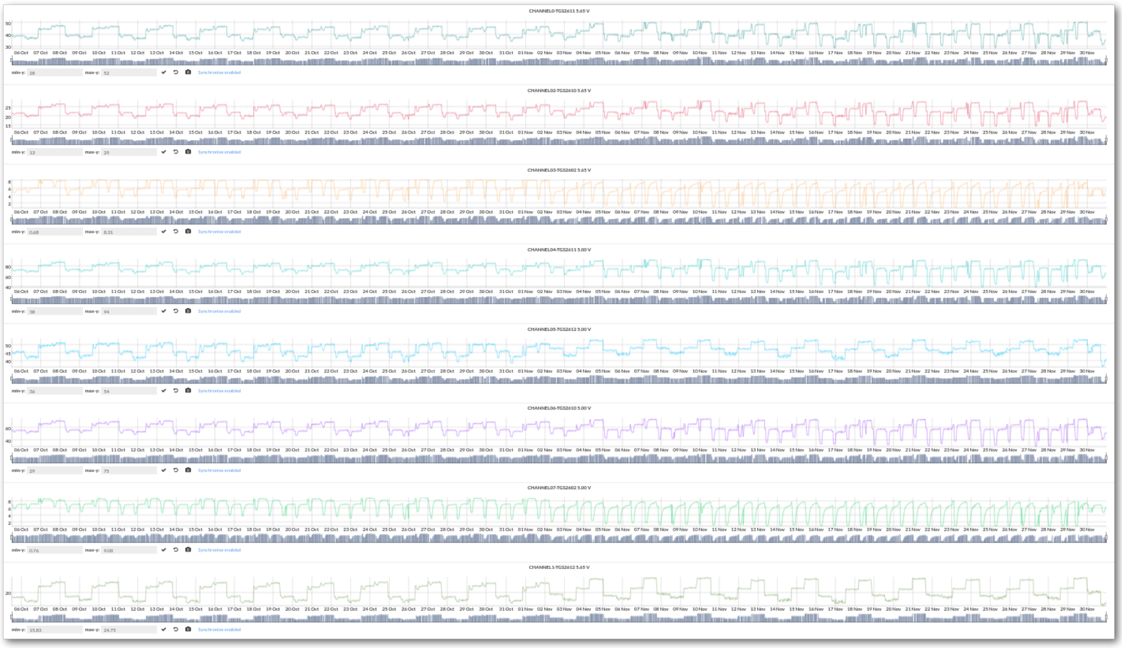


Figure 6.10: Single line charts displaying each sensor separately for better investigation.

some deviations to those patterns can be observed, especially in November. Therefore, the user zooms in to the end of November for detailed analysis. There, the user can see, that those deviations are caused by single measurements. This is shown in the first three images of Figure 6.11. To prove this also for other time ranges, the user selects another time range and receives the same result. This is shown in the lower right image of Figure 6.11. These peaks can be observed in several sensors, but not all. This might be due to coupling of several sensors to related production processes, while others are coupled to other processes. Single outliers within the measurements might be an indicator for failures within resulting products; these products need to be investigated further to ensure that there was no negative impact during the production process.

Afterwards, the user wants to investigate those measurements, where the first sensor in the list has a value larger than 40. Therefore, the user clicks on the filter button next to the sensor name and selects the corresponding value range, see Figure 6.12.

Figure 6.11: Zooming in loads details from the server and displays the corresponding measurements for each sensor separately.

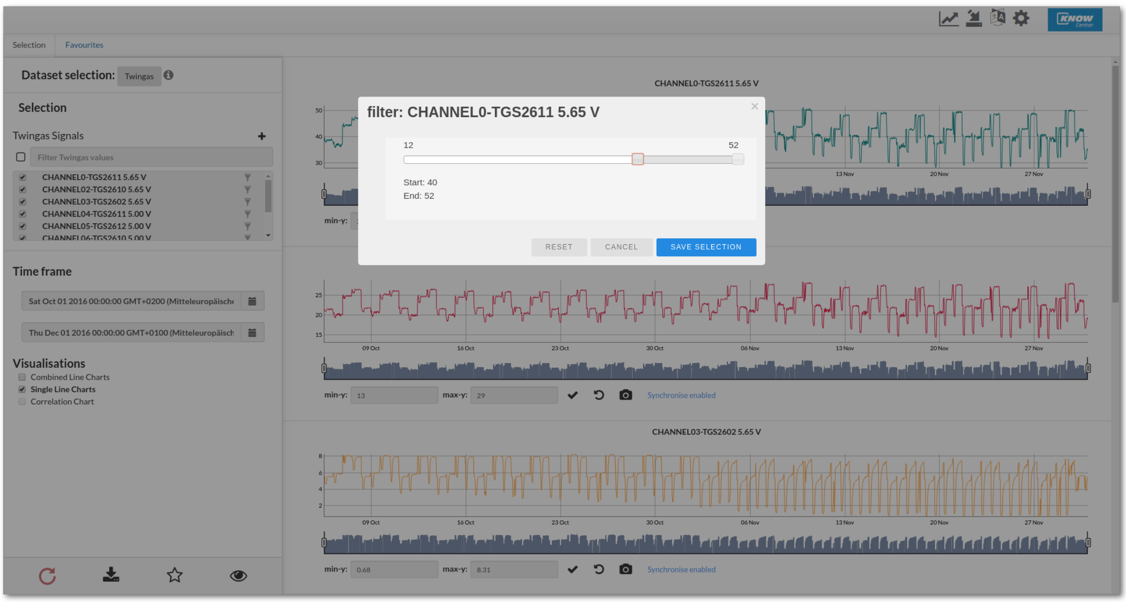


Figure 6.12: Filter range selection using the time series dashboard.

The resulting values are shown in Figure 6.13. The user can observe a certain pattern in October, where the sensor has first lower values and is increasing then. Afterwards, a gap is shown in the line chart. This gap is caused by the filtering feature, removing those values, which are smaller than 40. This pattern changes in November where each pattern first contains lower values, than higher values and afterwards again lower values before the next gap is shown. Similar patterns can be observed for other sensors.

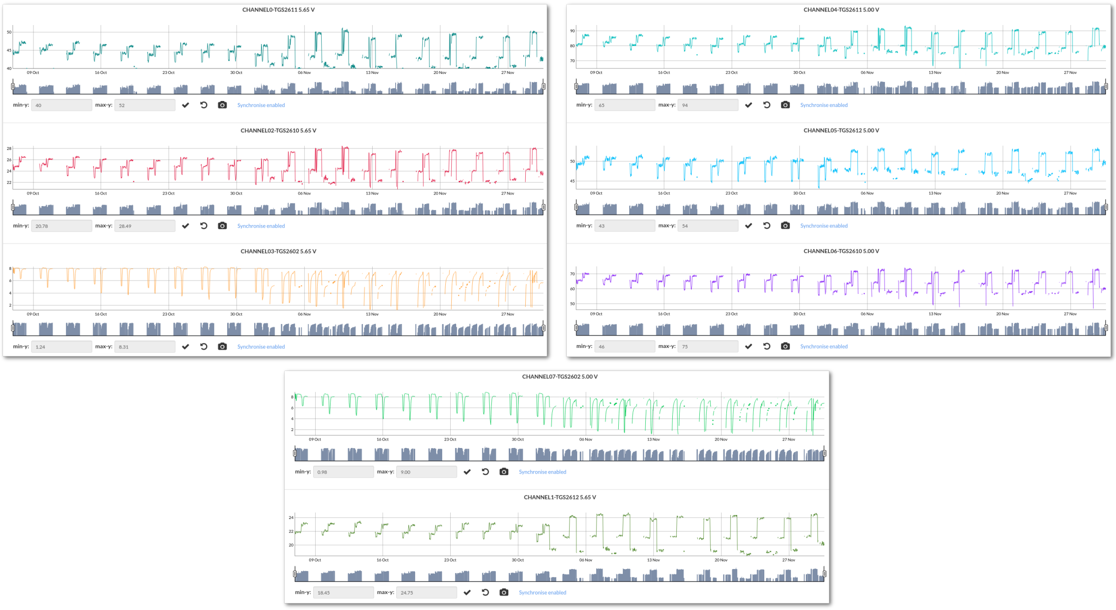


Figure 6.13: Line charts showing measurements after applying a filter.

Furthermore, the user wants to investigate correlations between two sensor measurements. First the user investigates the sensors three and seven. Therefore, the user selects the two sensors and the correlation chart in the selection area and displays the result. The left image of Figure 6.14 shows that there is a positive correlation between the two sensor measurements, observable by the diagonal from the lower left to the upper right. The right image in Figure 6.14 shows that no correlations can be observed for sensor five and seven. Thus, the user can conclude that sensor three and seven are related while sensor five and seven are not related to each other.

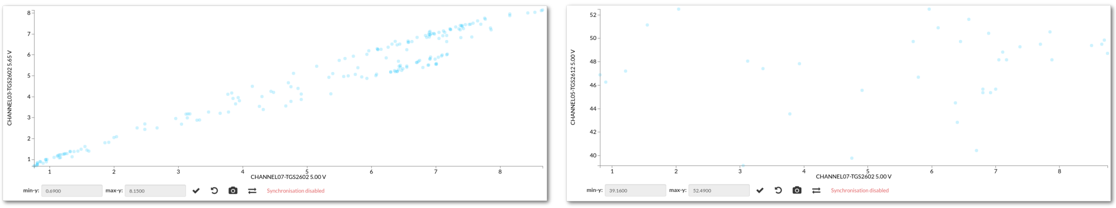


Figure 6.14: Correlation charts showing correlations between two selected sensors.

By investigating measurements in the line chart, the user identified single outliers for several sensors. In order to assign a certain problem identified in the previous process, the user provides an annotation for each outlier. This helps other users to receive further information if similar problems occur later on. Therefore, the user marks the area of interest in the line chart by pushing the Shift key and selecting the corresponding area. After clicking on the *plus* icon next to the orange area, a dialogue is opened which enables the user to provide detailed information, see Figure 6.15. After saving the annotation, it is highlighted blue in the line chart and can be modified or deleted any time.

In order to not always select each sensor separately within the dashboard when investigat-

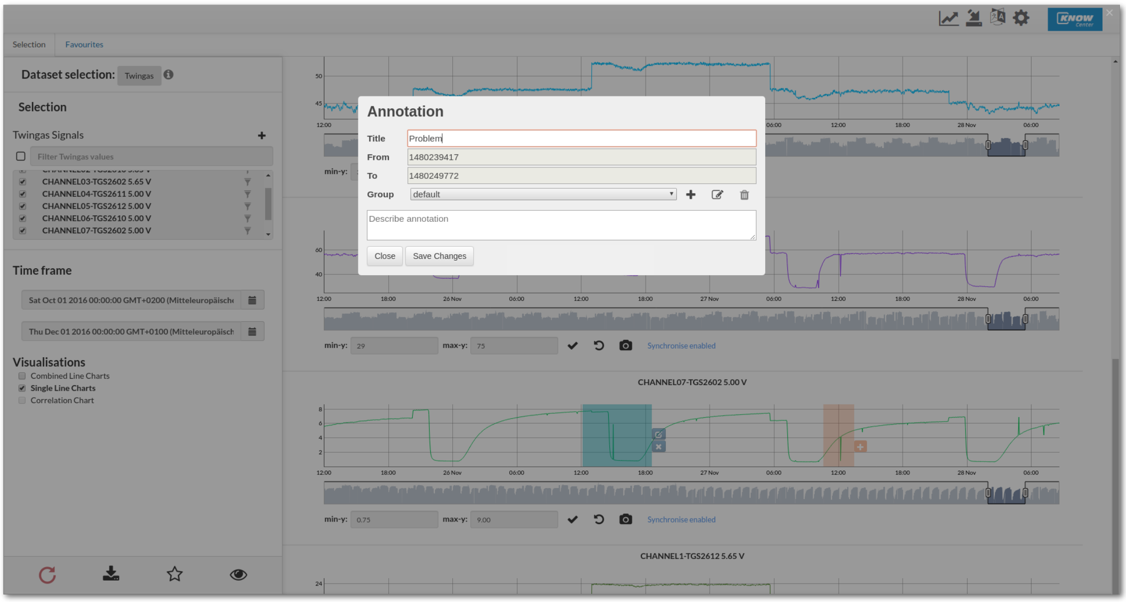


Figure 6.15: Annotations can be created in single line charts and can be used for further processing.

ing new measurements. The user saves the current configuration as a favourite. All stored favourites can be easily accessed by the user, see 6.16.

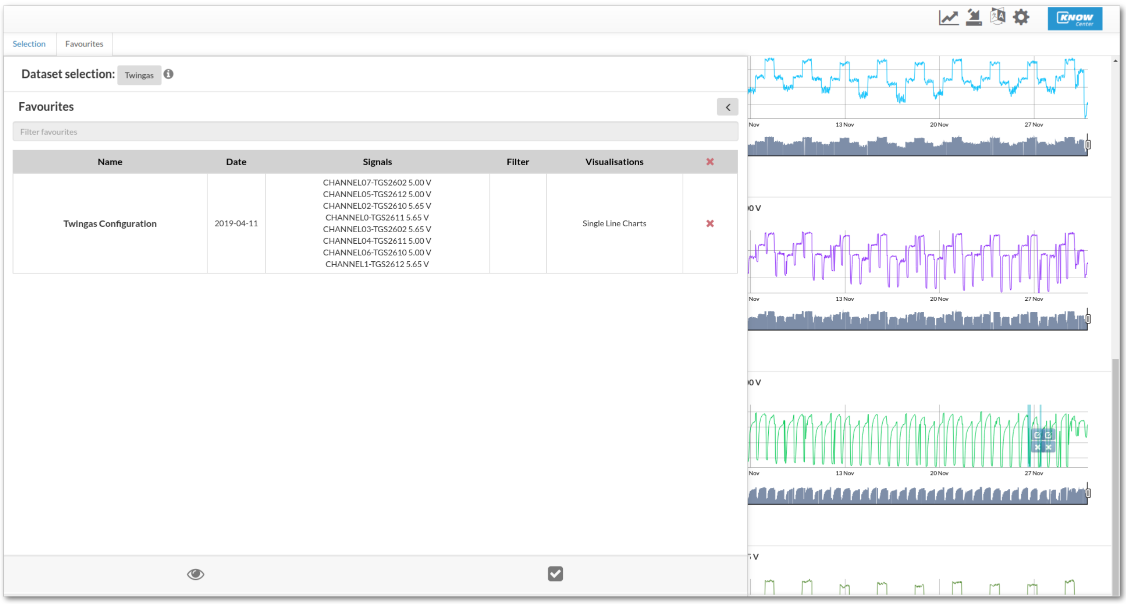


Figure 6.16: Favourite view containing configurations, which allow users to access sensor visualisations faster.

6.7 Software Design and Technical Information

The implementation of the dashboard is based on AngularJS [32]. Additionally, RequireJS [54] was used for loading libraries on demand. For line charts, Dygraphs [12] library is used. It supports multiple functionalities and can be extended on demand. Nevertheless, the most important reason to use Dygraphs was due to high performance while visualising many data points. For the correlation chart, the SVG-based library D3 [75] was selected.

As the number of data points returned to the client is reduced by applying a down sampling algorithm and on-the-fly down sampling calculation take some time depending on the selected time range, the server samples down and stores different resolutions in advance, during the data import process. Depending on the selected time range, the dashboard loads different resolutions from the server within a short period of time.

The back-end is implemented in Java, storing the data in a binary format called RTS2 (regular time series file format 2). This format was chosen since at that time, available time series databases did not yet fulfil all requirements like they do today. In order to allow users to filter numerical time series data, the data is additionally stored in Lucene [127] search engine, enabling users to filter data points which are within the requested value range.

6.8 Discussion

This section discusses whether and how the guidelines defined in section 2.4 were considered for this dashboard.

G1 (information): As this framework was considered to enable users to analyse sensor data, visualising any large-scale sensor data was the main focus of this dashboard. Thus, there was no need for selecting specific sensors. Selecting interesting sensors for analysing should be done by the users depending on their needs. Since the dashboard enables users to select sensors of interest, this guideline was fulfilled.

G2 (visualisation): In order to visualise sensor measurements, there are several options for displaying time series data. For this dashboard the most common visualisation for time series data was selected. Additionally, a visualisation enabling users to identify correlations between two sensors was provided. In order to identify suitable visualisations, several proposals were provided. Those which were most suitable for analysing the data were integrated in the dashboard. Thus, this guideline was fulfilled during dashboard design.

G3 (UI) and G5 (interaction): In order to investigate sensor measurements, simple interaction possibilities within the dashboard and the visualisations were provided. Since the provided visualisations and interaction possibilities enable users to analyse their data in detail without providing irrelevant information and interactions, these two guidelines were fulfilled.

G4 (clutter reduction): The dashboard enables users to select those sensors, which are interesting to them. Thus, it enables users to only investigate relevant data while other measurements are not shown in the visualisations. Therefore, this guideline was not fulfilled as the amount of data always depends on the user's selection.

G6 (overview-detail): Depending on the selected time range, the visualisations provide different resolutions of the underlying data. This enables users to analyse large time ranges using a low resolution while small time ranges can be investigated by zooming in and receiving a higher resolution of the data. This is required due to limited capabilities of time series libraries to visualise many sensor measurements. Loading many data points decreases the performance significantly while not being able to visualise them due to limited number of pixels available on the screen. Thus, different algorithms preserving the shape of the sensors were investigated and one suitable algorithm was implemented and applied on the original data on-the-fly. Since on-the-fly down sampling turned out to have two major drawbacks, considerations towards down sampling data before storing, were investigated. The first drawback is that down sampling consumes some time, especially for large time ranges. The second drawback is that values returned by the down sampling algorithm are not deterministic. This means that two users, requesting data for the same time range, receive different time points, both preserving the overall shape. Therefore, datasets were down sampled during data import using different resolutions. Depending on the selected time range in the dashboard, the server returns data from different resolutions without performing further operations. This solution overcomes both problems outlined before. This allows users to receive an overview and provide details on demand, therefore, this guideline was fulfilled.

G7 (colour): Since this is a generic framework supporting any kind of sensors, there is no specific assignment of colours to specific sensors. In order to visualise sensors using different colours, a set of 20 colours were selected and automatically assigned to the sensors for visualising. Thus, this framework only fulfils this guideline partially by only providing a limited number of different colours, while assignment of colours to sensors is done arbitrary not following any rule.

G8 (customisation): As the dashboard is a general approach, it only supports few customisation possibilities. Therefore, this guideline was not required and thus not fulfilled.

G9 (devices): Since mobile devices are less suitable for investigating large-scale sensor in detail, they are not supported by this framework. Furthermore, there was no demand for analysing data on mobile devices, therefore, this guideline was not fulfilled.

G10 (evaluation): For this dashboard no detailed user study was performed to investigate the usability. However, feedback from users confirmed that the dashboard enables users to investigate any large-scale sensor data, while not overloading the dashboard with too many configuration and interaction possibilities. Therefore, this guideline was not fulfilled.

6.9 Conclusion

This dashboard allows users to investigate large time ranges displayed in low resolution enabling users to receive an overview of the underlying dataset, zooming in provides a higher resolution of the sensor data enabling users to perform detailed analysis. The underlying server provides data for different resolutions depending on the selected time range. This allows the dashboard to load data within a short period of time as well as provide non-blocking interactions. The dashboard can be extended by different visualisations if required.

The proposed dashboard already allows to perform annotations on single sensors. Currently these annotations are only displayed in the dashboard. In the future, the system should be able to learn from annotations and automatically annotate sensor data based on previous annotations. This should allow users to recognise problems within the sensor measurements and provide potential solutions.

There are many possibilities for improvement and ideas for new features. One of them is the data import. In order to allow regular update of the latest sensor measurements, the framework needs the capability to perform data import automatically on a regular basis. Furthermore, time series databases have been improved since the project started, thus, replacing the RTS2 format by an open source time series database, for example InfluxDB [51], should be considered. Additionally, the provided filtering possibilities can be further improved by hierarchically indexing data and filtering depending on the selected time range as searching for value ranges within a large dataset is very time consuming.

Although there are many ideas for improvement and new features, the proposed dashboard is well suitable for analysing large-scale time series data, receiving an overview of large time ranges while loading details when zooming in.

Chapter 7

Analysis of Product Assembling and Testing Data

7.1 Introduction

This chapter discusses a dashboard where data from two different production processes, product assembling and testing, are combined within a dashboard. Due to different sources and data formats, users usually cannot investigate both datasets at the same time. Thus, problems in one dataset cannot be assigned to the other one, as no tool connecting the different datasets is available. Thus, the aim of this project was to select the right visualizations for the project partner enabling them to analyse their data by providing different interaction possibilities. Usually having one common data field for the two different datasets, enables users to evaluate the effects of product assembling data on testing results and vice versa. In contrast to other approaches, the dashboard used in this project only performs client-side data processing.

The author of this thesis was responsible for the conceptual design of the dashboard, implemented central visualisation components, designed and performed and evaluation and published the results in the poster *Evaluation of Visual Decision Support Systems used in Semiconductor Industry* [97] and contributed to the deliverable *Evaluation report on the effectiveness of the dashboard components* [26].

This project has been a research project with many external partners led by Roman Kern at Know-Center. The design and implementation of this dashboard has been a collaboration of several people. First of all, the tool used for creating this dashboard was proposed and implemented by Ilija Šimić in his Master's Thesis [135]. Choosing and implementing the right visualisation, suitable for connecting data from the two sources, was performed by the author of this thesis, by Ilija Šimić and Santokh Singh from Know-Center. The evaluation preparation, execution and analysis were mainly performed by the author of this thesis with support from Ilija Šimić and Vedran Sabol. The author of this thesis thanks everybody for collaborating, especially Ilija Šimić for providing Visualizer enabling users to create dashboards, all involved persons for implementing new visualisations and interaction possibilities as well as for providing advices during evaluation preparation, execution and result analysis phases.

7.2 Motivation

During production, usually many different measurements are performed to prove whether all components are assembled correctly within the production plant. Furthermore, tests are performed ensuring that the resulting products are working as expected. Often, measurements from different sources are analysed and evaluated separately. This is often due to different data formats, missing links between those datasets and missing tools enabling users to connect data. The main motivation for this project was to connect data from different sources and thus, considering data from all available sources when evaluating the resulting product. Linking different dataset with each other may outline effects of wrong assembling values in testing data, defect products can be identified during testing and can be assigned to certain assembling values and finally the combination of both values, assembling and testing data, may lead to adjustments of assembling ranges as certain values have no negative or even a positive impact on the resulting product. Thus, this project aimed to link data from different sources to gain full benefits of all available measurements during production and testing processes by only using a client-side tool.

7.3 Requirements

In order to connect data from different sources in general, different data sources require at least one common property, which can be used for coordinated linking and brushing [80]. Thus, production and testing processes are required to be set up in a way, which allows connecting these sources. This can be performed by using a single variable or, if nothing else is available, through timestamps if processes are executed successively.

Often, automatically generated data requires further processing due to changing data formats, different encodings or other reasons. Therefore, simple data cleaning and processing tasks are required to receive a common data format.

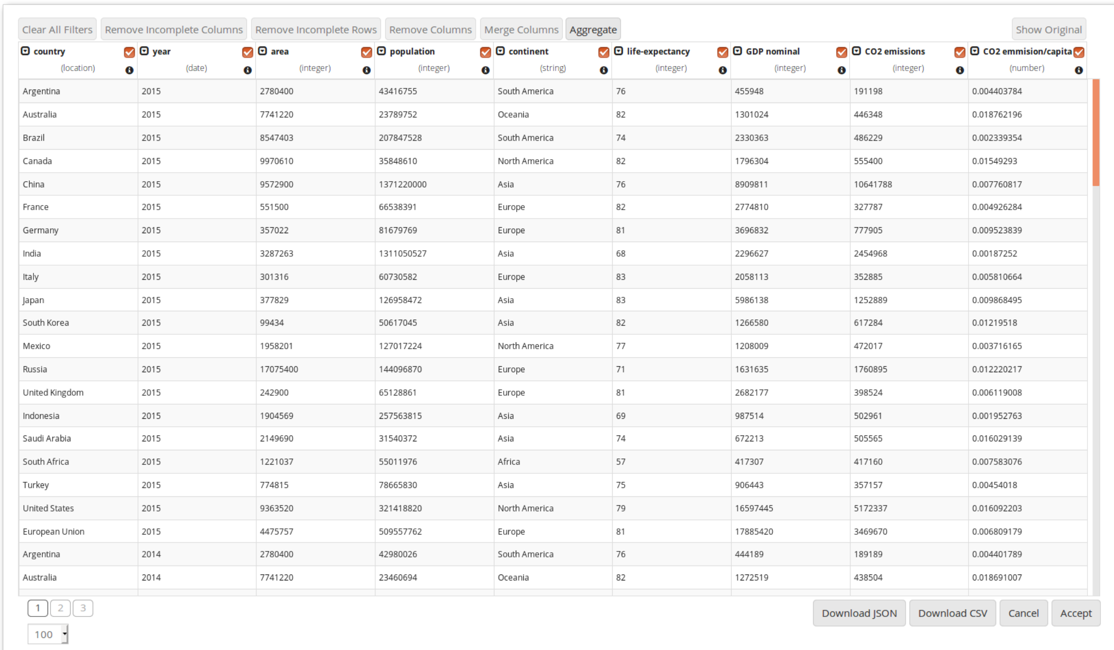
Besides linking the underlying datasets, a dashboard providing appropriate visualisations as well as supporting different datasets is required. In order to analyse the data, the dashboard should support different tasks. Performing these tasks is supported by choosing the right visualisation for the underlying dataset and allowing loading multiple datasets enable users to compare measurements. If these requirements are considered, the dashboard can provide an overview of the underlying data, allows to unveil data distributions and correlations, enables users to discover and monitor trends, allows to identifying anomalies and finally it enables users to identify cause effect relationships. Thus, visualisations within the dashboard are required to be interlinked. This enables them to select single elements within one visualisation and observe the corresponding values in other visualisations while greying out elements, which are not selected. Thus, providing a dashboard enabling users to investigate large dataset by linking data from different sources is an essential requirement.

7.4 Dashboard Design

The dashboard created in this project is based on Visualizer tool [135]. The development and design of the tool itself is not part of this thesis but will be first described shortly to receive an overview. This section concentrates on the dashboard creation and data analysis process using data from different production processes for a specific industry scenario.

7.4.1 Visualizer Description

Visualizer [135] is a web-based tool for analysing tabular datasets using multiple coordinated views [80]. Visualizer supports both, CSV and JSON format. It has two main views, a *Table View* for investigating and manipulating raw data and the *Dashboard View* for generating visualisations used for data analysis. The *Table View* allows several data cleaning and transformation operations. This includes removing incomplete rows or columns, activating/deactivating single columns, filtering data, replacing values and other features. The *Table View* does not support complex data operation methods as the proposed tool focuses on data visualisation rather than data manipulation operations. Table 7.1 shows an example of the *Table View*.



country (location)	year (date)	area (integer)	population (integer)	continent (string)	life-expectancy (integer)	GDP nominal (integer)	CO2 emissions (integer)	CO2 emmision/capita (number)
Argentina	2015	2780400	43416755	South America	76	455948	191198	0.004403784
Australia	2015	7741220	23789752	Oceania	82	1301024	446348	0.018762196
Brazil	2015	8547403	207847528	South America	74	2330363	486229	0.002339354
Canada	2015	9970610	35848610	North America	82	1796304	555400	0.01549293
China	2015	9572900	1371220000	Asia	76	8909811	10641788	0.007760817
France	2015	551500	66538391	Europe	82	2774810	327787	0.004926284
Germany	2015	357022	81679769	Europe	81	3696832	777905	0.009523839
India	2015	3287263	1311050527	Asia	68	2296627	2454968	0.00187252
Italy	2015	301316	60730582	Europe	83	2058113	352885	0.005810664
Japan	2015	377829	126958472	Asia	83	5986138	1252889	0.009868495
South Korea	2015	99434	50617045	Asia	82	1266580	617284	0.01219518
Mexico	2015	1958201	127017224	North America	77	1208009	472017	0.003716165
Russia	2015	17075400	144096870	Europe	71	1631635	1760895	0.012220217
United Kingdom	2015	242900	65128861	Europe	81	2682177	398524	0.006119008
Indonesia	2015	1904569	257563815	Asia	69	987514	502961	0.001952763
Saudi Arabia	2015	2149690	31540372	Asia	74	672213	505565	0.016029139
South Africa	2015	1221037	55011976	Africa	57	417307	417160	0.007583076
Turkey	2015	774815	78665830	Asia	75	906443	357157	0.00454018
United States	2015	9363520	321418820	North America	79	16597445	5172337	0.016092203
European Union	2015	4475757	509557762	Europe	81	17885420	3469670	0.006809179
Argentina	2014	2780400	42980026	South America	76	444189	189189	0.004401789
Australia	2014	7741220	23460694	Oceania	82	1272519	438504	0.018691007

Figure 7.1: Table view of Visualizer tool.

The *Dashboard View* enables users to create multiple visualisations based on suggestions or user's selection. In order to generate a visualisation, users select data fields they want to investigate. Generating a bar chart, for example requires two fields, one categorical and one numerical. Depending on the number and type of selected fields, different visualisations are enabled in the *VisPicker*, based on predefined rules. Depending on the selected visualisation, data aggregation methods are allowed or even required. Aggregation methods can be *average*, *minimum*, *maximum* or *the sum of all values*. Depending on the generated visualisation, different interaction possibilities are available. This ranges from *brushing*, *sorting* and *filtering* to the *assignment of data fields to visualisation channels*. Additionally, the proposed dashboard supports multiple diverse datasets. Each dataset (one to multiple files containing the same data fields) or even dataset file can be manipulated separately or together using the *Table View* before being visualised.

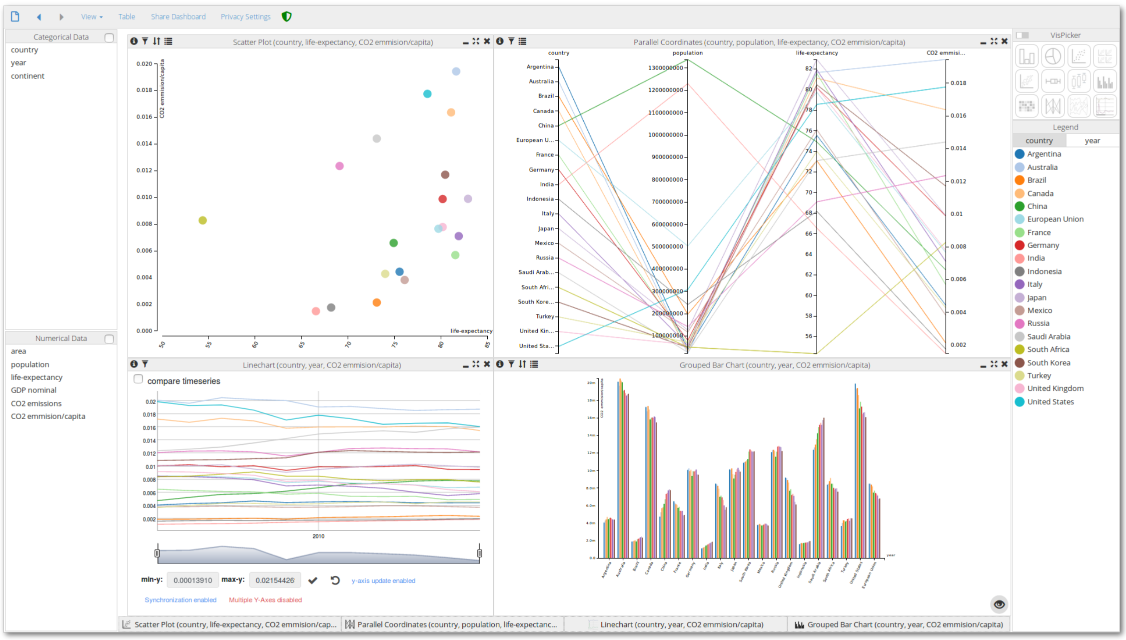


Figure 7.2: Dashboard View of Visualizer tool.

Figure 7.2 shows a sample dashboard using four different visualisations. Each chart provides different insights on the underlying dataset. Interactions in one visualisation highlight the corresponding data subset in all other visualisations. Figure 7.3 shows that interactions in the scatter plot highlights the corresponding subset in the parallel coordinates, the line chart and the grouped bar chart.

A more detailed description about Visualizer tool is provided in appendix C.1. This description was provided to the evaluation participants to receive an introduction to Visualizer including all necessary interaction possibilities.

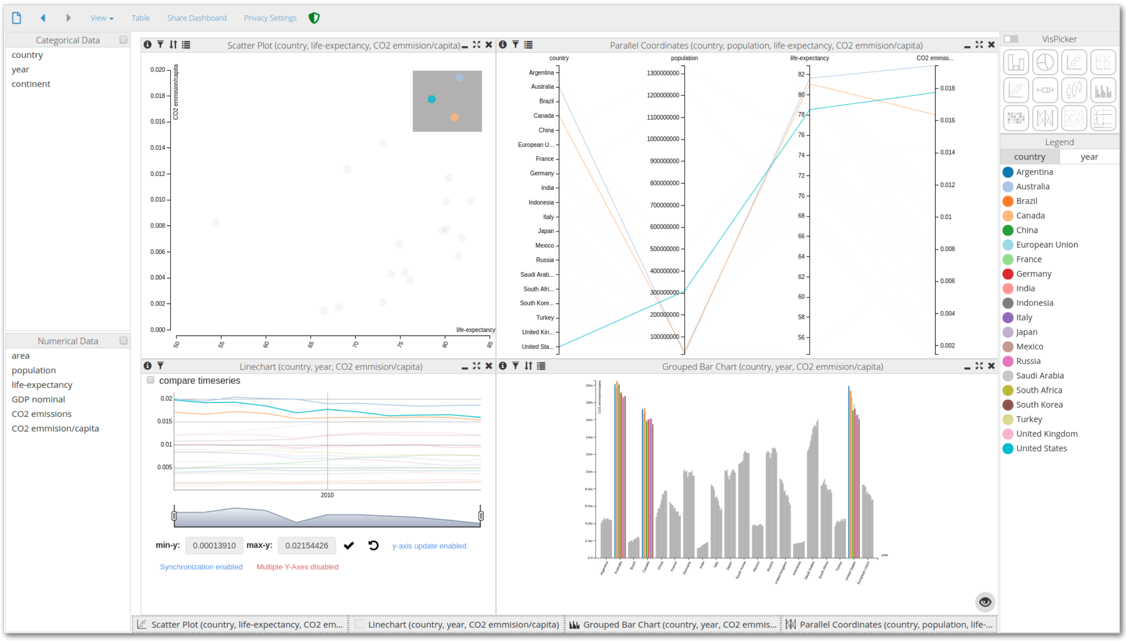


Figure 7.3: Interactions in one visualisation highlights the corresponding fields in all other visualisations.

7.4.2 Dashboard Creation

After receiving an overview of Visualizer, which is not the scope of this thesis, the dashboard design for assembling and testing data is described in the following.

As a result of the production process, the analysis of data gathered from different production processes should be merged and analysed to obtain new insights and to support users in deciding whether certain products match all requirements for the market.

After investigating and pre-processing the available data, an interesting subset of the provided data was used for analysis. Although Visualizer can cope with huge datasets containing up to several millions of rows, visualising large dataset is not possible for all visualisations. This is due to missing aggregation methods within some visualisations. Therefore, to demonstrate the functionality without spending too much time on optimisation, the underlying dataset was investigated and only a small interesting subset was selected to demonstrate the analysis process.

Before selecting a set of data files, the fields within the files were investigated to identify those, which might be interesting for further analysis. For the assembling data, several fields were identified as being suitable for visualising; these are the *id*, the *assembling description*, the *actual assembling value* as well as its *minimum* and *maximum ranges* and the *unit* of the assembling process. The same process was performed for the testing data, were all fields were considered for being suitable for visual analysis.

As the total number of files was quite large, not all products could be visualised at the same time. Therefore, to identify an appropriate data subset, a small analysis was performed to choose an interesting subset. The analysis revealed, which assembling operations contained values outside the expected range. The results outlined that several assembling values contained the value zero or one. According to the project partner, these values were stored due to missing measurements and can therefore be ignored. Other than these, only two assembling operations were outside the optimal range indicating that production processes are highly optimised. Thus, products with the deviating assembling values were selected as they are of special interest. In addition, some other arbitrary products, which were temporary correlated to the two deviating products, were selected for analysis. For some products only assembling or testing data was available and could therefore be visualised but not connected with the other data source. The dashboard discussed in the following contains 15 products.

In order to select appropriate visualisations for data analysis, several iterations of generating dashboards using visualisations inside Visualizer were required. Since the testing data contained time series data, a line chart visualisation was considered to be suitable for visual analysis. Visualizer already contained a line chart visualisation but the proposed implementation did not serve all requirements. Therefore, the decision to implement new line charts serving all requirements was taken.

For the assembling data, initially parallel coordinates were investigated. Using them, some drawbacks were revealed. Due to the large number of assembling operations and the thereof large number of resulting axes, using parallel coordinates for data analysis is too inefficient as horizontal scrolling is required for analysis. Additionally, an axis within the parallel coordinates can only encode one data field, thus displaying whether an assembling value is within the range or outside is not possible without adjustments for the specific use case. After some research, bullet charts were identified as serving all requirements while saving space as they contain a more compact visual representation.

Additionally, the project partner provided the information that products might be tested several times if some problems occurred during the first testing phase. Therefore, a stream graph showing the time of testing was considered suitable to prove whether a product was tested multiple times.

Since Visualizer did not contain bullet charts at all or appropriate line charts, these two visualisations were implemented for this dashboard. The stream graph was implemented for a different dashboard but reused in this case. After the first implementation phase, simple line charts without sufficient interaction possibilities were included in Visualizer.

During the next development phase, features for both visualisations, bullet chart and line chart, were extended and refined to serve all requirements.

7.5 Dashboard Description

In the following each visualisation and its interaction possibilities are described in detail.

7.5.1 Bullet Chart

The bullet charts allow visualising measurements according to certain ranges. Each line within the bullet charts represents an assembling operation. The corresponding operation name is shown on the left side. Depending on whether values are outside the optimal range or not, the name of the operation is displayed red or black. Clicking on the name provides the whole list of products and their corresponding measurements. Colour encoding is again the same as for the assembling operations. The value range for all assembling values is displayed on the top of the visualisation. If all values have about the same range, displaying them in one chart is sufficient. If these values differentiate much, separate bullet charts are more suitable for data analysis. Each assembling operation has an optimal range coloured green, lower as well as higher ranges are coloured red. As optimal ranges can change based on analysis results or optimisations, assembling operations can have multiple lines showing different optimal ranges and the corresponding measurements. Assembling measurements are displayed as small black or red lines depending on whether they are within the optimal range or not. Tooltips show the corresponding product ids and their assembling values. One or more measurements can be selected by brushing the area of interest for an assembling operation. Product ids, which are contained within the selected range, are highlighted in all other assembling operations within this visualisation as well as in all other visualisations, while all other ids are greyed out.

The bullet chart allows filtering those measurements, which are outside the optimal assembling range enabling users to concentrate on those measurements. Additional features provided for this visualisation are filtering for ids, assembling operations or value ranges. Data can be sorted according to the assembling operation name. Remapping of data fields to visualisation properties is done automatically on visualisation generation but can be changed by users. Since Visualizer does not automatically assigns values correct depending on the field names, adjustments of field mapping are required.

7.5.2 Line Chart

A line chart can contain one or multiple lines. Each line corresponds to one product and shows the measurements over time. Since the dashboard for analysing large-scale time series data, described in chapter 6, already contains line charts suitable for analysing sensor data, the line chart implementation was extracted from this dashboard and integrated to Visualizer. Thus, line charts within Visualizer provide the same interaction possibilities as described in section 6.5.2.1. In order to enable users to compare measurements of different products, the feature *compare timestamps* maps different timestamps to the same start time, usually to the first one.

7.5.3 Stream Graph

A stream graph [101] in general visualises the number of occurrences of a certain element for a given point in time, whereby each stream represents one element. Therefore, each product is represented by one stream, encoded by using a distinct colour. The thickness of each stream represents the number of occurrences at that time. For products, this corresponds to the number of measurements during testing. Thus, the stream graph allows visualising when a product was tested, how many measurements were performed as well as whether it was tested multiple times. Furthermore, it shows whether and if so, which other products were tested within a certain time range. The stream graph only provides few interaction possibilities. Products can be selected by clicking on the corresponding stream. Clicking again on a selected stream, deselects the corresponding product again. Clicking in the white area of the stream graph, removes all selections. Tooltips provide information on the time range of the measurements, product information and the number of measurements.

7.6 Case Study

7.6.1 Dashboard Generation

If a user wants to investigate product assembling and testing data, the following steps are required. After opening the corresponding URL in the browser, the user has to select the files for investigation. Therefore, the user opens the dashboard and selects a folder from the local file storage. The selected files can have multiple diverse fields as shown in Figure 7.4.

Figure 7.4: Opening Visualizer and selecting a folder.

After accepting the selected files, the user is forwarded to the *Table View*, see Figure 7.5. Each selected file can be investigated separately by switching between files using the dropdown menu in the upper left area of the Table View. Investigating the assembling and testing data, the user can observe that the tool automatically recognises data types correctly.

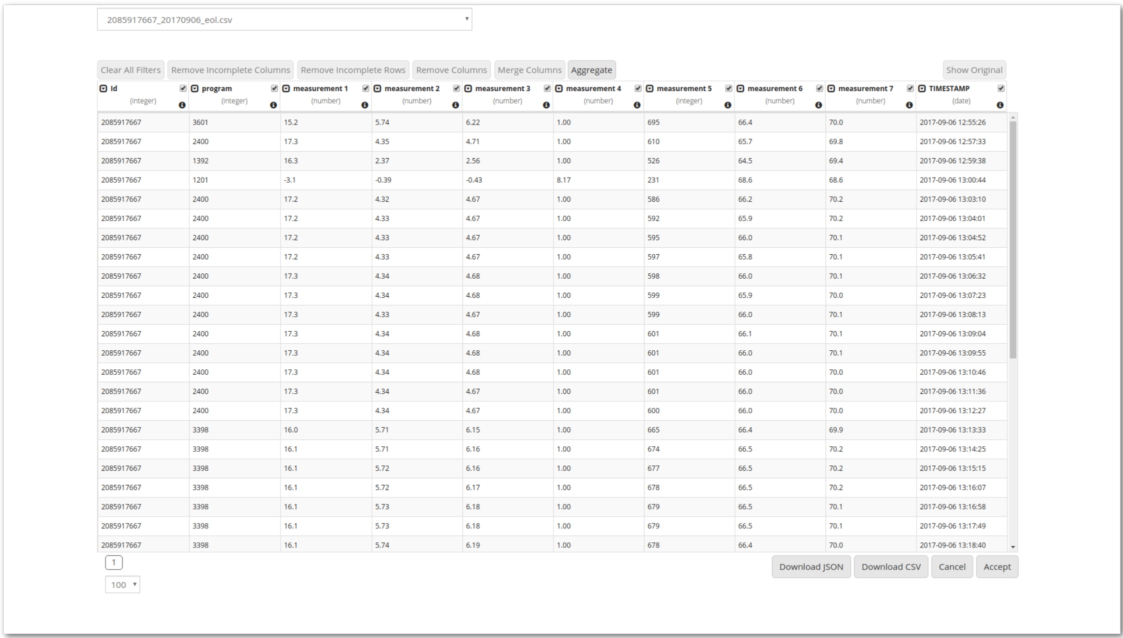


Figure 7.5: Table View displaying imported data.

Since the *id* is a numerical value, the tool assigns it to *integer* type. However, for visualising the data, the user needs to change the *id* to *string* type as shown in Figure 7.6. Although the *Table View* provides several data manipulation methods, no further operations are required. Thus, the user can continue with the actual dashboard by accepting the changes.

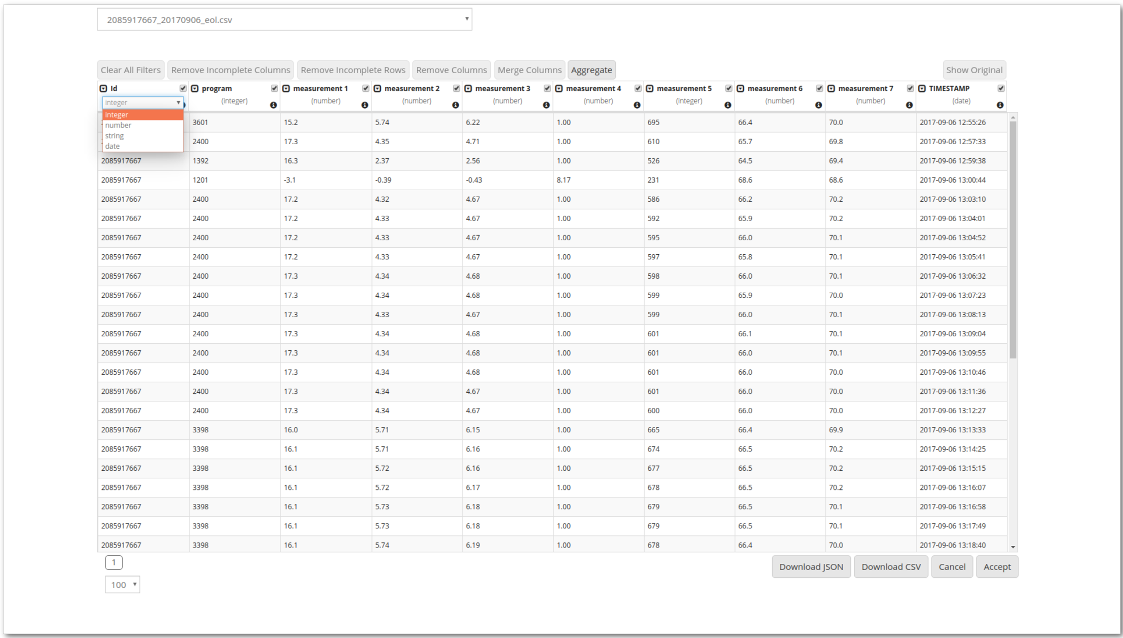


Figure 7.6: Changing automatically detected field types.

The *Dashboard View* requires several steps for visualisation generation. Initially, the user needs to select the data files for visualising. Depending on the selected files, only those data fields are shown in the left area, which are contained in all selected files. Selecting files with different fields for visualising may result in only few or no common fields for selection. Thus, selecting files containing similar fields enable users to generate visualisations. Therefore, the user initially selects all assembling files, see Figure 7.7.

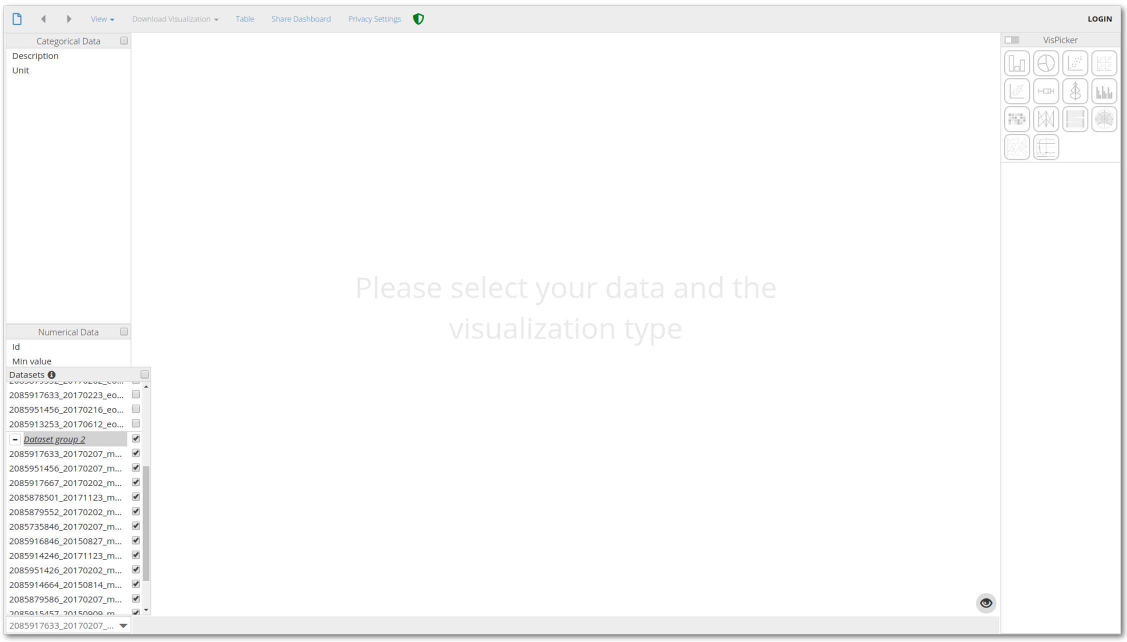


Figure 7.7: Selection of assembling data within the Dashboard View.

Afterwards, the user selects those fields, which should be visualised. These are the *id*, the *description*, the *minimum value*, the *maximum value*, the actual *assembling value* and the corresponding *assembling unit*. After selecting all fields, three visualisations are enabled in the *VisPicker*, positioned on the right side within the dashboard, indicating that for the current selection these visualisations are available, see Figure 7.8.

Since the user wants to investigate assembling ranges, the user selects the bullet chart without performing data aggregation, see Figure 7.9.

The visualisation tries to automatically assign the selected fields to certain visualisation channels. As there is no further intelligence than type assignment, generating the bullet chart for the underlying dataset requires remapping of available fields. This operation needs to be performed by the user. Therefore, the user selects the corresponding icon in the visualisation header and remaps the fields. After performing this, the visualisation is displayed as expected. The process of bullet chart generation, remapping of fields within the visualisation and resulting visualisation is shown in Figure 7.9 and 7.10.

Besides the assembling values, the user also wants to investigate the testing data. Therefore, the user selects the corresponding files in the file chooser. This updates the available fields in the left area. After selecting the *id* and *timestamp*, four visualisations are enabled

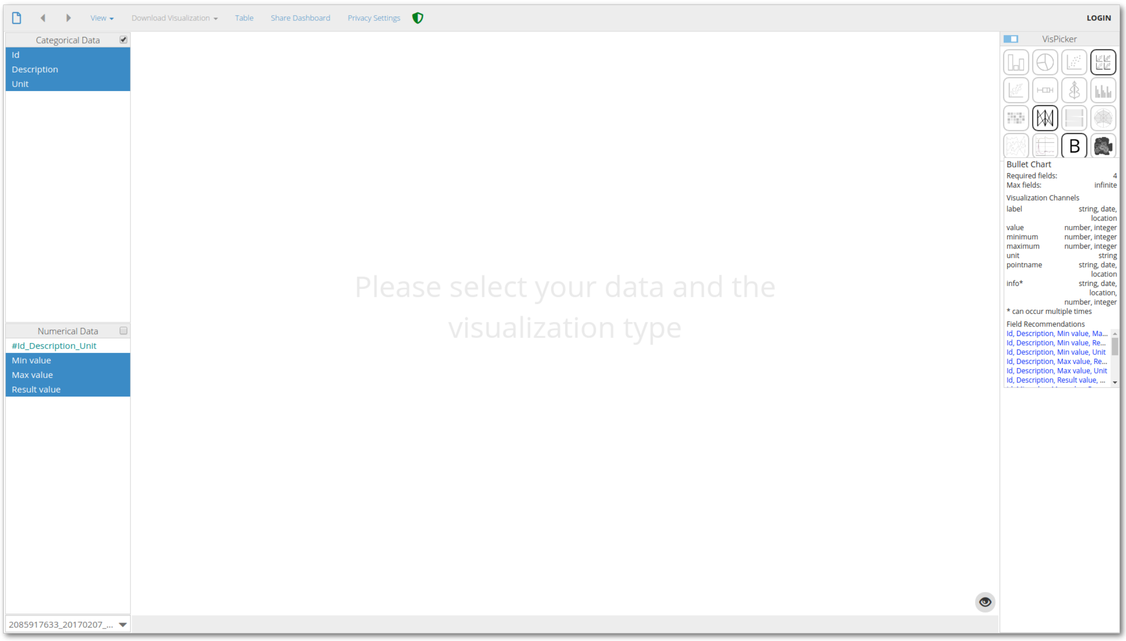


Figure 7.8: Depending on the selected fields, different visualisations are enabled within the VisPicker.

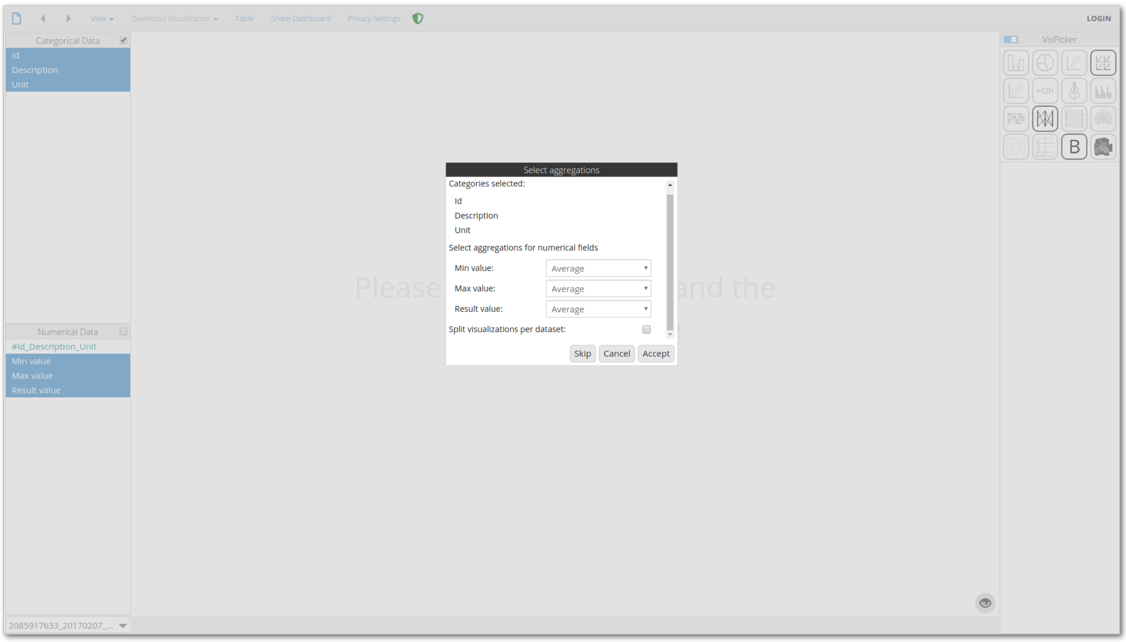


Figure 7.9: Dialogue displayed after selecting a visualisation enabling users to aggregate data.

in the *VisPicker*. Since the user wants to investigate when the products were tested, the stream graph without data aggregation is chosen. As both visualisations, bullet chart and stream graph, have one common property namely the *id*, interactions are linked by using this data field, see Figure 7.11.

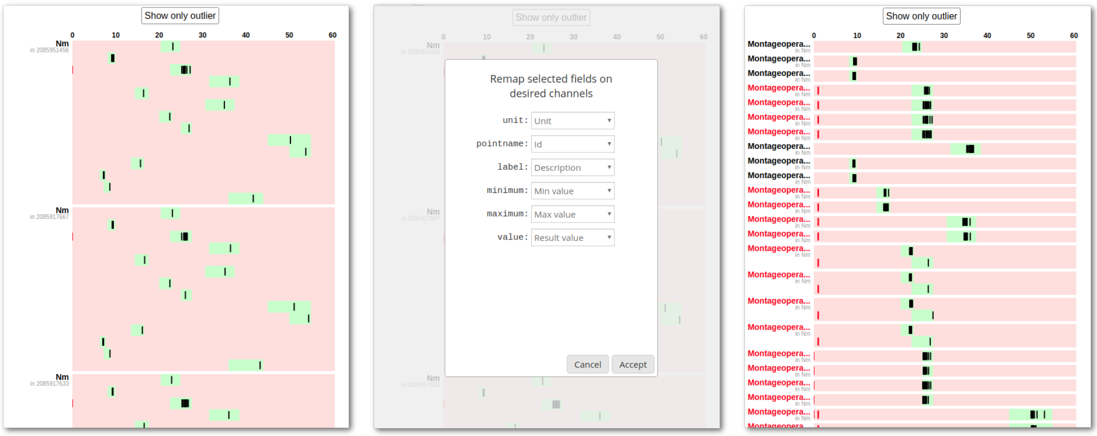


Figure 7.10: Generating bullet charts requires remapping of selected fields to visualisation fields.

Finally, the user also selects the field *program*, providing information about the product testing routine, and generates a line chart by skipping data aggregation. This shows multiple single lines within a time range from January to November 2017, outlined in the left image of Figure 7.11. Since the line chart does not support comparison of single values due to different times of measurements, a flag within the visualisation allows mapping all lines to one common starting point (usually the one of the first measurement). By setting this flag, measurements can be finally compared with each other. This is illustrated in the right image of Figure 7.11. The same procedure is performed using different measurements. Figure 7.12 shows the resulting dashboard.

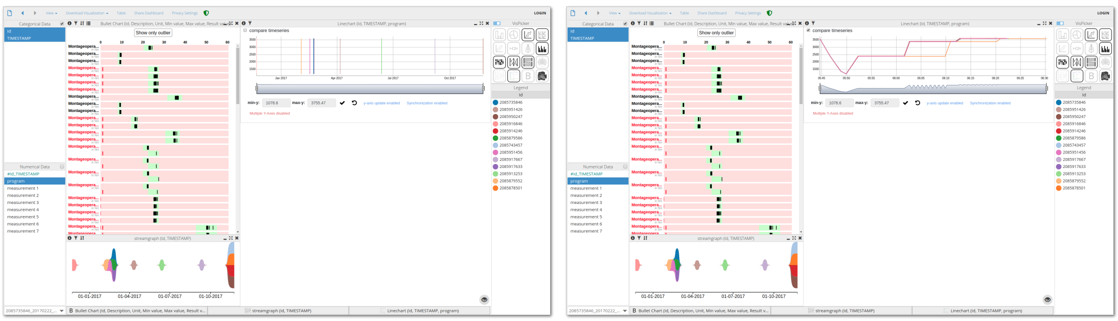


Figure 7.11: Generating a line chart by mapping all measurement to the same time range for comparison.

While generating the dashboard, the corresponding URL is constantly updating. This enables users to reuse the dashboard URL without performing all dashboard generation steps again. Furthermore, users can easily share the dashboard with colleagues.

Figure 7.12: Dashboard created for analysing product assembling and testing data.

7.6.2 Data Analysis

After dashboard generation, the user first investigates the assembling data. Most of the assembling operations are within the expected range not considering those measurements, which have a value of zero or one. There are two exceptions on this, assembling operation *2830589554* for product id *2085915457* and assembling operation *3498309064* for product id *2085913253*. Figure 7.14 shows how the user investigates assembling values. By selecting the value, which is outside the range, all corresponding assembling values for this product id are highlighted. Additionally, the corresponding stream in the stream graph, as well as the lines within the line charts, are highlighted. Figure 7.13 shows that there are no testing values for the first product, which is outside optimal assembling ranges.

In contrast to the previous product, testing values for the second assembling outlier, shown in Figure 7.14, are available indicated by the highlighted green area in the stream graph. Considering the corresponding line charts on the right, the user can observe that there are small deviations, compared to other measurements, during the testing procedure. These deviations are highlighted by the blue rectangles. They can be an indicator of wrong assembling values.

Investigating the different measurements for testing, see Figure 7.15, highlights only those assembling operations and testing results which are connected to the product with a deviating testing procedure compared to all others. This can be investigated in the first line chart, where the orange line deviates compared to all others after 15 minutes of testing. This of course also leads to deviating values for all testing measurements. Investigating the assembling values for this product, the bullet chart shows that all assembling values

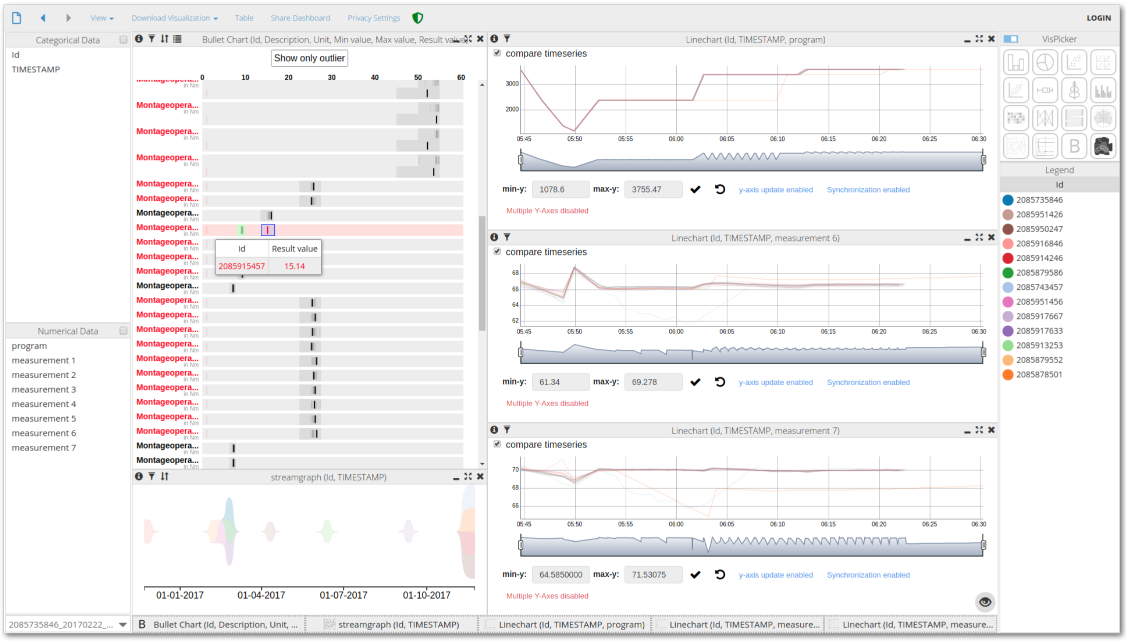


Figure 7.13: Selecting assembling values, which are outside the expected range using the bullet chart, shows that no testing data is available for the selected id.

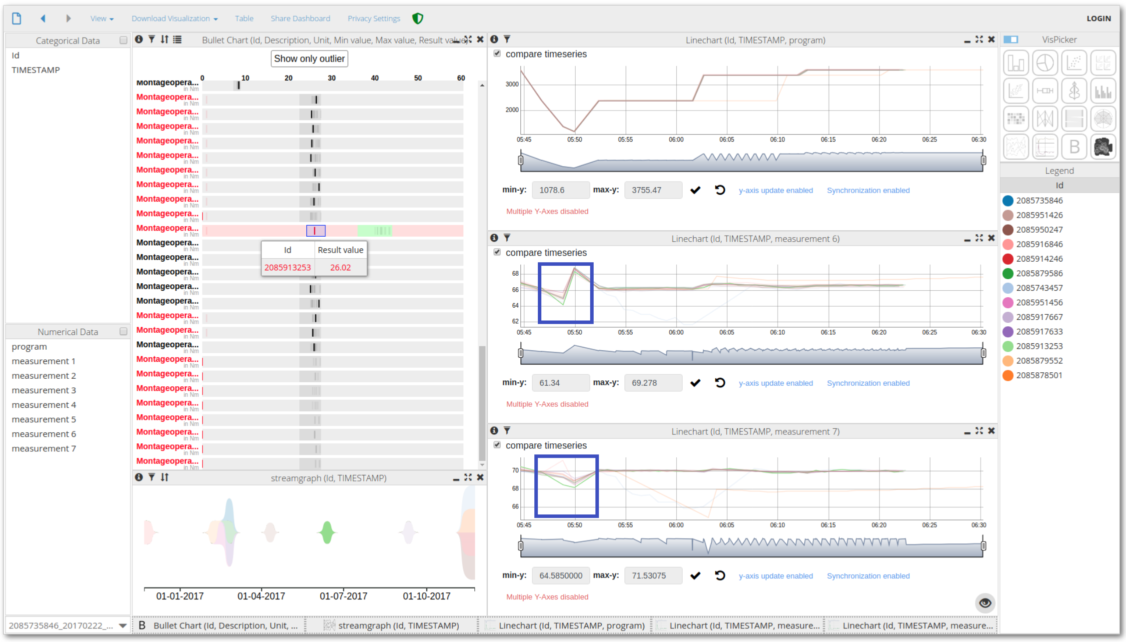


Figure 7.14: Selecting assembling values, which are outside the expected range using the bullet chart highlights the corresponding testing data for the selected id.

are within their expected ranges. Nevertheless, some are close to the minimum or maximum borders.

When investigating the second and third line chart, the user can observe that the blue coloured product deviates from the majority of measurements. Therefore, the user selects this line and investigates the assembling values, see Figure 7.16. Although the testing values deviate a lot, all assembling values are within the expected ranges. Therefore, assembling values should not be responsible for deviating testing measurements. In order to identify the actual problem, other production data need to be investigated.

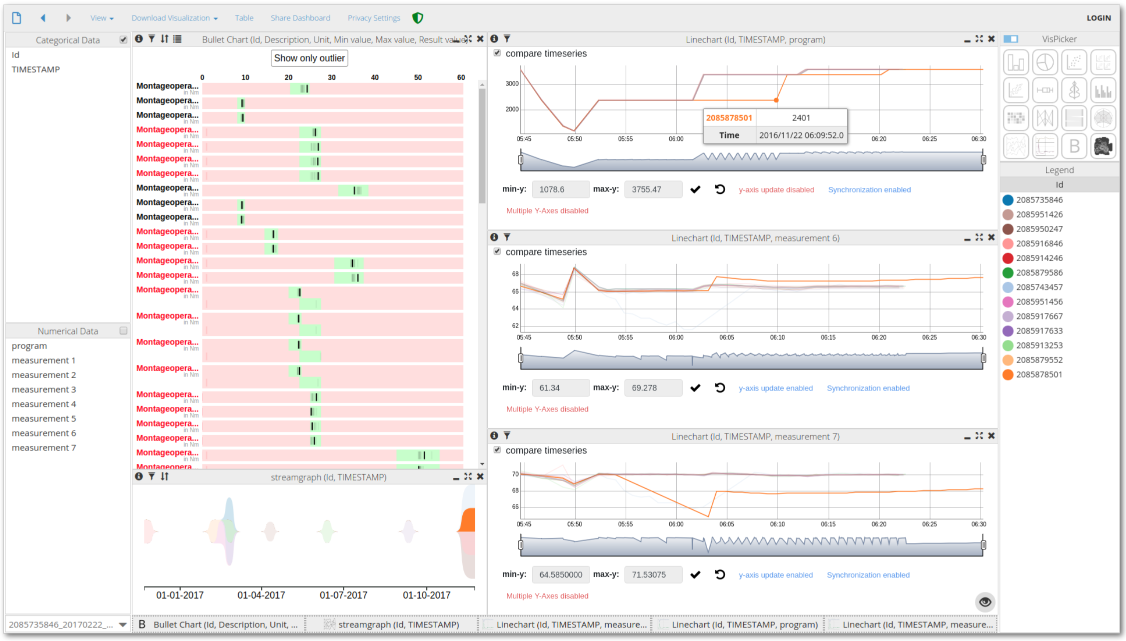


Figure 7.15: Investigating different test programs by using line charts.

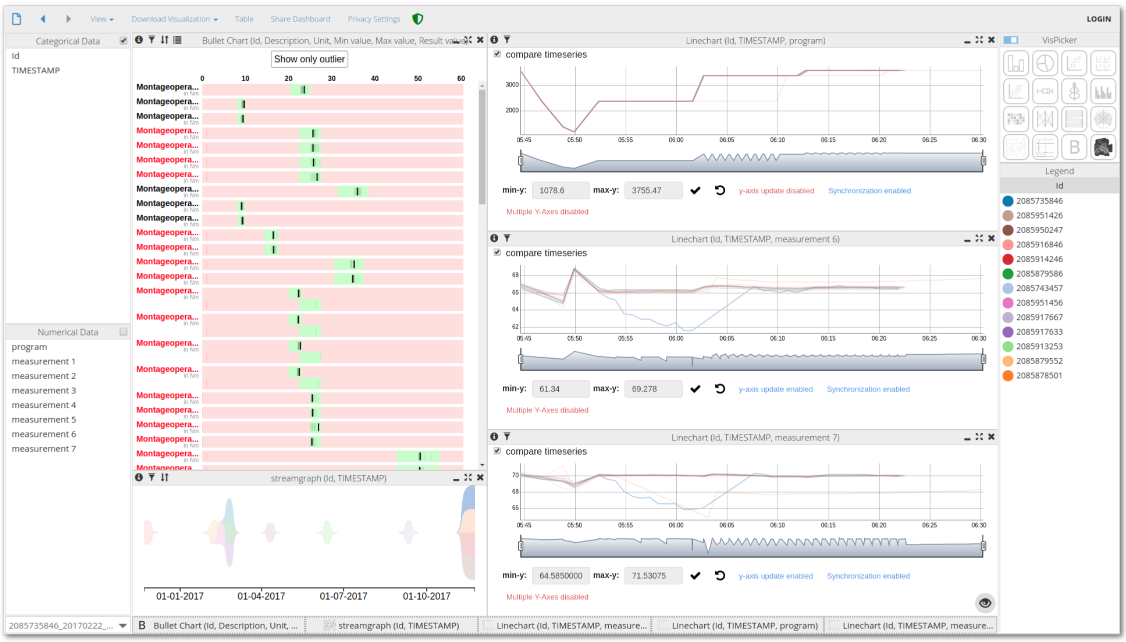


Figure 7.16: Investigating different test measurements by using line charts.

Figure 7.17 highlights another product, which has deviations at the beginning of the measurement in the third line chart. Analysing the corresponding assembling values, the user can observe that about one out of four assembling values are not available. Therefore, investigating this problem might be difficult due to missing values, but still might be possible by considering additional production data. Furthermore, the stream graph shows that this product was tested and therefore produced much earlier compared to other products.

Assuming that the majority of products with similar testing curves are more likely correct than those with deviating measurements, products with deviating measurements require further investigation. Analysing a larger set of testing measurements can prove this.

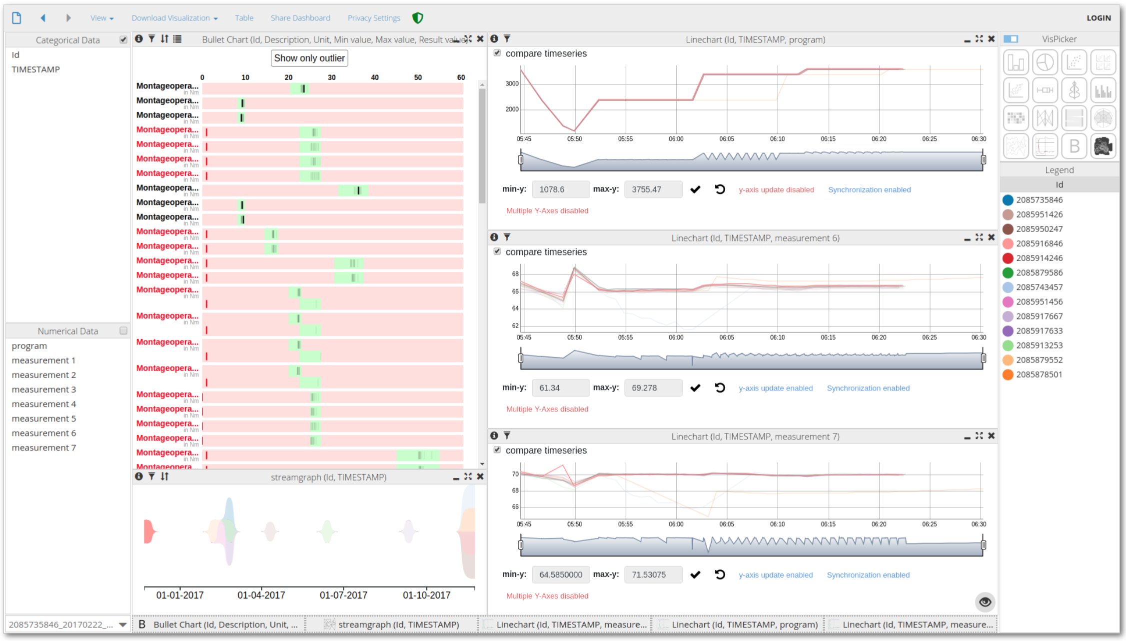


Figure 7.17: Investigating deviations in testing measurements by using line charts.

Investigating the stream graph in Figure 7.18, the user can see when the products were tested (and therefore produced) and whether they were tested only once or several times. Selecting the last four measurements where testing was performed within a short time period, the corresponding assembling values are mainly within the range, only few have measurements of zero or one. Considering the line charts, the user can observe that especially for these products tested within a short time period, measurements are deviating comparable much. This may lead to the conclusion, if all production values are okay, that the testing environment caused wrong measurements.

7.7 Software Design and Technical Information

As already mentioned before, the design and implementation of Visualizer itself, is not part of this thesis but Visualizer was used to configure a dashboard for a certain industry

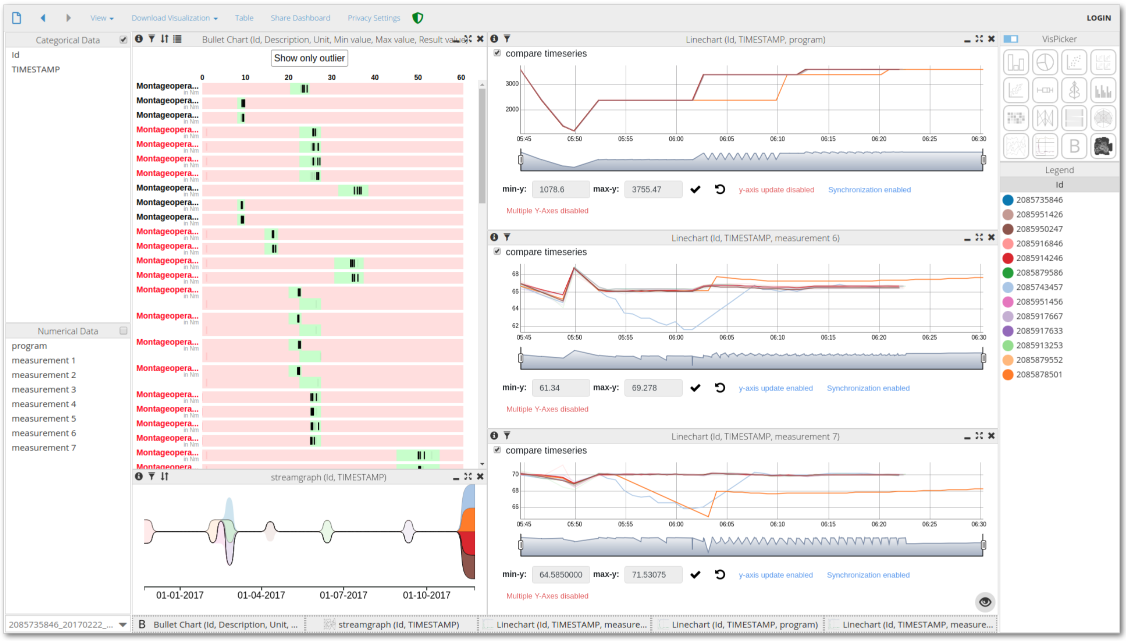


Figure 7.18: Investigating when and how often products were tested by using the stream graph.

scenario. Thus, detailed information about the server component can be investigated in the Master's Thesis written by Ilija Šimić [135]. Visualizer is set up in a way that all data processing is performed on the client, only for very specific features, information about the dataset is sent to the server. The dashboard is implemented using JavaScript [138], whereby each visualisation has its own environment using I-frames. Further details on the implementation are provided by Šimić [135].

7.8 Discussion

This section discusses whether and how the guidelines defined in section 2.4 were considered for this dashboard.

G1 (information): The main requirement within this project was to combine data from different sources enabling users to draw conclusions based a broader information base. Thus, data from different sources must be linked by using one common field and visualise them in a dashboard containing multiple coordinated views. Since the dashboard enables users to combine data from testing and assembling, this guideline was fulfilled.

G2 (visualisation): In order to identify the right representation possibilities, the data was investigated in detail and several visualisations provided in Visualizer were investigated. Results of this process revealed that line charts and stream graphs are suitable for visualising the testing data, while bullet charts were selected for assembling data. The final dashboard enabled users to analyse their data by combining data from different sources.

Therefore, this guideline was fulfilled.

G3 (UI): Due to visualising data from different sources, which by default raises complexity, the provided visualisations only show the required information. Less relevant information was not selected for visualising within the dashboard, as it did not provide and further insights and thus, no benefits. Therefore, this guideline was fulfilled.

G4 (clutter reduction): Within the visualisations, all measurements are displayed by default. Different operations like filters allow investigating data which for example was not within the expected range. Other than that, in general all data is visualised within the chart. As this dashboard initially shows all selected data, this guideline was not fulfilled.

G5 (interaction): Visualizer per default enables users to analyse their data by providing multiple coordinated views. This allows users to easily analyse their data by observing the effect of interactions in all connected visualisations. Therefore, this guideline was fulfilled.

G6 (overview-detail): Visualizer displays all provided data points without any exception. Thus, it does not provide visualisations enabling users to receive an overview and investigate details when performing interactions. However, it allows the user to create different visualisations suitable for different purposes. Therefore, this guideline was not fulfilled.

G7 (colour): Since Visualizer is a generic tool that makes it possible to visualise any tabular data, certain colours cannot be assigned to any data fields. In order to display data within Visualizer, the tool uses a colour scheme with 20 different colours, which are automatically assigned to the underlying data. If more colours are required, colours are repeated. Since colours are not assigned to specific data fields related to specific information or statuses, this guideline was only partially fulfilled by limiting the number of different colours.

G8 (customisation): Visualizer enables users to reuse and share dashboards after creating once. Thus, the tool enables different users to create a dashboard according to their purposes. However, restrictions are defined by the tool; including data processing, automatically assigned colours, and available visualisations as long as users do not create and upload their own visualisations. This tool enables users to create dashboards handling their specific requirements for data analysis and allow expert users to extend the framework by adding their own visualisations, therefore, this guideline was fulfilled.

G9 (devices): Although, Visualizer was initially not implemented to serve mobile devices, extensions allow analysing data using tablets and mobile phones. However, complex representations such as those proposed for this dashboard, are not supported by mobile devices and was also not a requirement. Therefore, this guideline was not fulfilled.

G10 (evaluation): In order to investigate whether the provided dashboard meets all requirements, an evaluation was conducted, which revealed that the dashboard was suitable for analysing data from different sources, see section 10.2 for details. However, ratings from the personal assessment revealed that for some users, who hardly ever use visualisation tools for data analysis, difficulties arose when analysing data within the dashboard. Therefore, this guideline was fulfilled.

7.9 Conclusion

In this project, a dashboard connecting data from different production processes was designed an analysed using Visualizer tool. In order to investigate whether Visualizer in general and the proposed dashboard specifically are suitable for data analysis, an evaluation with nine test users was conducted, see section 10.2. The evaluation was made up of two phases. The first phase included the usage of Visualizer while logging interactions. The second phase included a detailed questionnaire to gain further insights. The results from the evaluation showed that participants in general appreciate to have a tool for visually inspecting their data. Possibilities for interactions as well as data processing, especially within visualisations, require further improvements. Due to the fact that there is a lack of tools connecting data from different sources and enabling users to visually analyse their data, the proposed dashboard provided first visual insights on possible effects of wrong assembling values on testing measurements and vice versa. Additionally, the questionnaire provided many insights into what are the essential requirements on tool used for production processes, which visualisations are relevant for data analysis and users are familiar with, and which visualisations supported different tasks during data analysis.

Chapter 8

Monitoring Environmental Data

8.1 Introduction

This chapter discusses a dashboard created for analysing sensor measurements according to their availability and plausibility. While investigating availability of sensor measurements is simple, investigating whether available sensor measurements are plausible is more complicated. This approach focuses on evaluating the plausibility of sensors by investigating neighbouring sensors. This is done by analysing sensor measurements in relation to their geographical position.

The author of this thesis was responsible for the conceptual design of the dashboard as well as fully implemented the dashboard.

The project has been a research project in cooperation with Graz University of Technology and an external project partner led by Mark Kröll at Know-Center. While several employees from the Knowledge Discovery team, led by Roman Kern, investigated different plausibility algorithms, the server component and the resulting algorithms for calculating availability and plausibility were implemented by Michael Wittmayer. The dashboard has been designed and implemented by the author of this thesis. The author of this thesis thanks everybody for collaborating especially for providing the server component serving all data required within the dashboard and the research partner from Graz University of Technology.

8.2 Motivation

The main motivation for conduction this project was to visualise, compare and evaluate sensor data according to their plausibility and availability. Depending on sensor measurements, certain actions have to be conducted. Thus, investigating whether these measurements are plausible is essential. Therefore, this project was set up to enable users to:

- Prove whether sensor measurements are available.
- Prove the plausibility of sensor measurements by investigating close sensors.
- Observe measurements over time.
- Compare different sensors, measuring the same variable, with each other, and thus identifying high quality sensors.
- Prove whether the right actions are taken depending on sensor measurements and their interpretation.

8.3 Requirements

The main requirements on the dashboard are to visually identify sensors which are either not available or implausible. By providing a dashboard supporting to expose these sensors, decisions can be taken accordingly. Thus, wrong measurements can be ignored and sensors can be replaced while correct measurements can lead to certain actions.

After the problem definition, several visualisations have been proposed to the project partner enabling them to refine their requirements. Details on these visualisations are described in section 8.4. Based on the selected visualisations, small use cases were described. Each of them was rated according to its priority. Since not all of them could be implemented in the first prototype, the following use cases have been selected:

- First of all, the dashboard should provide a simple overview of all stations. The geographical visualisation is considered suitable for providing an overview of most important sensor information. A sensor can work correctly, can be defect, can contain outliers or cannot be available. A simple table representation allows to analyse selected or defect sensors including most important data fields. Line chart representation allows observing the historical development of single or multiple selected sensors.
- Depending on certain roles, different interaction possibilities should be available. The *simple view* only provides limited analysis methods, whereas the *expert view* provides more analysis methods as well as changing the status of sensors. In contrast to the previous two, the *admin view* allows users to modify sensor information within the database. Thus, handling different views depending on their tasks was a major requirement on the dashboard.
- The geographical visualisation should display a map of Austria allowing users to zoom in on demand. Additional mouse interactions should enable users to select sensor stations within the map. Selecting one or more stations should open a sunburst visualisation displaying all sensors of a station.
- The availability and plausibility of sensors should be outlined in the geographical visualisations by using different colours.

- Selecting one of the stations in the geographical visualisation opens the sunburst view and immediately enables users to see which sensors are not available and which are implausible.
- Neighbouring sensors measuring the same variable can be compared by using line charts.
- In order to compare sensors, the line chart visualisation should enable users to compare up to four sensors of the same unit.
- Outliers should be highlighted within visualisations. Additionally, the dashboard should also enable users to define outliers by providing simple constraints. This was not required for the first prototype.
- The dashboard should enable expert users to passivate single sensors or whole stations. Availability and plausibility algorithms are not applied on passivated sensors or stations. Passive stations are displayed in the geographical visualisation by using a separate colour.

8.4 Dashboard Design

After problem definition, several visualisations were investigated and proposed to the project partner for data analysis. The first proposed visualisation was a geographical visualisation, which aggregates underlying information and provides details when zooming in. As in the final implementation each station was represented by a simple shape, this was not required. Filtering operations enable users to only investigate sensors of interest, for example sensors that are implausible.

As a station can contain many sensors, providing a compact visualisation enabling users to analyse all sensors, was proposed by introducing the sunburst visualisations. The sunburst visualisation was initially considered to be directly positioned within the geographical visualisation. Testing the resulting implementation revealed two main problems. First, the problem of overlapping close locations and second the readability due to small representations. Thus, the sunburst visualisation was positioned outside the geographical visualisation within a dialogue, while still enabling users to investigate the geographical visualisation. Additionally, the sunburst visualisation can be accessed directly from the selection menu without using the geographical visualisation.

Additionally, a table representation providing important information about selected sensors and its measurement was considered to be suitable for analysing data.

A line chart representation highlighting implausible values, the availability of sensors but also enabling users to defined annotations was proposed and accepted.

A correlation chart, similar to the one implemented within the time series analysis framework shown in Figure 6.5, outlining the correlations of two measurements, was proposed but not considered for the first implementation phase.

Additionally, a calendar view, similar to the one shown in Figure 4.2, was introduced enabling users to analyse availability and plausibility for each day of a year separately. As this representation did not match the expectations and other visualisations had a higher relevance, it was not selected for being implemented.

Thus, four visual representations, the geographical visualisation, the sunburst visualisation, the table representation and the line charts, were selected.

In order to create a dashboard, a design, similar to the one implemented for analysing time series data was proposed, see Figure. 6.3, since requirements are similar. The dashboard contains two main areas, the selection area on the left side and the visualisation area for visually inspecting the data. Different views and the thereof resulting user roles define different levels of interactions possibilities. While the *simple view* only provides limited analysis methods, the *expert view* additionally provides further functionalities for data analysis. This includes the passivation of stations and sensors as well as annotation of sensor measurements. In contrast to this, the *admin view* can also modify station and sensor data stored in the database.

8.5 Dashboard Implementation

As already specified earlier, there are three main views: simple, expert and admin. Each of them allowing different interaction possibilities and thus, different tabs are enabled depending on the view. Regarding the simple view only *HOME* and *NEW SENSORS* are available, both of them allow investigating the data using the different visualisations. The only difference is that *NEW SENSORS* additionally allows users to filter for the first measurements for each sensor. The export view additionally provides the *ANALYSIS* and *DATA EXPORT* tab. The analysis tab enables users to investigate outliers and annotations in detail while the export tab enables users to download original sensor data or statistics on sensor data. Finally, the admin view allows modifying data by using the *DATA UPDATE* tab. In general, the structure of all tabs is similar, allowing to select sensors, time ranges, visualisations and specify additional configuration values on the left side while the corresponding information is than displayed in the main area of the dashboard. Thus, this thesis concentrates on the main functionalities of the *HOME* and *ANALYSIS* tab.

Apart from this, users have to select between current and historical data, which was required due to changes in data structure.

The main components within the dashboard are the four visualisations provided within the dashboard.

8.5.1 Geographical Visualisation

The geographical visualisation, based on Leaflet [134] library, provides an overview of all sensor stations see Figure 8.1. Different station types are visualised by using different geometric shapes within the visualisation. In this case, squares and rectangles are used to indicate different station types. Depending on the statuses of sensors within a station, the colour coding changes. Green indicates that all sensors are okay, grey that sensors are passivated, red that sensors are not okay (either values are not available or implausible) and blue means that there are no sensors defined for the station. Users can filter for stations within the visualisation, by for example selecting only those, which are not okay. The following interactions are available within the geographic visualisation:

- Activation: In order to activate the visualisation, users are required to click on the map once.

- Zooming: can be performed by either using the navigation in the upper left of the visualisation, double-click or by using the mouse-wheel.

- Panning: can be performed by dragging within the geographical visualisation.

- Open station details: requires a single click on the corresponding station within the geographical visualisation. This enlarges the geometric shape and opens a dialogue visualising the corresponding station in a sunburst visualisation, station without sensor information are displayed blue and cannot be opened.

- Close station details: requires a single click on the selected station within the geographical visualisation or clicking on the icon in the dialogue.

- Close all station details: push Shift key and click in the empty area of the geographical visualisation or click on the icon in the dialogue.

- Mark station: push the Ctrl key while clicking on a station within the geographical visualisation.

- Remove station marking: push the Ctrl key while clicking on the enlarged station within the geographical visualisation.

- Remove marking all stations: push the Ctrl key and click in the empty area of the geographical visualisation.

- Remove marking and close all stations: push the Ctrl and Shift key and click in the empty area of the geographical visualisation.

- Taking over selection from the geographical visualisation to the sensor selection area on the left: mark stations in the geographical visualisation and use corresponding icon in the visualisation header.

The geographical visualisation helps in getting an overview of those stations, which are not okay and allows to directly investigate the problems by clicking on the corresponding geometric shape.

8.5.2 Line Chart

The line chart, based on Dygraphs [12], shows the historical development of sensor values. It allows comparing neighbouring as well as similar measurements by selecting sensors of one station or sensors having the same unit. The following interactions are possible within the line chart:

- Zooming: can be performed by either using the range selector below the chart or by pushing the Shift key and selecting the area of interest by using the mouse.
- Panning: can be performed by either using the range selector below the chart or by dragging within the chart.
- Show original time range: can be performed by either using the range selector below the chart or by double-click in the chart.
- Select area for annotation: can be performed by pushing the Ctrl key and selecting the area of interest.
- Create/Change annotation: can be performed by pushing the Ctrl and Shift key and clicking on the selected area or on an existing annotation. This opens a dialogue for entering detailed information. Annotations can only be created by expert and admin users.

In the visualisation header users can select whether they want to see existing annotations, plausibility problems or areas where the sensor value was not available. Additionally, in the *ANALYSIS* tab, users can navigate through annotations or automatic detected outliers within the sensors values and get further information. For annotations, this might be the date of generation and additional information provided by the user. Regarding implausibilities, this might be information on which sensors were compared and additional statistics.

8.5.3 Sunburst

The sunburst visualisation, based on D3 [75] library, provides an overview of all sensors within a station and allows comparing sensors of different stations, see in Figure 8.2. The innermost circle describes the status of the station, which can be *okay*, *not okay* or *passivated*. The next level defines the units of the contained sensors using a predefined colour scheme. The third level describes the status of a sensor. This can be either *active and okay*, *active and not okay* or *passive* and finally the outermost level shows the sensors themselves. The colour coding provides information on the sensor data, grey means the sensor is *passivated*, red means the sensor is *implausible*, yellow means the sensor is *not available* and finally green means the sensor is *okay*.

The following interactions are possible within the sunburst:

- Zoom in: can be performed by clicking on the corresponding arcs of the sunburst.

- Show original sunburst: can be performed by clicking on innermost circle.
- Mark a single sensor or an area: can be performed by pushing the Ctrl key and clicking on corresponding arc.
- Remove marking of single sensors or area: can be performed by push the Ctrl key and click on the selected arc.
- Taking over selection from the sunburst visualisation to the sensor selection area on the left: mark sensors and use the corresponding icon in the visualisation header. For sunburst, opened within the dialogue, this needs to be done via the header of the geographical visualisation.

8.5.4 Table

The table representation is used for displaying information about selected sensors as well as providing statistical information over the selected time range. The following information is displayed for each selected sensor:

- Station id
- Sensor name
- Sensor alias
- Sensor unit
- Station type
- Sensor status (active or passive)
- Plausibility of the sensor for the selected time range
- Availability of the sensor for the selected time range
- Average sensor value of the selected time range
- Mean sensor value of the selected time range
- Mode sensor value of the selected time range
- Minimum sensor value of the selected time range
- Maximum sensor value of the selected time range
- Standard deviation of the selected time range
- Difference between minimum and maximum value of the selected time range

8.6 Case Study

A user wants to receive an overview of all stations and wants to observe whether there are some stations, which are not okay. Additionally, the user further wants to investigate those stations, which are not okay.

In order to receive an overview of all stations, the user opens the dashboard. Initially the *HOME* tab is opened and shows all stations in Austria, see Figure 8.1. Thereby the user observes that there is one station coloured red indicating that there is a problem.

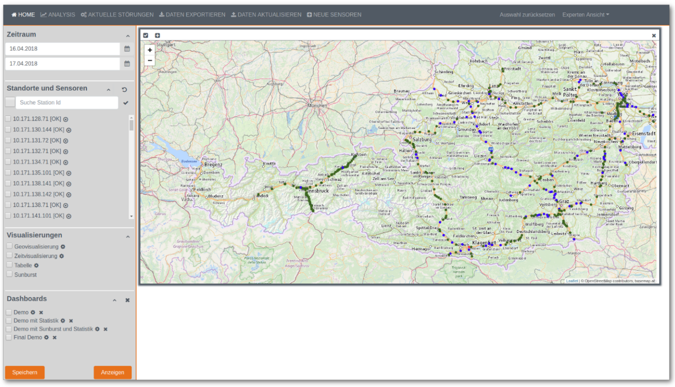


Figure 8.1: Geographical visualisation showing the status of all sensor stations.

The user zooms into the geographical visualisation and selects the corresponding station by simply clicking on the corresponding geometric shape. After clicking, the station enlarges in the geographical visualisation and a dialogue opens, containing all sensors of this station represented in a sunburst visualisation, see Figure 8.2. Now, the user can see that there are two sensors within this station which are implausible indicated by the two red areas in the outermost circle.

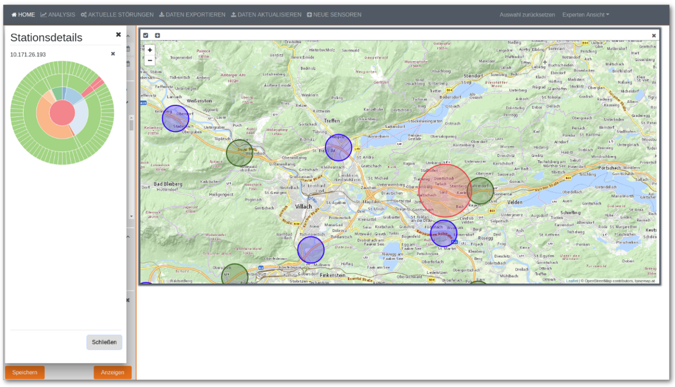


Figure 8.2: Clicking on one station within the geographical visualisation opens the sunburst visualisations containing the corresponding station details.

As the user can observe that the two sensors have the same unit, the user selects all sensors of this unit within the station by pushing the Ctrl key while clicking on the corresponding arc. This highlights the area as shown in Figure 8.3 for further comparison.

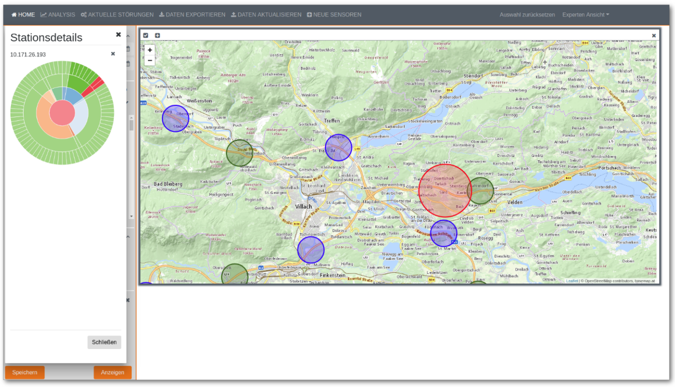


Figure 8.3: Selecting areas within the sunburst visualisation for further investigation.

Afterwards, the user closes the sunburst dialogue and applies the selection to the sensor selection view by clicking on the left icon in the header of the geographical visualisation. In order to further investigate the results, the user selects the corresponding time range in April 2018 (4th to 11th) and chooses the line chart by combining all sensors of this station in one chart. Figure 8.4 shows the resulting line chart. It shows that there are two sensors, which have several outliers.

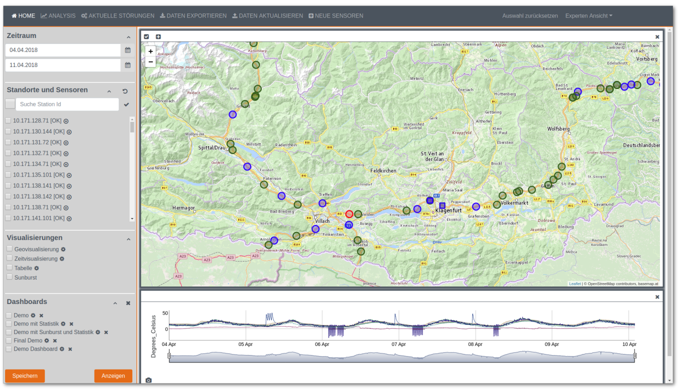


Figure 8.4: Visualisation of sensor values within the line chart.

As the user is interested in concrete values, the user zooms in to receive further details. This can be done by using the range selector below the chart or by pushing the Shift key and select the corresponding area using the mouse. Details when zooming in are illustrated in Figure 8.5.

As the user is especially interested in the two implausible sensors, the user again opens the corresponding sunburst visualisation and only selects these two sensors and again applies this in the selection view, shown in Figure 8.6.

Afterwards the user selects the table representation to receive statistical information. This is illustrated in Figure 8.7.

As the measurements of one sensor can be considered as uncritical due to former experience and the plausibility of the sensor changes to okay again, this sensor is not further

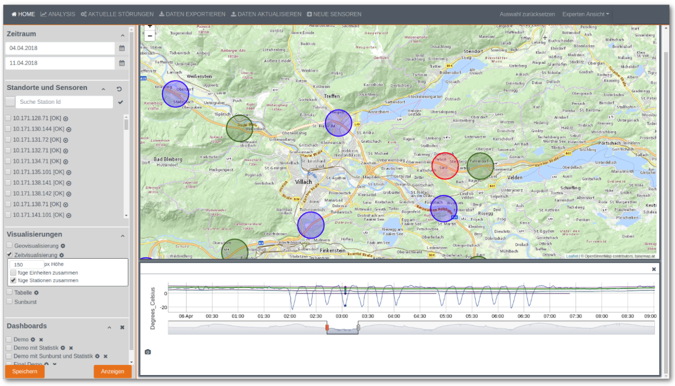


Figure 8.5: Zooming in the line charts enables users to investigate sensor measurements.

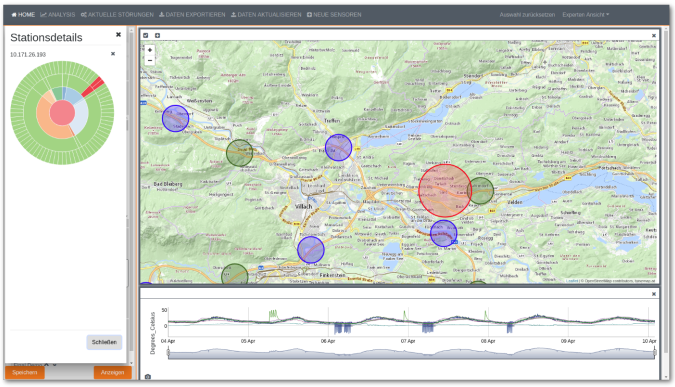


Figure 8.6: Selection of implausible sensors within the sunburst visualisation.

Figure 8.7: Visualisation of sensors within the table.

investigated. However, the other sensor is still implausible and needs further investigation. Now the user wants to investigate the same sensor in neighbouring stations. Figure 8.8 shows that there are only two stations close to the investigated one. The blue one does not contain any sensors. Therefore, the user can only select two stations and open the corresponding sunburst visualisations. After investigation, the user identifies the same sensor in both stations and selects them, see Figure 8.8.

After selecting the same sensor in both sunburst visualisations, the user selects the line chart for combining sensors of the same unit. The line chart shows that only the implausible sensor has deviating values, the other one does not have any outliers, see Figure 8.9.

As the user wants to annotate this outlier, the user switches to the *admin* view and opens the *ANALYSIS* tab to create an annotation for these outliers. This is done by first selecting the area of interest and then opening the dialogue to provide further information, see Figure 8.10.

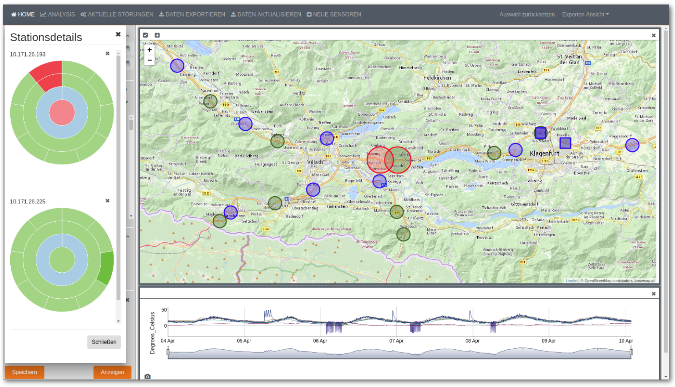


Figure 8.8: Selecting the same sensor in two neighbouring stations within the sunburst visualisations.

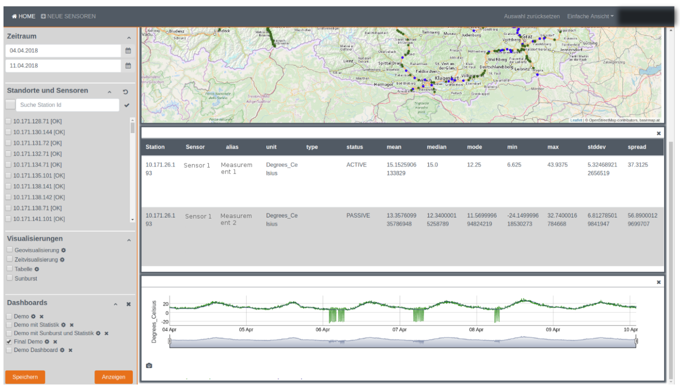


Figure 8.9: Visualising the same sensor for two neighbouring stations within the line chart.

Afterwards, the corresponding area is highlighted in orange and can be investigated in detail using the navigation on the right side. This is illustrated in Figure 8.11. Information provided by the user can support other users in analysing similar patterns.

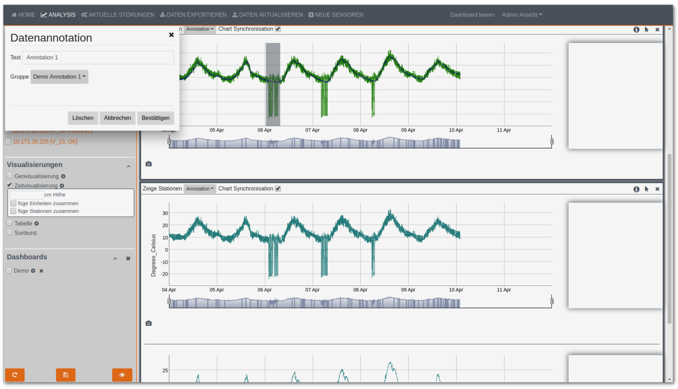


Figure 8.10: Creating an annotation for a specific sensor and a selected time range in the line chart.

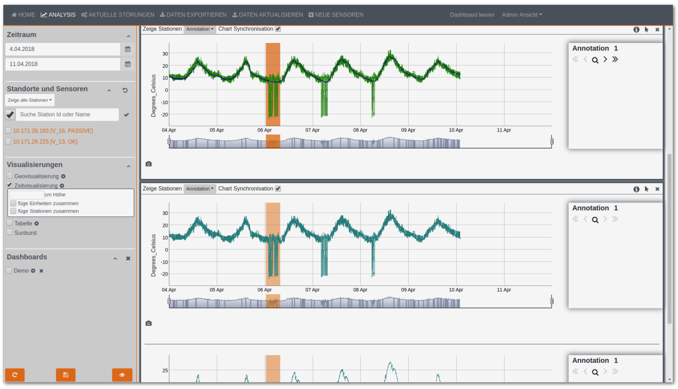


Figure 8.11: Sensor annotations are highlighted in the line chart.

8.7 Software Design and Technical Information

The dashboard has been implemented using the Angular 5 framework [31] to provide a well-structured, extendible and re-usable framework for sensor data analysis. The back-end has been implemented using Java programming language. Bundling all components was performed by using Docker [20]. Furthermore, some additional frameworks were required, including MongoDB [77] for storing general data, InfluxDB [51] for storing time series data and Kapacitor [52], which is part of InfluxDB and is responsible for processing time series data. Additionally to the provided Kapacitor tasks, which allow users to define minimum and maximum ranges for sensor values, different algorithms have been implemented by Know-Center to gain further insights on the plausibility of sensor values by considering close sensor stations.

8.8 Discussion

This section discusses whether and how the guidelines defined in section 2.4 were considered for this dashboard.

G1 (information): The dashboard enables user to investigate available sensor stations, their status and measurements by using different visualisations. As all required data can be accessed using the dashboard, this guideline was fulfilled.

G2 (visualisation): After proposing several visualisations, those, which matches the project partners expectations best, were selected. Initially, six different visualisations were suggested, a table representation, a line chart, a sunburst visualisation, a geographical visualisation, a correlation chart and a calendar visualisation. The first four of them were selected for being implemented. In order to define all required interactions, short use cases were described. While the correlation chart was not considered for being implemented in the first phase, the calendar visualisation did not match the expectations of the project partner. Those which were selected for being implemented were considered as being suitable for analysing geo-temporal data. Therefore, this guideline was fulfilled.

G3 (UI) and G4 (clutter reduction): The provided dashboard enables users to select those visualisations and data fields they want to investigate. Thus, only important data fields can be displayed by selecting the right visualisation. By enabling users to only select those visualisation and data fields they are interested in, this guideline was fulfilled.

G5 (interaction): The provided visualisations include many interaction possibilities for analysing the underlying data. As the project partner was not that familiar with using visualisations for data analysis, the visualisations and their interaction possibilities turned out to be overall complex requiring user training. Thus, providing too many interactions within the visualisations, this guideline was not fulfilled.

G6 (overview-detail): Depending on the visualisation, different insights are provided. While the geographical and sunburst visualisations always show current values, the table and line charts also allow visualising historical data. The geographical visualisation provides a good overview of all sensor stations as simple geometric shapes, coloured depending on their status. This enables the user to receive an overview without overloading the visualisation with detailed sensor information of each station. Other visualisations, like sunburst and line charts, provide details about each sensor on demand by zooming in. Therefore, this guideline was fulfilled.

G7 (colour): The colour design of the dashboard is based on the overall design of the project partner. For the different visualisations, colours were carefully selected. While the colour green indicates that everything is okay within the geographical and sunburst visualisation, the colour red is mainly used to outline problems. Colours within the line charts are automatically assigned by the library using a predefined colour scheme. Therefore, this guideline was fulfilled in most cases, except for line charts, where sense-making colour assignment is not possible.

G8 (customisation): The dashboard enables users to only select those sensors, which are interesting for them, by creating visualisations on demand. Dashboard configurations can be stored for proceeding with the analysis later on without generating each visualisation again from start. Additionally, colours and layouts were selected according to the project partner's corporate design. Therefore, this guideline was fulfilled.

G9 (devices): Since there was no need for providing a mobile application, this guideline was not fulfilled.

G10: The provided dashboard was not evaluated with test users within a study, however, feedback from the project partner revealed that some interaction possibilities were not obvious to users, who need further explanation and training. Other than that, the dashboard was considered to match expectations and thus, a plan for creating a product is being considered. Therefore, this guideline was not fulfilled.

8.9 Conclusion

This project has been conducted to enable users to investigate environmental data. By investigating spatially close sensors measuring the same variables, conclusions about the plausibility of certain sensor measurements can be drawn and visualised. This enables users to take the right actions, for example replacing or disabling a sensor if it produces wrong values on a regular basis.

The proposed dashboard can be used for different purposes, including production plants were the geographical visualisation can be replaced by a map of the corresponding production plant, highlighting specific sensor stations.

Chapter 9

Competitor Monitoring and Analysis

9.1 Introduction

This chapter discusses a dashboard created for monitoring and analysis of competitors.

The author of this thesis was responsible for the conceptual design of the dashboard, developed major dashboard and visualisation components, designed and performed an evaluation and was the first author of two full papers, one describing the search dashboard *Knowminer Search - A Multi-visualisation Collaborative Approach to Search Result Analysis* [99], and one describing the graph approach discussed in section 9.5.5.3 *Semantic Blossom Graph: A New Approach for Visual Graph Exploration* [100].

The design and implementation of this project has been a collaboration of several people. Initially, a design was proposed by the author of this thesis with contributions from Vedran Sabol and Werner Klieber. The dashboard was implemented in close cooperation with Joanneum Research [55]. The server component has been developed over several years involving many people, especially Werner Klieber and Franjo Bratić. The dashboard implementation extended by the author of this thesis, including several visualizations and interaction possibilities. This process was supported by Ralph Wozelka and Santokh Singh also working at Know-Center by adding visualizations they implemented in other projects, which are suitable for exploration. An evaluation was designed, executed and evaluated by the author of this thesis. Results of the evaluation have been published by all involved persons from Know-Center [99]. The author of this thesis thanks all involved people for collaborating throughout the whole design, implementation and evaluation phase.

9.2 Motivation

In order to monitor competitors, there are different approaches. This can range from simple information retrieval of information published on the web up to criminal actions like espionage. Retrieving the right information on the web, in enterprise storages or on private mass storage devices, is getting more difficult due to the huge amount of data. Therefore, one of the main challenges is to find the right information within acceptable time. In order to support this task, powerful text analysis methods, a powerful search engine as well as a dashboard supporting different retrieval analysis tasks is required. According to Marchionini [66] there are two ways of searching depending on the tasks: The first approach includes simple information lookup, which is usually supported by standard search engines. While the second approach, exploratory search can be split up in learning and investigation [44]. In contrast to information lookup, exploratory search requires a longer period of time while involving more interactions. Thus, competitor analysis can be assigned to exploratory search. Therefore, dashboards supporting these tasks provide different insights and interaction possibilities on the underlying result set enabling uses to draw the right conclusions based on huge amount of information available from different sources.

9.3 Requirements

The need to retrieve and analyse information on demand, search engines as well as dashboards handling huge amount of information, are required. The dashboard allows users executing search queries, but also analysing results using several visualisations, supporting the analysis process. As information on the web is usually not enriched with semantic information, powerful information retrieval and NLP (natural language processing) methods are required, for example to extract entities like organisations, persons or locations, allowing to refine search result but also for visualising. As competitor analysis requires detailed analysis of available data, the proposed dashboard must support both search approaches. Simple information lookup can be supported by providing a list of search results, as well as a faceted view for refining search results. In contrast to this, exploratory search can be supported by different visualisations providing several interaction possibilities, for example drilling-down, filtering or tasks such as getting an overview of the result data, unveiling distributions and correlations, discovering and monitoring trends, identifying anomalies as well as understanding cause and effect relationships.

The need for collecting information for further investigation can be supported by providing a portfolio collection enabling users to collect and organise results as well as share and annotate portfolios for further processing. Social features like rating as well as boosting (push a result to the top for the corresponding search query) can additionally support the analysis process.

9.4 Dashboard Design

The dashboard design was improved over several iterations, while the requirements on the proposed dashboards did not change. The iterative design process of all dashboard components is discussed in the following. The dashboard design follows well-known design proposal from search interfaces on the web:

- One of the main parts of the search dashboard is the possibility for users to perform their own search query as well as to select the corresponding source. These two components are usually positioned in the upper part of the search dashboard.
- The filter bar, which is supposed to be positioned below the search query input, shows all active filters and enables users to remove them individually or all together.
- Filters can be selected via the facet view positioned on the left side. Collapsible tree representations allow to gain an overview of the most frequent facets but enable users to also investigate less frequent facets. Additional filter opportunities are shown below the facet tree. These can include the possibility to filter for ratings.
- The actual search results are positioned in the centre of the dashboard. The information shown for each search result is similar to well-known web search tools and described later in this section.
- The portfolio view, positioned on the right of the search results, is only visible for registered users. It enables users to define their own collections of data for further processing or sharing.
- The proposed visualisations are positioned in the centre below the search results. Four visualisations were considered as being suitable for visualising search results: a geographic visualisation, a stream graph, a graph visualisation and an information landscape.

Figure 9.1 shows the first mock-up of the search dashboard. This dashboard should serve four main use cases defined by the project.

The first use case includes the need for executing search queries and visualising extracted entities. This can be performed by providing a keyword-based search UI including facets for filtering as well as visualisations for data analysis. The specification of search terms is supported by query suggestions. After selecting the right search term, a list of search results is displayed in the centre, the corresponding extracted entities are shown in facets as well as in different visualisations for further investigation. Social features like user ratings support the user in selecting the right search results. Visualisations additionally support the analysis process by displaying extracted entities and outlining correlations to other results.

The second use case is related to the portfolio view. It allows users to create topics based on their interests and makes it possible to manually collect documents related to certain topics. Their primary goal is to collect and easily access results afterwards. Optionally, documents can be shared with other users. As this is a personalised view, users are required to register before customising their own portfolios. Search results can be easily added via drag and drop to a portfolio.

The third use case handles the upload of new data to existing search indices. This feature is restricted to administrators as the search index is modified. The mock-ups pro-

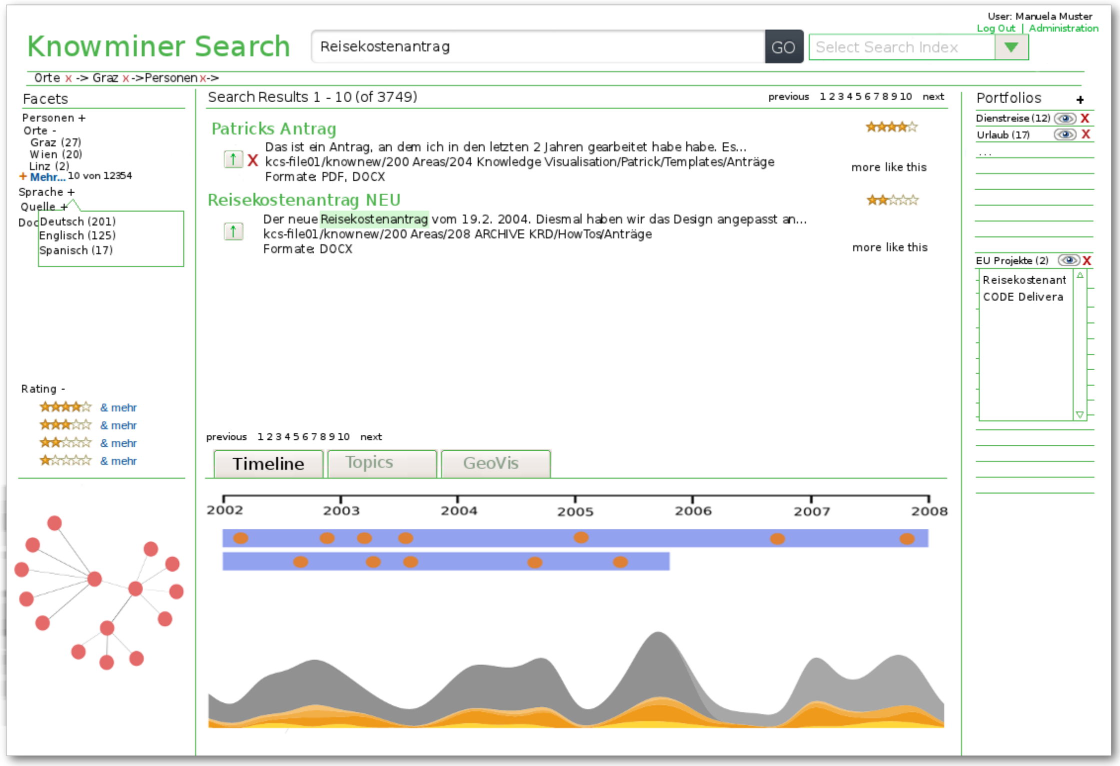


Figure 9.1: First mock-up of the search dashboard.

posed in Figure 9.2 provide five different data sources. These include simple task file uploads, web sites, file systems, databases and media wiki sources. For the task file upload only an appropriate XML (extensible markup language) file needs to be passed to the server. For web content, a distinct name, the corresponding URL, a list of linked domains, which should also be accessed and the maximum number of crawling hops. For file system crawling, one or multiple file system paths can be provided, which are recursively crawled. Connecting the data crawler to a database, detailed information needs to be provided including information about database access, table schemes and mappings and finally how data is indexed. For media wiki content, configurations are similar to those of databases.

The fourth use case is more on the technical side and concentrates on the possibility of engineers to easily set up a new version of the whole system for presentation and use case specific purposes. This starts from deploying the system, running a new indexer considering use case specific or uncritical data for demonstration purposes allowing to perform search operations on the underlying datasets. In contrast to the previous use cases this one includes the set up and deployment process but no dashboard functionalities.

As this thesis concentrates on dashboards for data analysis, instead of project set up and data import, this section concentrates on the first two use cases.

In order to receive the final dashboard, several implementation iterations were required. Figure 9.3 shows an early implementation. The first layout already includes an initial

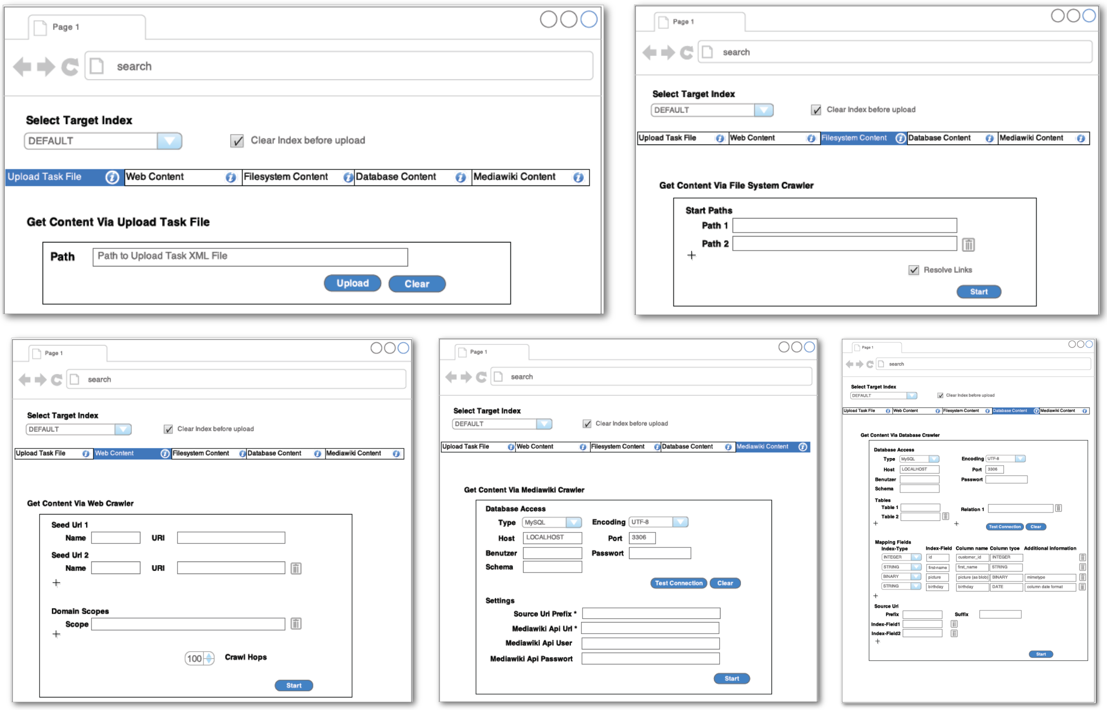


Figure 9.2: Mock-ups for uploading data from different sources.

version for user and portfolio management. No visualisations were included.

Besides improving the overall design of the dashboard, first visualisations were proposed during the second phase of the implementation, see Figure 9.4 and 9.5. In contrast to later improvements of the visualisations, their interaction possibilities were still limited. While the functionalities of the geographical visualisation, stream graph and information landscape were later extended, the D3 [75] graph visualisation, was replaced by a more sophisticated approach allowing detailed analysis of extracted entities and their relationships. Additionally, colours were assigned to specific facet groups, which were reused within the visualisations enabling users to identify certain facet groups more easily. This is also outlined by bar charts displayed in the background of each facet within the facet view.

During the last implementation phase, the existing dashboard was further refined and extended, to provide a clear structure enabling users to perform their analysis tasks. The improved visualisations are now positioned above the result set, see Figure 9.6. Detailed information about all dashboard components are described in the following.

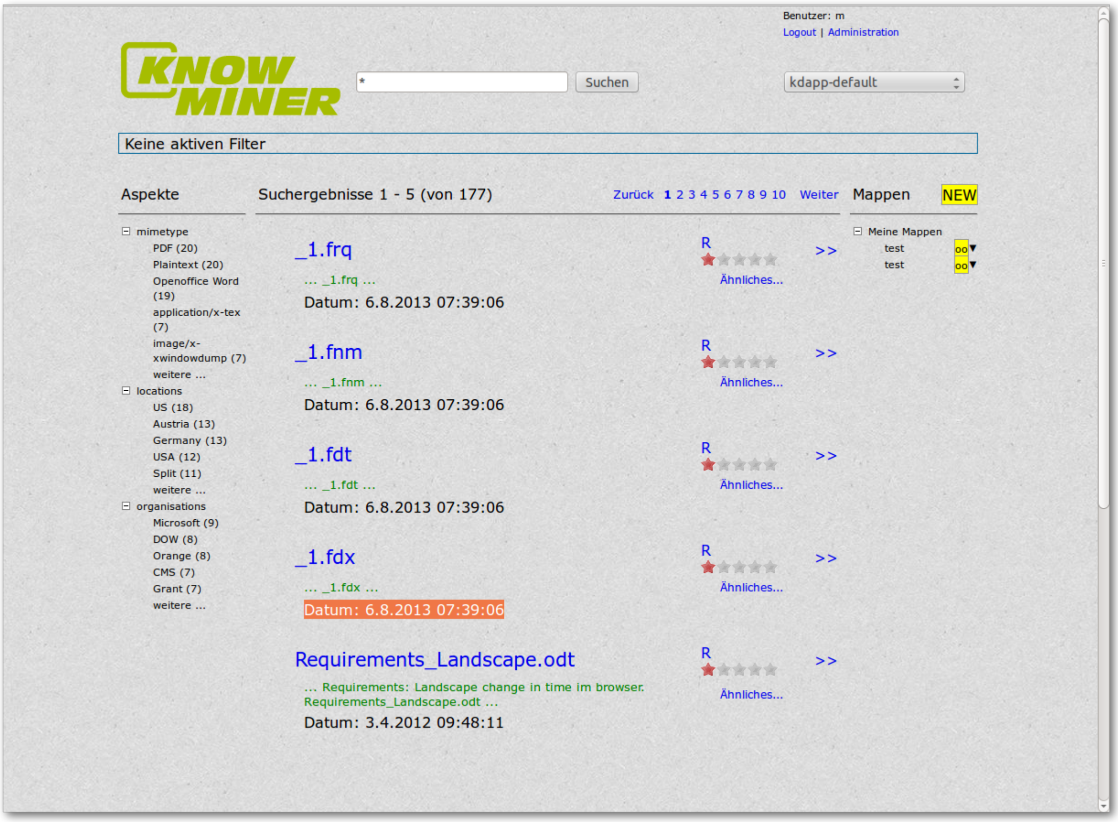


Figure 9.3: Dashboard after the first implementation phase.

9.5 Dashboard Description

9.5.1 Query Input Field and Index Selection

Users can either specify simple queries, for example *football*, more complex queries, for example *football OR tennis*, *football AND tennis* or queries even containing filters, for example *football AND source:CNN* which already contains a restriction by data source. Additionally, different indices can be used for indexing and searching data from various sources, for example web content, data from shared resources or even data from local devices. Furthermore, different indices can be used for access restriction. Finally, the query suggestion feature proposes different suggestions depending on the user input.

9.5.2 Facets and Filter Bar

The facets on the left side show extracted entities from the data sources for the corresponding search query. In general, there are seven facet groups. The most important ones are *Persons*, *Organisations* and *Locations*. Additionally, *Scope*, *Source*, *Language* and *Date* are available. The first three facet groups are expanded showing the five most fre-

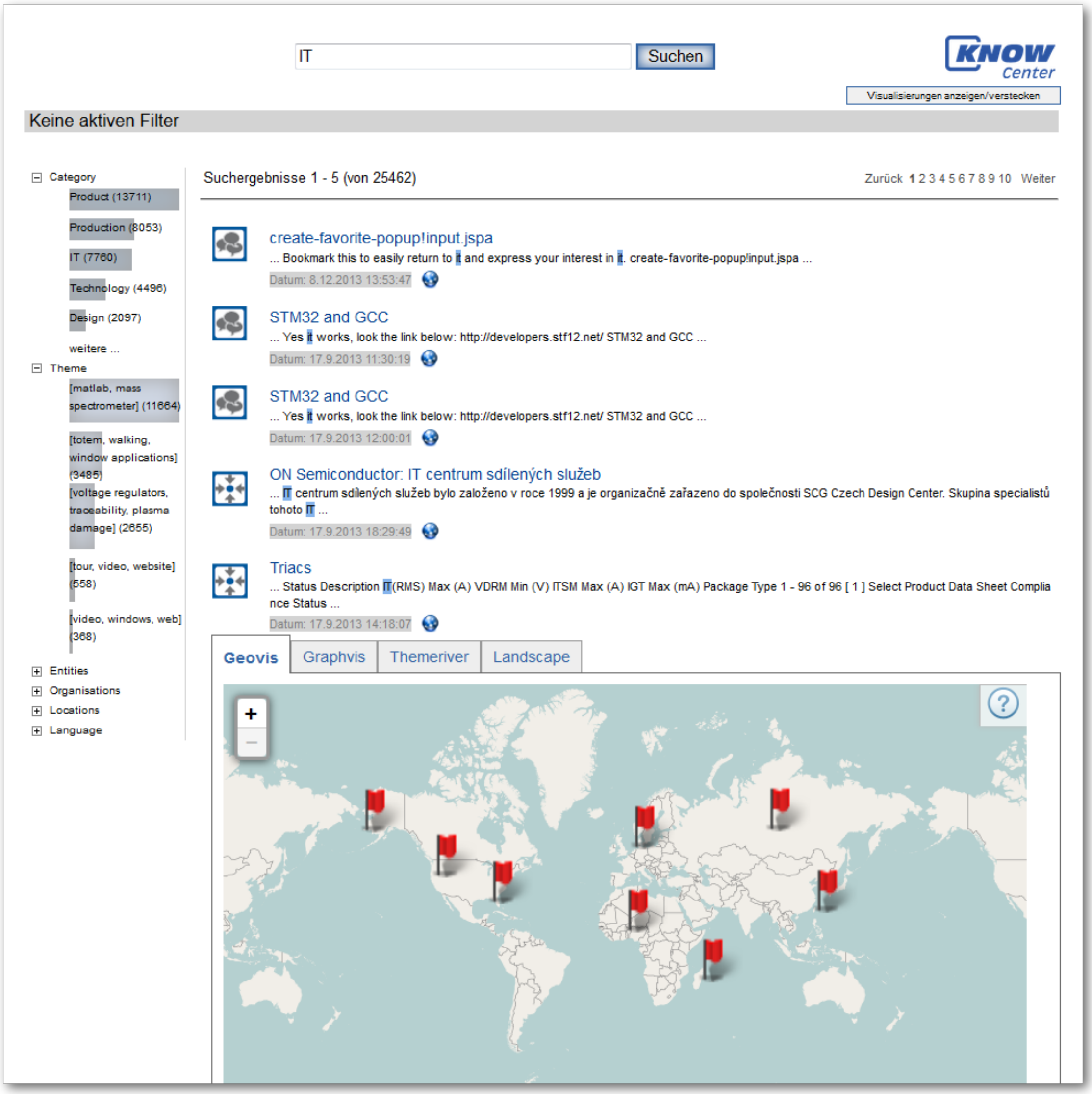


Figure 9.4: Dashboard after the second implementation phase.

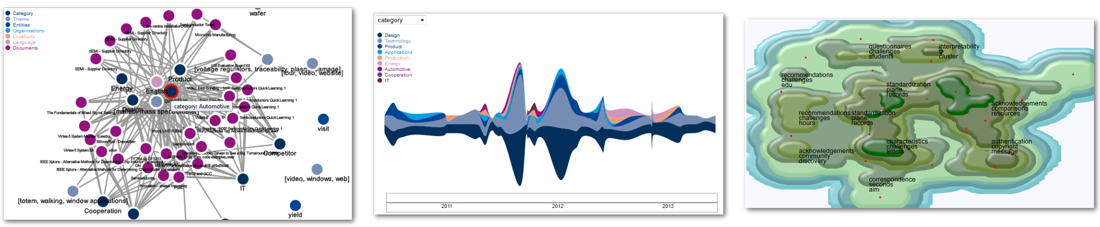


Figure 9.5: Visualisations within the dashboard after the second implementation phase.

quent terms, further groups and less frequent facets can be investigated on demand. Each facet group has its own colour, which is reused within the visualisations. This supports recognition of different facet groups more easily. Selecting one of the facets refines the search by the corresponding filter term and pushes it to the filter bar. The filter bar al-

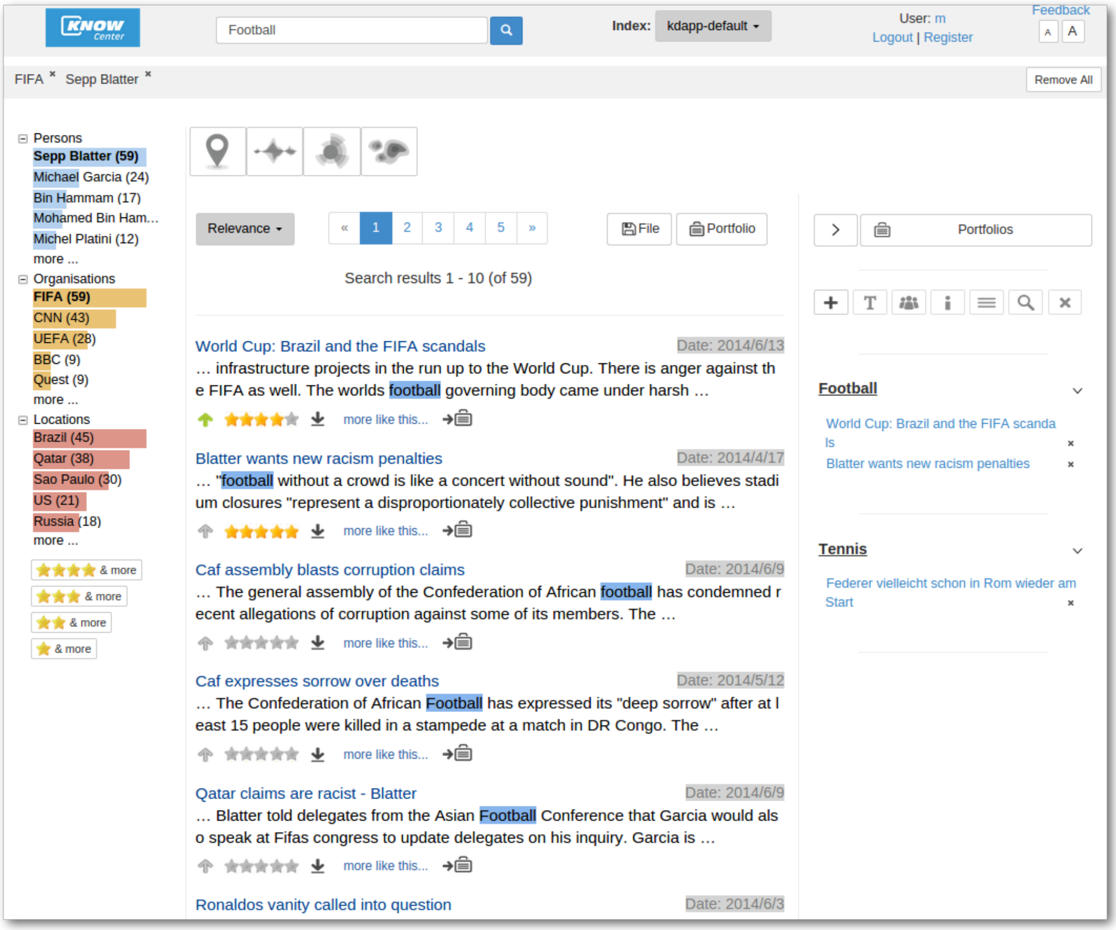


Figure 9.6: The final web-based faceted search interface.

lows users to manage all filters applied on top of the search query. Filter from the same facet group are per default *OR* connected while filters from different facet groups are *AND* connected. An example therefore is the following query *(football) AND (organisations:"FIFA") AND (persons:"Sepp Blatter" OR persons:"Michel Platini")*. Each facet in the facet view has a bar chart in the background of the name. This bar, additionally to the number in brackets, visually indicates the number of results for this facet within the facet group. Furthermore, the dashboard provides the possibility to refine the search according to user ratings. Users can select one to four stars indicating the minimum average rating of a certain search result.

9.5.3 Search Results

In the centre of the dashboard, search results are displayed according to their relevance in relation to the provided query input. Besides the relevance, results can be sorted according to user ratings or dates. The date displayed for each search result is either the modification date for documents on a storage device, the extracted date for web content or, if nothing else is available, the crawling date. In order to allow users to investigate search results different information is available. This includes the title, a snippet text providing additional content and highlighting search terms, the date of a search result and the average rating provided by all users. The dashboard provides the possibility for each search result to investigate similar results. For registered users, additional features are available. These range from boosting documents (setting the relevance of a result to the highest value for the corresponding search query) over rating of search results to organising search results in personalised portfolios.

9.5.4 Portfolios

Portfolios are shown on the right side of the dashboard and require registration since this is a personalised feature. The portfolio view provides a variety of features including standard features as adding, renaming and deleting of portfolios but also sharing of portfolios, portfolio annotations, which may be used for searching, and the visualisation of a portfolio. Portfolios are per default private and not accessible by other users. Sharing can be done globally or on user level by providing read only or read and write access. After collecting relevant search results in a new portfolio, users can annotate them. Portfolio annotations can later be used for increasing the relevance of certain results when searching for distinct terms. This means, if a portfolio is annotated using a certain term and other users search for this term, results in the portfolio receive a higher relevance compared to other results not contained in the portfolio. If users collect many search results over a longer period of time related to a certain topic, they may want to investigate these portfolios sometimes. Therefore, besides search results, the dashboard also allows to visualise single portfolios for analysis.

9.5.5 Visualisations

There are four different visualisations available within the dashboard to gain deeper insights of the underlying result set. Their purposes and interaction possibilities are described in the following. Visualisations usually update on search refinements, this can be activated or deactivated on demand for each visualisation separately. As already announced, some of the proposed visualisations were implemented for a different scope and considered as being suitable for search result exploration.

9.5.5.1 Geographical Visualisation

The geographical visualisation, see Figure 9.7 shows the distribution of locations extracted from the search results. The implementation is based on Leaflet [134] library. In general locations are extracted in the underlying dataset using gazetteer lists. Gazetteer lists are dictionaries containing location names and their geographical position using latitude and longitude values. This enables systems first, to easily identify locations by simple

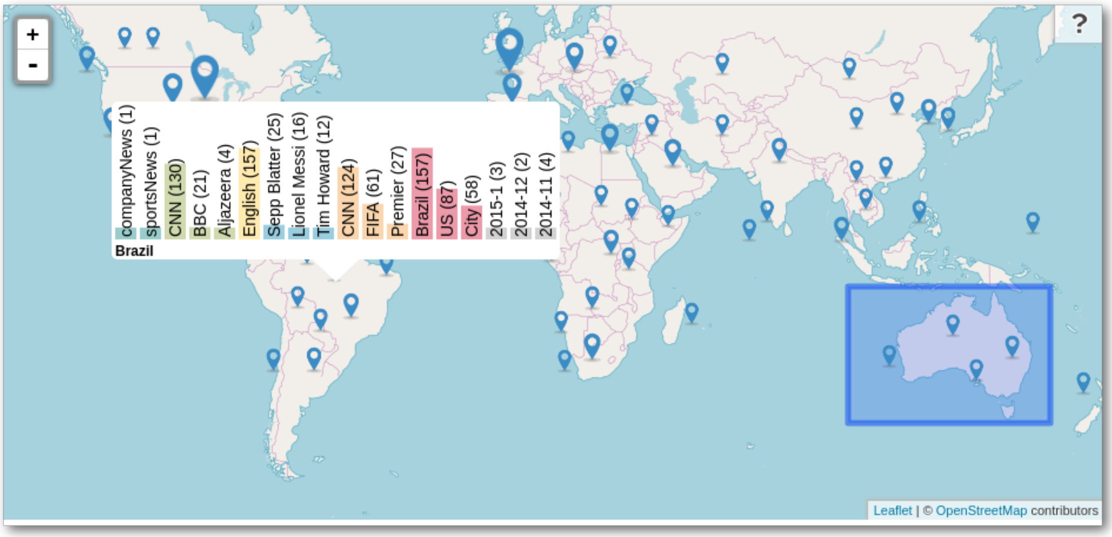


Figure 9.7: Geographical visualisation displays locations extracted from the search results.

information lookup and second, by positioning them within a geographical visualisation. Locations are encoded using simple location icons. Due to potential high density within the visualisation, close locations are aggregated indicated by the icon size. Depending on the zoom level, aggregations are resolved and locations are displayed separately. This allows users to investigate the resulting locations easily without overloading the visualisation. Tooltips provide additional information. For aggregated locations, they provide the number of locations hidden by the certain icon. For single locations, the three most relevant facets for each group are shown using bar charts. The geographical visualisation allows several interaction possibilities. Search refinement can be performed by clicking on a single or aggregated location icon. Additionally, several locations can be selected using the rectangular area selection. This can be performed by pushing the Ctrl key and selecting the area of interest. An example therefore is shown in Figure 9.7. After mouse release, all selected locations are used for search refinement; all visualisations are updated except those, which disabled updates before. Selected locations within the geographic visualisations are displayed orange. This is also true for aggregated location if it contains at least one selected location. Moving the mouse over facets in the facet view highlights the locations, which are highly correlated with the current facet. Highly correlated means that two facets often occur together with low distances in search results.

9.5.5.2 Stream Graph

The stream graph, see Figure 9.8, shows extracted facets in relation to time. The implementation is based on D3 [75] library. Each stream within the visualisation corresponds to one facet whereas streams having the same colour but only using different opacity levels always correspond to the same facet group. If more than five facets of one group are visualised, opacity levels are repeated. The thickness of each stream corresponds to

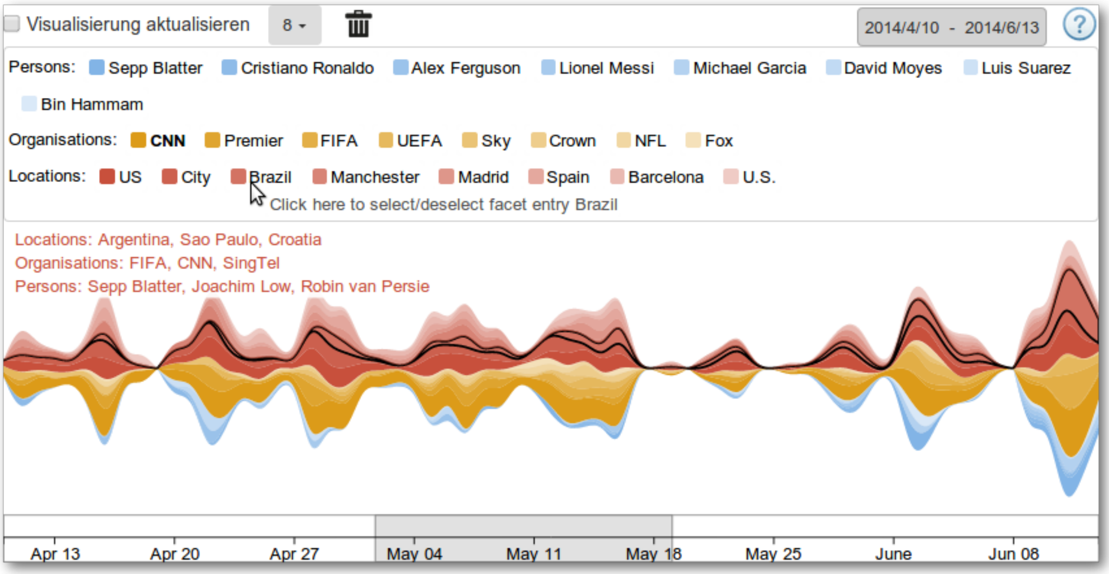


Figure 9.8: Stream graph shows most relevant facets over time.

the number of occurrences at a certain point in time. This enables users to investigate when certain facets occur but also whether there are time related correlations between two or more facets. Furthermore, users can investigate trends over time. Initially, only the three most frequent facets of each facet group are shown in the stream graph. The corresponding legend is shown above the streams. Single facets or facet groups can be easily removed by drag and drop to the corresponding icon but also by using the facet view. Additional facets can be added by increasing the number in the corresponding drop-down menu but also by adding single elements from the facet view via drag and drop or using the icon appearing next to the facet entry. Clicking on one of the legend entries or single streams allows refining search results, equally to search result refinement in the facet view. Tooltips on single streams in the stream graphs show other facets, which are highly correlated to this facet in the underlying result set, similar to tooltips on location icons within the geographical visualisation. Additionally, interactions in the facet view highlight the corresponding stream, if available, in the stream graph. The time range selection box below the stream graph shows the corresponding time range of the result set. Furthermore, it allows users to refine the search in terms of time.

9.5.5.3 Graph Visualisation

The graph visualisations, implemented in Java and translated to JavaScript by using GWT [34] (see Figure 9.9), shows relations between search results as well as facets. Initially, only the top six search results are shown within the graph. This limitation is due to users usually investigating the top search results. If the expected information is not available, the probability to refine the search by using facets or to reformulate the query is much higher than investigating less relevant results. Thus, the focus of the graph is to provide less information initially but enable users to explore the provided results. In contrast to

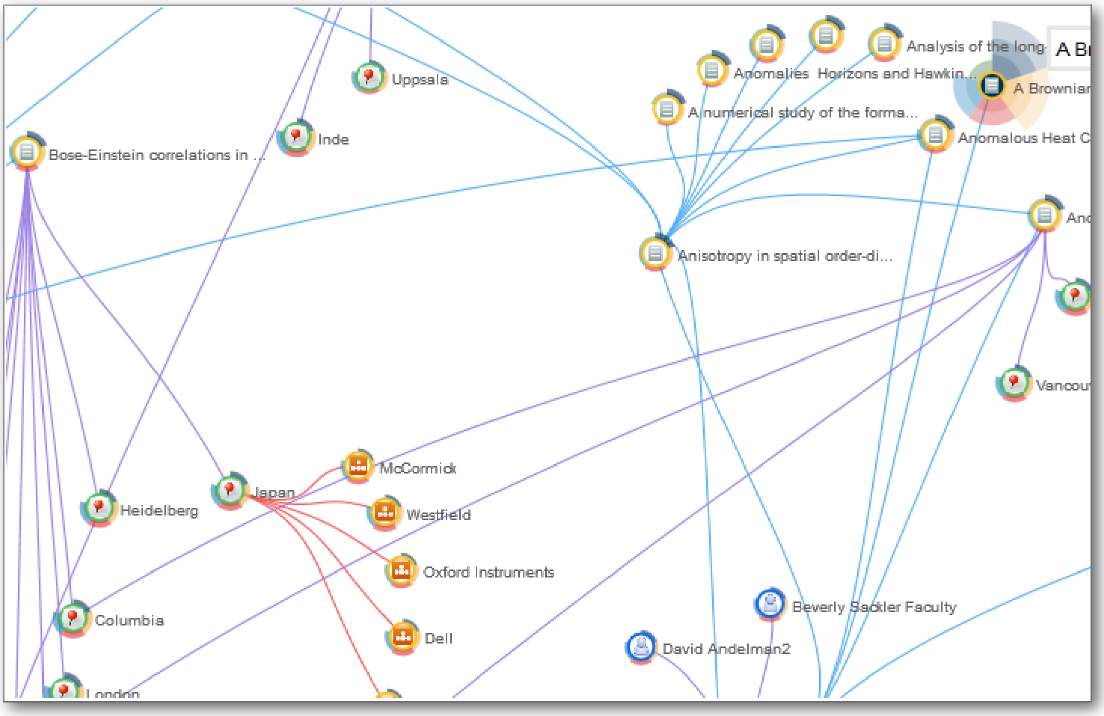


Figure 9.9: Graph visualisation shows relations between nodes representing results and entities of different types.

this, the first graph approach proposed for this dashboard contained many search results; relation to each other and to most important extracted facets, see Figure 9.5. The first approach was implemented using D3 [75] library including the force-directed placement algorithm proposed by the library. The approach discussed in Figure 9.9 provides a more sophisticated approach by only providing a small set of results and exploring relationships in distinct directions. The new graph approach distinguishes three different types of relationships. Two search results can be connected to each other if they are *similar to* each other. This relationship is based on similarities calculations considering document content. On expanding similar results in the graph visualisation, a similarity search is executed. Connections between documents and facets are called *contained in* as the facets are extracted from the documents. Finally, the third type of connection is between two extracted facets if they are *associated with* each other. Connections are shown between two facets if they often co-occur in documents. Thus, for association calculations, a co-occurrence analysis is performed on the server.

Initially, only the resulting documents are shown, these are indicated by a certain icon and colour. Several blossoms are shown in the surroundings of the document nodes, one to similar documents, all others correspond to a certain facet group. On mouse-over, the blossoms are enlarged, enabling users to perform navigation operations. There is the possibility to open the five or ten most relevant documents or facets connected to the current node. Furthermore, expanding results by considering already displayed nodes is also provided by the approach. Clicking on one of the document nodes opens the corresponding result, clicking on one of the facet nodes, performs a search refinement. Facets can be added to the graph visualisation by simple drag and drop from the facet view. This enables users to investigate any facet on demand. The proposed graph approach has been described and evaluated in context of the search dashboard [98] but also separately [100]. While evaluation results using the search dashboard are part of this thesis, results from the second evaluation only investigating the graph approach are not discussed in this thesis.

9.5.5.4 Information Landscape

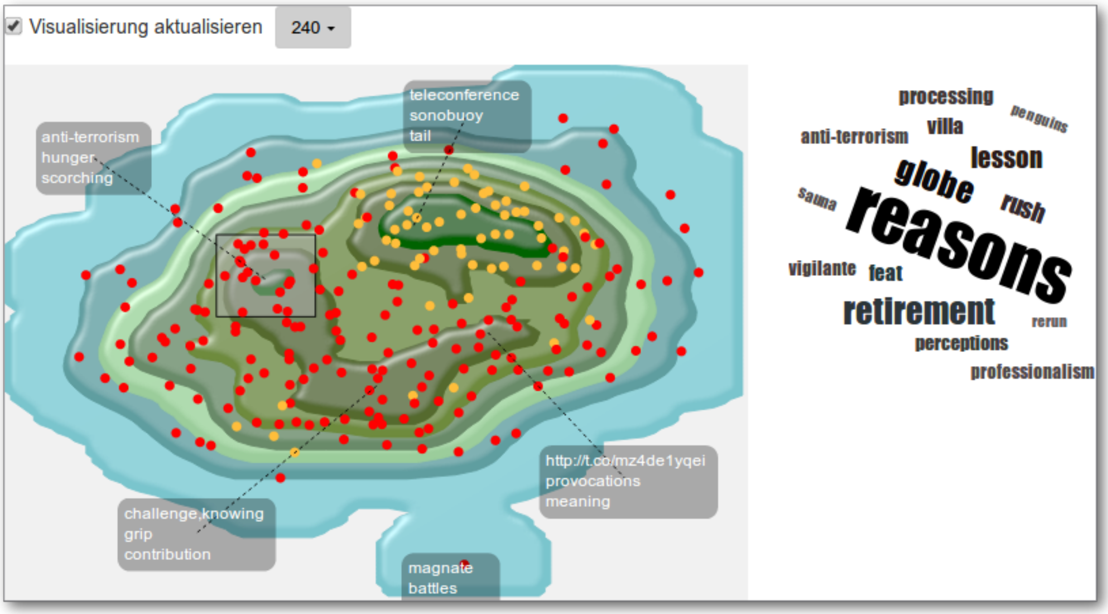


Figure 9.10: Information landscape provides a thematic overview of the search results.

The information landscape is implemented in JavaScript using SVG see Figure 9.10. It provides an overview of the topical distribution of search results. The higher the similarity of two result documents, the closer they are positioned to each other. Each search result is represented by a red dot. Areas with high density of similar projects (clusters) are surrounded by contour lines indicating hills, whereas single documents are displayed as islands. Since computing the information landscape can take some time, per default only the top 20 search results are shown. Additionally, frequent terms are displayed as labels. These labels can be repositioned while investigating the result set. Tooltips on single results provide their title, clicking on them, opens the corresponding search result. In contrast to previous versions of the information landscape this version allows to perform an area selection, displaying most important keywords in a tag cloud next to the information landscape. These keywords can be used for further search refinement. Mouse-over interactions on facets in the facet view highlight those search results in the information landscape, which contain the specific facet.

9.6 Case Study

In order to provide a demonstration application containing non-sensitive data, web pages mainly containing news articles were crawled from 2013 to 2015. The following description provides an example for analysing competitors using these news articles.

In the following scenario, the manager of a football team in Europe wants to investigate the transfer market of its competitors right after the football world championship in Brazil 2014. Thus, the manager types in ”football transfer” in the query input field and executes the search, see Figure 9.11.

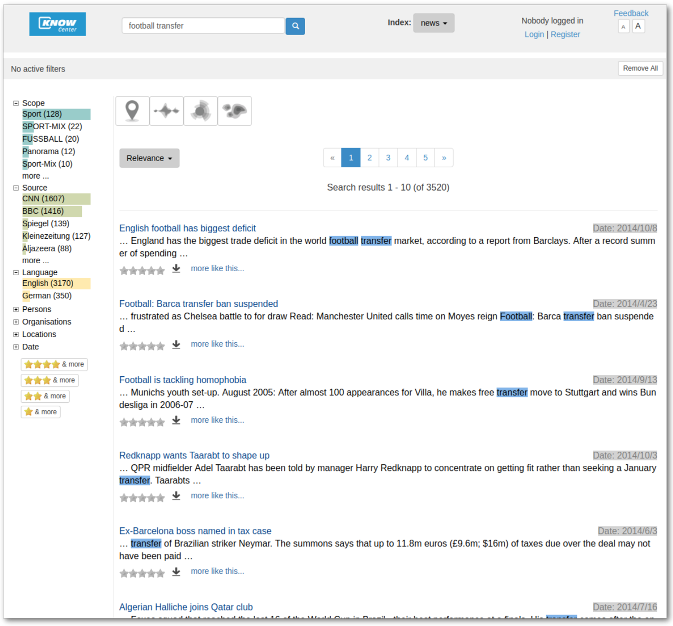


Figure 9.11: Dashboard showing search results for search query *football transfer*.

As the manager is interested in clubs located in Europe, the manager first investigates the organisation facet. As the underlying dataset was not specified for football analysis, extracted organisations are not related to football, thus the manager again closes the organisation facet group.

In order to investigate clubs by their location, the manager can see that *US*, *City*, *Brazil*, *Manchester* and *U.S.* are occurring most frequent indicated by the numbers in bracket as well as the bar charts in the background. Afterwards, the user opens the geographical visualisation, showing all locations extracted from the underlying result set, see Figure 9.12.

As the manager is interested in European clubs, the manager zooms in by using the mouse wheel. As *Real Madrid* is one of the most competitive football clubs in the world, the manager wants to start with this club. By zooming in to the highest granularity level, all locations mentioned in the result set are shown in the map, see Figure 9.13.

In order to investigate *Real Madrid*, the manager selects the corresponding location icon

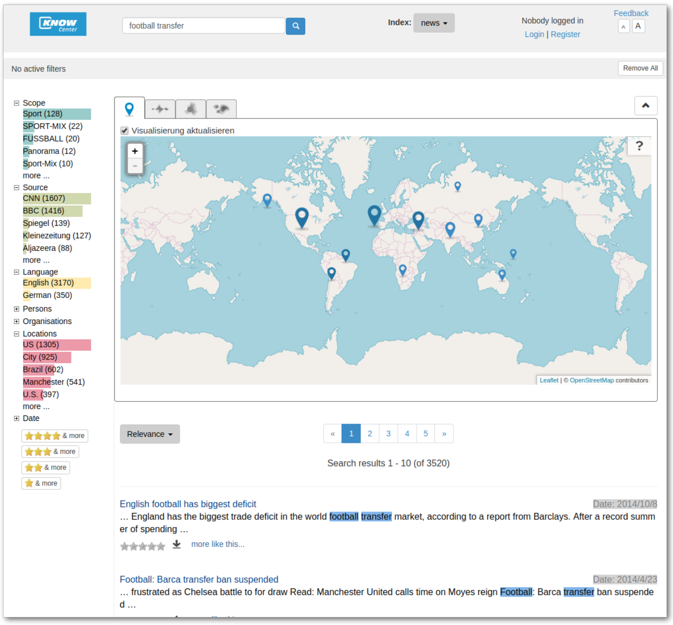


Figure 9.12: Dashboard showing geographical visualisation for search query *football transfer*.

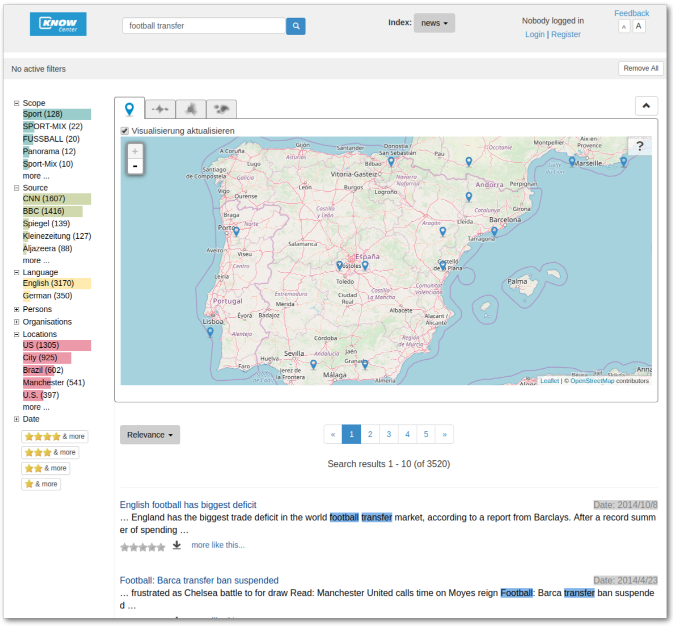


Figure 9.13: Zooming in, reveals details about locations facets.

for *Madrid* and performs a search refinement. Search results, as well as facets and visualisations, are updated. Zooming in again, shows that the corresponding location *Madrid* is highlighted in orange, see Figure 9.14.

Afterwards, the manager switches to the stream graph and sees that the results range from August 2013 to January 2015, see Figure 9.15.

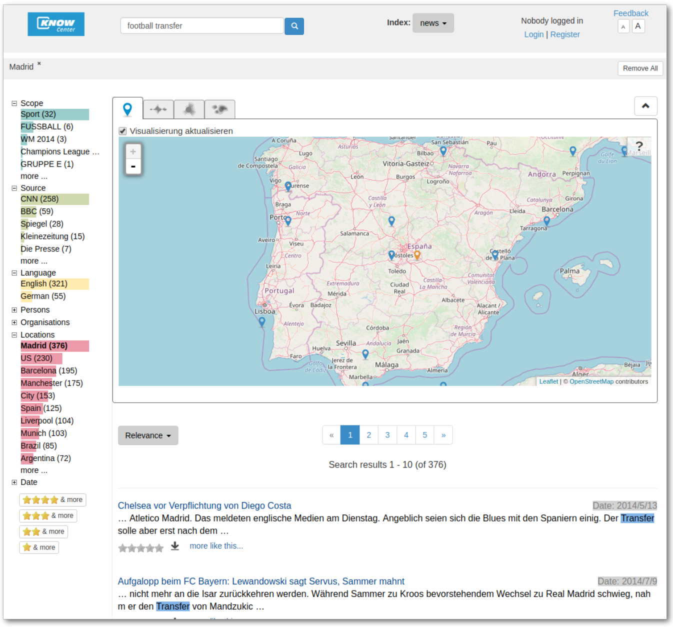


Figure 9.14: Selected locations are highlighted orange within the geographical visualisation.

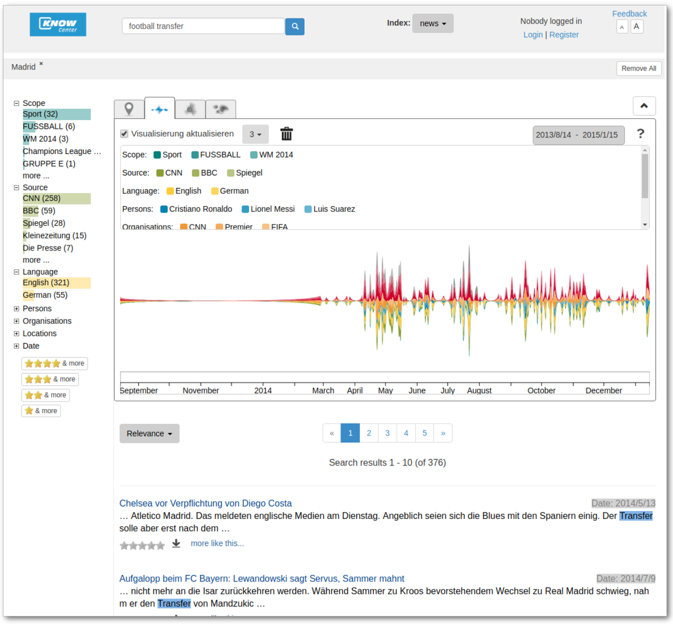


Figure 9.15: Stream graph for search query *football transfer* and facet *location: Madrid*.

Since the manager is interested in the time range right after the world championship 2014, the manager refines the search by selecting the time range of July 2014, see Figure 9.16.

For investigating more than three results, the manager sets the count for visualising to 10 in the stream graph. Additionally, the manager removes some facet groups from the

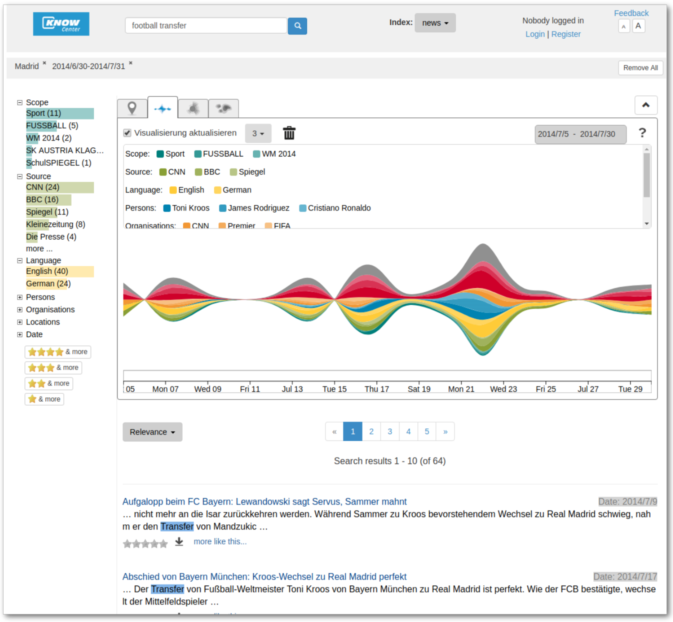


Figure 9.16: Stream graph showing football transfers in July 2014.

stream graph as they are currently not relevant. These are *scope*, *source*, *language*, *organisation* and *date*. Figure 9.17 shows the resulting stream graph. The five most frequent persons mentioned within this time range are *Toni Kroos*, *James Rodriguez*, *Cristiano Ronaldo*, *Luis Suarez* and *Gareth Bale*.

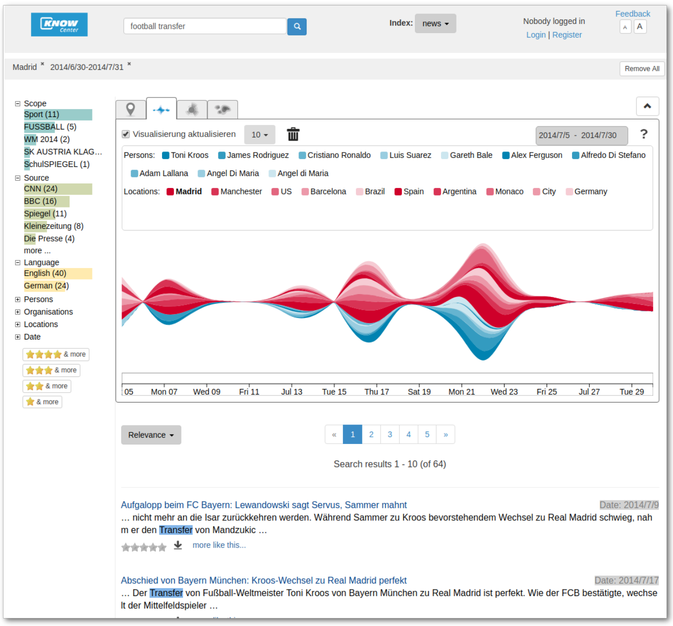


Figure 9.17: Stream graph showing football transfers in July 2014 only containing persons and locations.

Since the manager knows that *Cristiano Ronaldo* and *Gareth Bale* have both already been playing for *Real Madrid* for some time, the manager switches to the graph visualisation to investigate the three other persons. When switching to the graph visualisation, the first six results already indicate that *Toni Kroos* left *Bayern München* for joining *Real Madrid* in July 2014, see Figure 9.18.

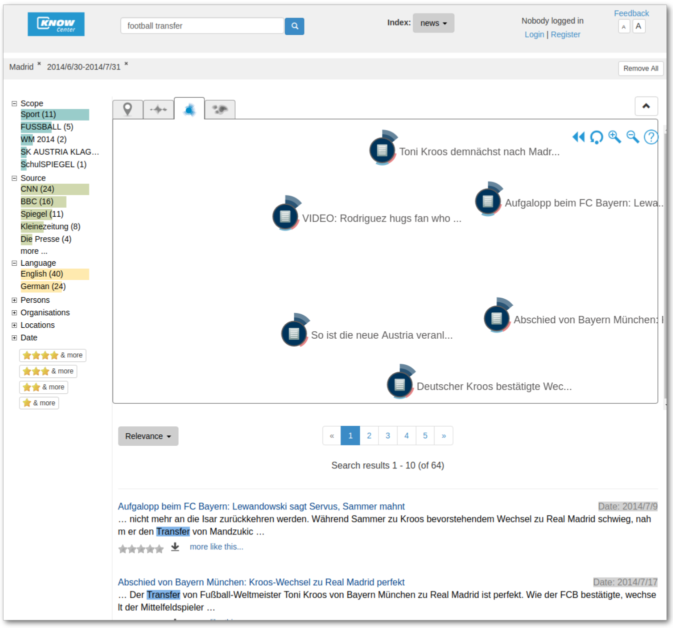


Figure 9.18: Graph visualisation showing the top six search results.

Furthermore, the manager adds *James Rodriguez* and *Luis Suarez* to the graph by drag and drop from the facet view. The manager wants to investigate whether these two also joined the club. When opening documents related to *James Rodriguez*, results show that he also joined the club in July 2014. Results for *Luis Suarez* are different. There is no indication that he joined *Real Madrid*, see Figure 9.19.

There are many news articles about *Luis Suarez* containing rumours that he might join *Real Madrid*, *FC Barcelona* or *FC Arsenal*, but there is no confirmation that he joined a specific club within the most relevant results. Thus, the manager refines the search by selecting *Luis Suarez* within the graph visualisation. After selection, the dashboard is updated and most relevant results are shown in the graph visualisation. Now, one of the results provides information that *Luis Suarez* joined *FC Barcelona*, see Figure 9.20.

Finally, the manager wants to store investigated results in a portfolio. Therefore, the manager needs to register, if not done before, and login. Afterwards, the manager creates a new portfolio called *Real Madrid* and adds relevant articles for further investigation, see Figure 9.21.

In order to investigate whether *Cristiano Ronaldo* or *Gareth Bale* leaves *Real Madrid* or why there are so many articles about them, further investigations are required. Additionally, the manager can repeat the whole procedure for further football clubs to investigate their transfers.

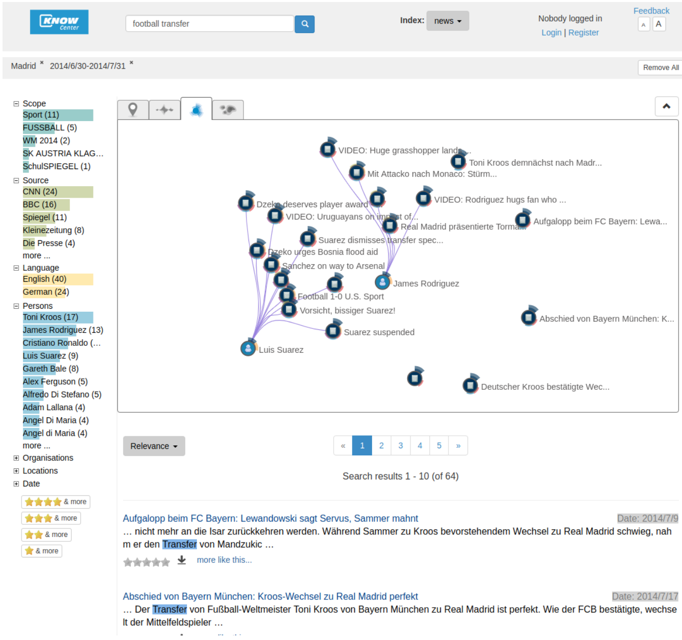


Figure 9.19: Additional graph elements can be added via drag and drop from the facet view for further exploration.

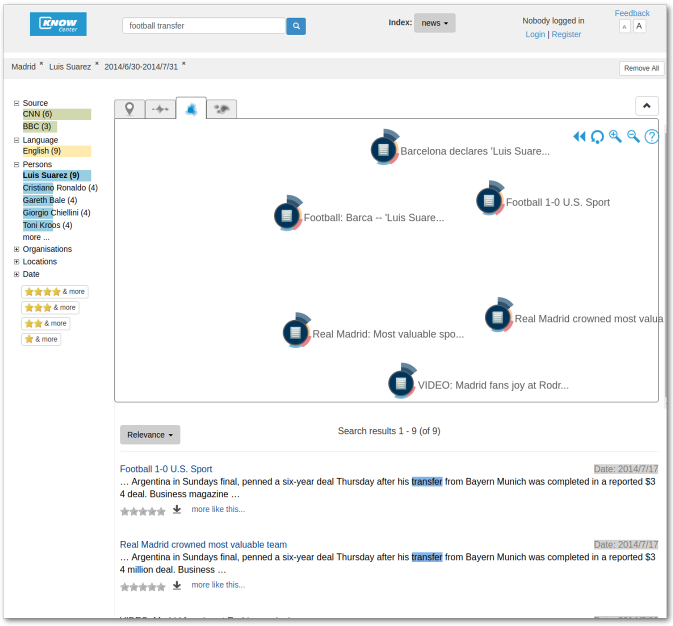


Figure 9.20: Further refinement of the search reveals that *Luis Suarez* joined *FC Barcelona*.

9.7 Software Design and Technical Information

The search dashboard was implemented in JavaScript [138] and extended by using jQuery [129]. The geographical visualisation is based on Leaflet [134] library, the stream graph and the

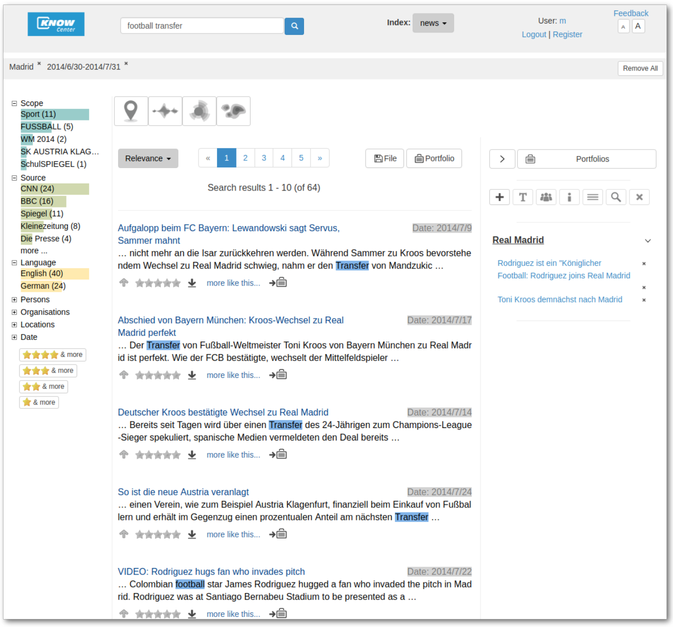


Figure 9.21: Organising interesting news article concerning football transfers in a portfolio.

information landscape are both based on SVG and implemented by using D3 [75] library. The graph visualisation was implemented in Java and compiled to JavaScript using GWT [34].

In order to provide this complex search dashboard, many different functionalities are required. This includes a search engine as well as powerful crawling and data extraction methods in the back-end. These features can be mainly categorised into five different groups illustrated in Figure 9.22:

- Data crawler/importer: These components can import data from different sources including file systems, web pages, e-mails, databases, Microsoft SharePoint and Media Wiki.
- Data conversion: This component allows converting different file types to one common format used for data indexing. Many different types are supported, for example HTML, Microsoft Office, PDF (portable document format) or Wiki.
- Semantic enrichment: This component provides many powerful methods for extracting and processing information, including language detection, key phrase detection, named entity recognition, feature engineering, natural language processing, clustering and classification of data.
- Customisation: This component enables developers to adjust the proposed dashboard and back-end for serving requirements from different domains.
- Social personalisation: This component includes user management, search personalisation, portfolio sharing as well as document boosting and rating.

Figure 9.22: Back-end components provided by Knowminer.

All proposed components are built upon a plug-in architecture supporting different server configurations depending on the underlying requirements. Since the back-end and all its requirements are not the scope of this thesis, further details are not provided.

9.8 Discussion

This section discusses whether and how the guidelines defined in section 2.4 were considered for this dashboard.

G1 (information): The dashboard enables users to search for data, which has been previously indexed. Thus, only indexed information is investigated when performing a search. Therefore, depending on the use cases, the right sources need to be added to the search index. Otherwise, the search dashboard will not provide required information. Thus, the fulfilment of this guideline highly depends on the underlying data sources, the search engine and algorithms used for data extraction. Therefore, this guideline was fulfilled.

G2 (visualisation): The dashboard provides four completely diverse visualisations enabling users to investigate different types of information. The geographical visualisation enables users to investigate extracted locations, the stream graph allows to investigate time related data, the graph visualisation highlights connections between search results and extracted entities and the information landscape outlines document similarities. Therefore, this guideline was fulfilled.

G3 (UI): In order to not overload the provided dashboard, visualisation can be minimised and enlarged depending on the users demands. This hides complexity if users only want to investigate search results. When performing more complex operations, visualisations can be enlarged and investigated. Therefore, this guideline was fulfilled.

G4 (reduce clutter) and G5 (interactions): Since the visualisations contain information extracted from search results, they can, depending on the visualisation, contain many elements. Additionally, the provided visualisations enable users to perform several interaction possibilities. The visualisations provided within the dashboard reduce clutter by aggregating locations within the geographical visualisation, by summarising results within the stream graph, by only showing the top 20 results within the information landscape and by initially showing only six documents within the graph visualisation. However, depending on the search results, many locations can be displayed within the geographical visualisation. Additionally, the visualisations provide and overwhelming number of interaction. This can be addressed by removing less relevant interaction possibilities and reducing the information provided within tooltips. Therefore, these two guidelines are only partially fulfilled.

G6 (overview-detail): The geographical as well as the graph visualisation, provide an overview by showing aggregated data (geographical visualisation) or only few elements (graph visualisation). Users can zoom in or expand elements depending on their demands. While graph approach is suitable for exploring details, the geographical visualisation can, depending on the underlying result set, contain many locations thus overloading the visualisation. The other two visualisations, stream graph and information landscape, do not provide an overview but can limit the number of displayed elements. Therefore, this guideline was only fulfilled partially.

G7 (colour): Within the search dashboard, colours were carefully selected distinguishing each facet group. Colours are reused within the visualisations enabling users to easily recognise certain groups. The set of colours can be easily configured depending on the dataset and available facet groups. Instances of the same facet type are shown using different shades of the same colour within the stream graph. Therefore, this guideline was fulfilled.

G8 (customisation): The implementation of the whole framework including server and dashboard can be customised depending on certain needs. This includes extracted facets, colours for facets and information shown in the search results. Whereas, repositioning of components within the dashboard is not possible. Therefore, this guideline was partially fulfilled.

G9 (devices): When creating the dashboard, there was no requirement on providing a mobile version, thus this guideline was not fulfilled.

G10 (evaluation): The provided dashboard was evaluated with 17 test users, see section 10.3, to investigate which visualisations and interaction possibilities are useful. The evaluation revealed that, visualisations (stream graph and graph visualisation) allowing only few interactions received a higher acceptance rate than those, which contain many interaction possibilities. Performing different tasks using different visualisation and interaction possibilities requires training to successfully perform tasks and to gain full benefits of the dashboard. Considering the visualisations, results revealed that the acceptance of the geographical visualisation was lowest. This was mainly due to displaying many locations, also including those, which have a lower significance, and due to providing too many interaction possibilities. Therefore, the geographical visualisation did not convince the participants, still, they could, in general, identify several advantages. The stream graph and graph visualisation received a higher rating as the conveyed information is easily readable, and the visualisations include simple interaction possibilities. In contrast, participants were not familiar with the concept of the information landscape. Therefore, and due to the too few interaction possibilities within the information landscape, the acceptance rate for this visualisation was lower than for the previous two. However, participants could identify some advantages in using all visualisation. As using the visualisations for data analysis is time consuming, the amount of information within the visualisations and interaction possibilities should be reduced to gain full benefits without overwhelming users. Therefore, this guideline was fulfilled.

9.9 Conclusion

The proposed dashboard was designed and developed during several iterations. Besides competitor monitoring and analysis, the proposed dashboard was created to also serve other application domains. Thus, the visualisations were carefully selected and are applicable in different domains, not focusing on competitor analysis. The dashboard design was based on well-known web-based search interfaces providing a faceted view, detailed information in the search results as well as rating and portfolio functionalities. The dashboard enables users to monitor their competitors by crawling their web pages, news articles and other sources such as social media. The huge amount of information received during these crawling activities can be investigated by using the search dashboard including various visualisations for performing different tasks. As this dashboard and the server-side functionalities can be used for different scenarios, methods for extracting information from the underlying dataset and finding information with retrieval methods is essential to raise acceptance of this search-based dashboard.

Chapter 10

Evaluation of Selected Dashboards

This chapter discusses the evaluation of three dashboards described within this thesis. These are project portfolio monitoring, analysis of production assembling and testing data and competitor monitoring and analysis. For each of them, the evaluation goals, the methodology and results are discussed in detail. The evaluation results of all three dashboards have already been published. Details on this are provided in the corresponding sections.

10.1 Software Project and Portfolio Monitoring

The dashboard design of this project has been discussed in detail in chapter 3. The evaluation results discussed in the following have been published by Rauch et al. [98]. For this evaluation a dataset was generated containing 20 artificial projects including all relevant data values required for the project symbol.

10.1.1 Goals

An evaluation was conducted to investigate

- whether the proposed symbol is appropriate for conveying project information.
- whether the amount of information encoded in the symbol is sufficient.
- whether the proposed portfolio layouts are suitable for analysing projects.
- which proposed portfolio layout is preferred for analysing projects.
- whether the proposed portfolio layouts are suitable for identifying outliers.

10.1.2 Methodology

10.1.2.1 Procedure

The evaluation was set up as follows. Initially the participants had to fill in a demographic and experience questionnaire. Afterwards they received an introduction on the purpose of the evaluation, on the project symbol as well as the provided portfolio layouts. In addition to the description, three Excel charts were proposed containing project information using off-the-shelf technologies. This included two bar charts showing the project progress for certain properties as well as a stacked bar chart for visualising work packages. Finally, the evaluation form contained two types of questions, free text as well as responses based on a Likert scale. The questionnaire as well as all additional information provided to the participants is attached to appendix A.

10.1.2.2 Tasks and Questions

In order to investigate whether the proposed project symbol and its components are suitable for conveying relevant information, participants were requested to rate the whole symbol and its components on five-point Likert scale. Additionally, participants had to score whether it is an advantage or disadvantage in showing many properties at the same time, as well as provide information on which project variables are suitable for analysis, which information is missing in the proposed symbol, and suggestions on better representation possibilities.

After proposing the two portfolio layouts, participants were requested to analyse the projects displayed in two portfolios and discuss salience. The proposed portfolio layouts were rated according to their ability to support the analysis process, whether they allow to identify groups, and which other layouts they would choose for analysing projects to gain further insights. The complete list of questions and tasks is attached to appendix A.2.

10.1.2.3 Participants

Fourteen participants performed the evaluation. Eleven out of them were males and three females. Their age ranged from 25 to 42 with a mean average age of 32.9 years. The evaluation was conducted in German as most of the participants were native speakers or had fluent German knowledge. All of them were experienced computer users, 12 out of 14 reported of having medium to high experience in using Excel-like visualizations. Only one participant had much experience in using project visualizations, eight had some experience and five only little or no experience.

10.1.2.4 Measurements

In order to investigate previously defined goals participants received an introduction to the provided project symbol and the two portfolio layouts. Ratings on a five-point Likert scale allow to conclude whether the proposed symbol and portfolios layouts are suitable for analysing projects. Additionally, small tasks identifying the root cause for a problem within a project symbol and identifying groups and anomalies within the portfolio layouts allow to conclude whether the proposed approaches are suitable for analysing projects.

10.1.3 Evaluation Results

10.1.3.1 Quantitative Results

	very good (5)	good (4)	medium (3)	bad (2)	very bad (1)	mean average
1. complete symbol	5	5	3	0	0	4.15
2. risk sign	3	8	2	1	0	3.93
3. overall progress symbol	6	3	3	2	0	3.93
4. time, money, resource symbols	8	3	3	0	0	4.36
5. project stage symbol	2	5	4	3	0	3.43
6. work package symbols	5	3	5	1	0	3.86
7. conveyed information amount	5	4	3	2	0	3.86
8. usefulness of similarity layout	2	6	3	3	0	3.5
9. find groups in similarity vis.	2	5	3	4	0	3.36
10. identify patterns in matrix vis.	2	5	3	3	1	3.26

Table 10.1: Distribution of user ratings and their average values for the project symbol, its single components and usage of the two portfolio layouts, using a five-point Likert scale.

The evaluation results base on a five-point Likert scale are shown in Table 10.1. The first six questions correspond to the project symbol and its components, question seven corresponds to the amount of information conveyed by the symbol and the last three questions handle the usage of the proposed portfolio layouts. Results show, that users highly appreciate the project symbol and their single components with a mean average rating of 3.43 to 4.36. The project stage received the lowest rating. This can be explained by the fact that the participants are not familiar with the five-stage concept used by the project partner. Users also rated the amount of information conveyed by the symbol comparable high with a mean average of 3.86. When it comes to rate the usability of the two different portfolio layouts, the mean average results were slightly lower whereas users rated the usage of the similarity layout higher than the matrix layout.

10.1.3.2 Qualitative Results

This section handles responses to the open questions. Considering the project symbol, participants appreciated that many project properties are encoded within the symbol, as it provides a quick overview of all relevant data fields, the provided design is visually appealing and the proposed symbol can be read easily without further information. In contrast to the provided positive comments, some participants criticised that the amount of information is limited due to the design of the symbol, whereas for others the amount of information conveyed within the symbol was already too high. They stated that comparisons of single data fields might be easier using bar charts and that they are not familiar with the proposed symbol. Therefore, it requires some learning to enable user in analysing projects.

The two portfolio layouts were also rated positive as they enable users in recognising project related information at one glance and better support users in identifying deviations compared to Excel sheets. Besides receiving a fast overview of the whole portfolio, the majority of participants also appreciated to be able to identify and focus on specific project groups within a portfolio layout. As for some participants the amount of information conveyed by the symbol was too much, analysing these projects within a portfolio layout was too complex and inefficient for them.

Participants were requested to identify different groups within both portfolio layouts. All participants were able to recognise the grouping of projects according to their progress within the similarity layout. Additionally, the majority of participants also recognised that projects were grouped in those, which had a risk, and those without any risk. In the matrix layout, also the majority of participants were able to identify at least two groups and that projects were mainly aligned along the diagonal indicating that project resources and time are correlated with each other. Some participants also identified the two outliers, project 11 and 14 in the matrix layout. Comparing the two portfolio layouts, participants preferred the matrix layout, as they are more familiar with aligning projects according to two data fields. Some participants did not appreciate the similarity layout, as the underlying similarity measurement was not transparent to them. Furthermore, participants would appreciate to define their own similarity measurement for visualising their portfolio.

Finally, many further suggestions for improving the project symbol have been provided. This ranged from adding information about the project priority, employees assigned to the projects or milestone information. All this information is not included in the current symbol. Some participants criticised the intuitiveness of arcs for money, time and resources increasing during the project while the actual values are decreasing over time. Other than that, for some participants it was difficult to identify project or work package sizes as they do not deviate or associate colours to certain project properties.

10.2 Analysis of Product Assembling and Testing Data

The dashboard design of this project has been discussed in detail in chapter 7. The evaluation results discussed in the following have been published by Rauch et al. [97] and Gaal et al. [26]. For this evaluation a dataset containing assembling and testing data for 15 different products was used.

10.2.1 Goals

An evaluation was conducted to investigate the following questions:

- Which data analysis and visualisation tools are currently used for analysing data?
- What are essential requirements on using tools for (visual) data analysis?
- Which visualisations have the highest relevance for the participants?
- Is Visualizer suitable for analysing data?
- Which of the provided visualisations within the dashboard are suitable for performing the following tasks:
  - Getting an overview of the data
  - Unveiling data distributions and correlations
  - Discovering and monitoring trends
  - Identifying anomalies
  - Understand cause effect relationships
- Which visualisations and interaction possibilities were used for data analysis?

10.2.2 Methodology

10.2.2.1 Procedure

The evaluation was set up in two phases. The first phase included the usage of Visualizer using the pre-configured dashboard as shown in Figure 7.12 as well as free usage to receive a good impression of the functionalities provided by the tool. In order enable participants to use the tool, they received a detailed instruction on how to use the *Table* and *Dashboard View*, see appendix C.1. Additionally, further instructions on which steps are required to transform the data to the right format, generate a dashboard and analyse the data, were provided. Section 7.6 describes how the dashboard is generated and how data can be analysed using multiple coordinated views. Participants additionally received a link for easily generating the preconfigured dashboard without having to create and configure visualisations by themselves. Participants were requested to enable interaction logging when using Visualizer to provide information about what visualisations were used and which additional interactions were performed frequently. Interaction logging was only activated on voluntary basis and could be disabled any time if concerns arose while analysing data. None of the participants raised any concerns.

The second phase of the evaluation included a detailed questionnaire to receive further insights on participants' current data analysis behaviour as well as feedback on using Visualizer while performing analysis tasks. The questionnaire contained more than 200 questions attached to appendix C.2. This included responses based on a Likert scale but also free text answers.

10.2.2.2 Tasks and Questions

The evaluation of this use case was part of a larger evaluation investigating different application domains. For each use case, different dashboards were generated tailored to their specific needs and requirements. Therefore, the author of this thesis decided against providing specific tasks for analysing data but providing instruction for dashboard generation and data analysis and request participants to analyse the provided data based on these instructions. This enabled the author of this thesis to easily analyse results over all use cases and draw the right conclusions from the results. Nevertheless, the evaluation within this thesis only includes the use case, which is described in chapter 7.

Since analysing logging interactions was not sufficient to draw conclusions, a detailed questionnaire was set up to receive as much insights as possible.

The questionnaire can be split up into the following categories:

- Demographic information
- Information about analysis and visualisation tools used before
- Questions about Visualizer including expectations, data analysis, the provided instructions as well as further functionalities
- Use case specific questions including visualisation supporting different tasks for data analysis

Filling in this questionnaire enabled participants to provide detailed feedback related to data processing, dashboard generation, data analysis, analysis by using the predefined dashboard and much more.

10.2.2.3 Participants

The questionnaire was filled in by nine male participants. Their age ranged from 24 to 47 years. Six of them originally come from Austria, one equally from France, Germany and Slovenia. Six out of them stated to be engineers, two managers and one participant was a researcher. Considering the education level, one participant has a Doctorate, five have a master's degree, one has a bachelor's degree and two finished secondary education or high school. When investigating their experience in analysing data, it turned out that four participants equally use data analysis tools *several times a week* or *once a month or less often*. Only one participant stated to *never* use data analysis tools. When analysing their data, all except one stated to use Excel, other than that only one participant stated to additionally use Power BI. When asking participants how frequent they use visualisation tools for analysing their data, five participants stated that they *never* use visualisation tools, three stated to use visualisation tools *once a month or less often* and only one participant stated to use visualisation tools *several times a month*. Tableau and Excel are respectively used by two participants whereas Power BI and think-cell are used by one participant respectively.

10.2.2.4 Measurements

Interaction logs are collected to investigate, which visualisations were generated by the participants and which further interactions were performed while analysing the data. The questionnaire investigated which proposed visualisations were rated as being suitable for different tasks, which additional visualisations are considered to be suitable for analysing their data and what additional features are required within the dashboard for analysing data. Detailed insights were provided by further ratings and free text responses.

Combining the results from both, interaction logging and questionnaire, allows to prove whether responses from the questionnaire are confirmed by tool usage and vice versa.

Furthermore, the subjective assessment for using Visualizer as well as for following the provided instructions for dashboard generation and visual analysis on a five-point Likert scale, provided information on the usability of the tool.

10.2.3 Evaluation Results

10.2.3.1 User Interactions

Participants were requested to enable interaction logging while analysing their data, to provide information on which visualisation were generated as well as which additional features and interaction possibilities were used for data analysis.


```
{
        "_id" : ObjectId("5c3ddc98bd249d04da217b84"),
        "timestamp" : "1547558167400",
        "sessionId" : "c1e85e23-23cb-4e7b-8a31-e587a358d1
           31",
        "context" : "visualization",
        "action" : "open-close-window-dialog",
        "name" : "",
        "screenWidth" : "1366",
        "screenHeight" : "768",
        "bodyWidth" : "1349",
        "bodyHeight" : "800",
        "userAgent" : "Mozilla/5.0 (Windows NT 6.3; Win64
           ; x64) AppleWebKit/537.36 (KHTML, like Gecko)
           Chrome/71.0.3578.98 Safari/537.36",
        "clientX" : "838",
        "clientY" : "86",
```

```
        "useCase" : "test",
        "visualizationId" : "1"
        "visualizationName" : "lineChart"
}
```

Listing 10.1: Example of log entry generated while performing interactions within Visualizer.

Listening 10.1 shows an example log entry for generating a line chart visualisation. The visualisation id tells analysts that this is the first visualisation generated within the dashboard. Furthermore, the dashboard belongs to the *test* use case. Additional information about the web browser and window sizes is also available for analysis. However, no sensitive information is provided by the log entry.

During the first phase of the evaluation 2084 interaction logs were generated. Figure 10.1 shows the visualisations generated (left image) and additional interactions performed (right image) while analysing data. Using the pre-defined configuration link for data analysis, did not increase the visualisation count. This count was only increased by actively generating a visualisation within the dashboard. The results show that besides proposed visualisations also scatter plot, scatter plot matrix, bubble chart, parallel coordinates and others were generated at least several times. Investigating most important interactions, participants actively changed the automatically detected field type 378 times. The brush functionality within visualisations was applied 110 times. Other interactions were executed less frequently. Two of these interactions are data aggregation (27 times) and changing of channel mapping (11 times). All these interactions were required for generating the proposed dashboard by themselves and analysing data. Other interactions, which were introduced but not required for analysing the data, were performed less frequently. One of these is sorting the data, the other is filtering.

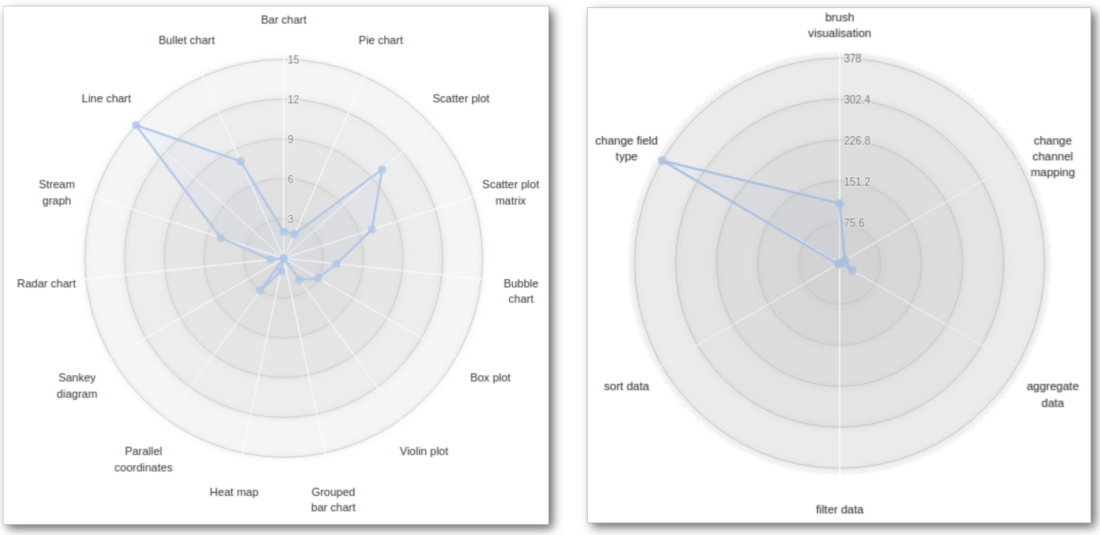


Figure 10.1: Number of interactions for generating visualisations (left) and performing additional interactions for data analysis (right).

10.2.3.2 Questionnaire

Participants stated that their primary interest when using visualisation tools are to *investigate new datasets*, *receiving a good overview of large data amounts in an easily readable format within a short period of time*, as well as *being able to efficiently explore information*.

Participants were requested for which processing steps they would use visualisation tools in general. The majority, namely five, would use them for *management presentations*. Three participants would use them for *quality assessment/assurance tasks* as well as for *investigation of production problems*. Two reported that they would use them for *getting an overview of data they are unacquainted with* and no participants would use visualisation tools for *predictive analytics* or *configuration of production steps*.

Three participants equally stated that they would use Visualizer for *quality assessment/assurance*, for *investigation of production problems*, for *predictive analytics* and for *gaining an understanding of new datasets they are unacquainted with*. Only one participant respectively would use Visualizer for *management presentations* and no one for *configuration of production setups*.

Figure 10.2 shows the average relevance of visualisation for the participants. Results show that bar charts are most important followed by line chart and pie charts. Whereas the geographic visualisation took the last place in the importance rating followed by node-link diagrams, Sankey diagrams and wafer maps.

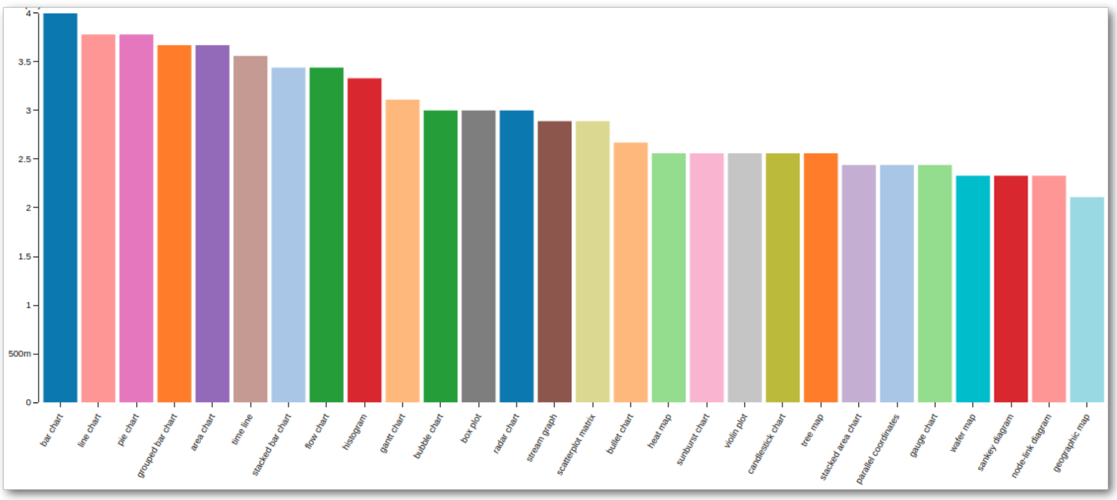


Figure 10.2: Average relevance voting of visualisation by the participants.

Limitations concerning their current data analysis tools are mainly related to processing a large number of complex data sets, which cannot be performed efficiently using for example Excel. Investigating the limitations of visualisation tools, participants stated that *identifying the right visualisation for a certain dataset is difficult or sometimes not possible at all due to not offering appropriate visualisations within these tools*. Participants appreciated the *high usability without any training*. Nevertheless, they criticised the *low* *processing time* as well as that *data cannot be displayed as preferred due to complexity*. One participant stated that he *would like to have the possibility to easily load and filter input data and automatically display data in graph representations*. Users additionally would like to *easily observe changing data fields*, *improve and optimise processes* and *identify correlations between different datasets*.

The demands on visualisation tools range from *high usability without much training*, *high performance in executing different operations* but also in receiving *a design, which highlights the data in a suitable way*. Essential requirements on a visualisation tool are, that *the tool must run on every computer*, it *considers high safety regulations due to processing sensitive data* and *should be used throughout the company or also by partners and therefore to simplify communication*.

The participants were requested what they expected when getting Visualizer introduced. Some expected *a simple tool*, some thought that Visualizer *might be a complicated application with many interaction possibilities*. Another participant expected that Visualizer *supports him in understanding the data structure and relations by using visualisation*. This participant also stated, that Visualizer matches his expectations *more than expected*. Ratings from other users were more on the average side. They ranged from *good* (1), over *average* (5) to *bad* (2). No participant provided the worst rating. For some, Visualizer was *easy to handle*, others stated that *some features were too complicated* and for some, *the instructions did not really support them in generating the dashboard and analysing data*. In order to improve the tool, *additional tooltips should be provided*, *some operations should be simplified* and *dashboard components should be repositioned, as they are not familiar with the current design*, although it reflects leading products in the market, for example Tableau. The majority of participants (5) stated that Visualizer could be an improvement/extension to their current tool, two equally agreed and disagreed. They stated that Visualizer provides *much more possibilities to show data compared to their current tool* and the tool *provides an overview of the data*. In contrast to positive comments, one participant criticised that *the amount of data currently supported by Visualizer is comparably small*.

Additionally, the participants provided ratings on the usefulness of certain functionalities. The majority of participants provided ratings of *very useful* to *medium* for all interaction possibilities. One participant respectively provided worse ratings for some interactions. Using a five-point Likert scale, average values ranged from 3.0 to 3.89. Participants rated the *reconfiguration of automatically assigned data fields to visualisation fields* as least useful. An example of reconfiguration can be the data field which is used for colour mapping within parallel coordinates or the remapping for visualising bullet chart as shown in Figure 7.10. Furthermore, *multiple coordinated views* and *table column merging* also received a comparable bad rating. The highest values were achieved by five interaction possibilities equally. These are the *filtering in the Table View*, *the dashboard creation using multiple visualisations* and *sorting*, *filtering* and *zooming* within the visualisations.

Besides all features and visualisations, also the subjective task load was measured using the NASA Task Load Index [81]. Figure 10.3 shows the distribution of personal assessment. Low values indicate low mental and physical demand, high performance, no stress, frustration or difficulty, whereas high values indicate high mental and physical demand, low performance, high stress, frustration and difficulty. The left image, outlining results for data analysis, shows that distributions are mainly between three and four on a five-point scale. Except for the physical demand, where values are mainly between two and three. These comparable bad results can be explained due to participants hardly ever use visualisation tools for data analysis.

For the evaluation, instructions on how to generate the dashboard and analyse the data were provided. Personal assessments on following these instructions are shown in the right image. Here, results are slightly better. All values are close to the average, except the frustration level, which was higher in average.

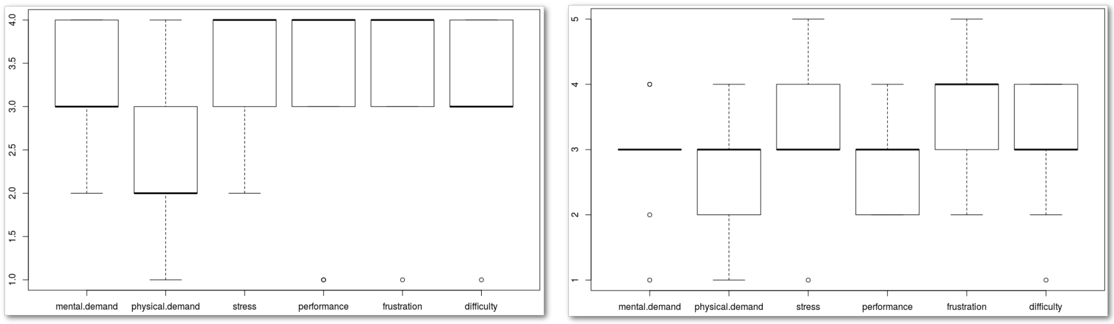


Figure 10.3: Distribution of mental and physical demand, stress, performance, frustration and difficulty while analysing the data (left) and following the provided instructions (right) on a five-point Likert scale (1: very low demand/high performance/not frustrated/not difficult 5: very high demand/low performance/very frustrated/very difficult).

Furthermore, participants were requested to rate simple statements. Users provide an *average* rating when asking *whether they want to use Visualizer frequently*. The same is true when asking for its *usability*, which was slightly better. Results concerning the *positive learning curve* and the *usability without instructions* were rated between *good* and *average*. All questions concerning the *Table View* and its features were rated *very good* to *average*. The *Dashboard View* and its interaction possibilities also mainly received *very good* to *average* ratings. Only one user respectively rated the *interactions in general* and *interactions for generating a visualisation* not that intuitive.

When investigating the usability of certain visualisations to perform distinct tasks, users rated as follows. All three visualisations, bullet chart, line chart and stream graph, were rated *average* to *very useful* in supporting all different tasks. Only one participant stated that bullet charts are *less useful* for discovering and monitoring trends. Tasks discussed in these questions are *getting an overview of the data*, *unveiling data distributions and correlations*, *discovering and monitoring trends*, *identifying anomalies* and *understanding cause effect relationships*. Finally, some other visualisations were rated as being *very useful* but only by one participant respectively. These include scatter plot matrix, heat map, bubble chart, box plot, Sankey diagram and radar chart. This is also reflected by the visualisations generated while analysing data in Visualizer.

10.3 Competitor Monitoring and Analysis

The dashboard design of this project has been discussed in detail in chapter 9. The evaluation results discussed in the following have been published by Rauch et al. [99, 100]. For this evaluation web pages mainly containing news articles were crawled and indexed from 2013 to 2015.

10.3.1 Goals

An evaluation was conducted to investigate which visualisations support data analysis and refinement and whether the provided interaction possibilities are useful and sufficient for performing analysis tasks. Additionally, analysis methods using the facet view were compared to those using different visualisations.

10.3.2 Methodology

10.3.2.1 Procedure

The evaluation was set up as follows. Initially, all participants received an introduction to the provided dashboard showing all interaction possibilities and functionalities required for the evaluation. In addition, they received a printed description on all dashboard components as described in appendix D.1. This description was provided for information lookup throughout the evaluation and includes information about the query input, the search results, the facet view, the four visualisations and the portfolio view. After the introduction of the tool, participants were encouraged to ask questions as well as to try all functionalities without time restriction. After they felt secure in handling the dashboard, the actual evaluation started.

The evaluation was made up of 24 short tasks using different components within the dashboard. By following the thinking aloud protocol, deeper insights on participants perception, while performing the evaluation, were provided. Each task either included a short instruction on what to perform or what information to investigate using the corresponding dashboard component. If users were not able to perform the corresponding task, additional information on how to perform the task was provided. While defining the tasks, an optimal procedure for performing each task was defined. Therefore, during the evaluation, tasks for each participant were evaluated whether they followed the expected procedure as for some tasks also different, mainly more complex approaches, were possible to complete the task correctly. After each task group, questions on the subjective task load using the NASA Task Load Index [81] on a five-point Likert scale were used. While one indicated the lowest values for mental and temporal demand, effort, frustration and performance, five was the highest value for these measurements. After performing all tasks, participants were requested to fill in a questionnaire to gain further insights.

10.3.2.2 Tasks and Questions

The evaluation consisted of 24 tasks. Each task is assigned to one of the six groups, handling interactions using a certain dashboard component. The original questionnaire is available in appendix D.3 also including the information for supporting the tasks. A translation of all tasks is provided in the following. The first six tasks aimed to execute search queries, analyse search results as well as perform simple filter operations using the facet view. The following tasks correspond to the first group:

1. Search for the term *Fußball* and pronounce how many results are there.
2. Which different facet types are there?
3. What is the most frequent facet?
4. Refine your search by selecting the Facet *David Alaba*.
5. What is the rating of the first search result?
6. Remove the facet *David Alaba*.

The second task group contained the following three questions and instructions using the geographical visualisation:

7. Open the geographical visualisation and investigate how many locations in Australia were found and what are their names.
8. Refine your search by selecting *Melbourne* and investigate how many search results are soccer related by investigating the search results.
9. Select all locations in Europe by using the brush functionality of the geographical visualisation. Name all selected locations.

The third task group were related to the stream graph:

10. Search again for *Fußball* and open the stream graph. Investigate which time range is covered by the search results.
11. Refine the search by selecting the time range *May 25th to June 22nd* to investigate which locations occurred most frequent, how often they occurred and name the corresponding dates.
12. Move your mouse over the stream of the most frequent location and investigate which organisations are highly correlated to this location.
13. This task contains three instructions for the stream graph:
    (a) Show ten entries for all facet groups in the stream graph.
    (b) Remove all locations from the stream graph.
    (c) Add the location *Österreich* to the stream graph.

The fourth task group investigates the usability of the proposed graph visualisation:

14. Investigate how many locations are extracted from the document *Manipulationen um Fußball-WM 2022: Auch Australien soll bestochen haben* by using the graph visualisation and name these locations.

15. Refine the current search by selecting *Katar* using the graph visualisation and investigate how many similar results are found for the result *WM in Katar: Flossen fünf Millionen Dollar Schmiergeld?*, open the ten most similar results in the graph visualisation.

16. Drag and drop the organisation *FIFA* from the facet view to the graph and investigate in how many search results it is contained.

The fifth task group consider interactions using the information landscape:

17. Reapply the search term *Fußball* and show the top 200 results in the information landscape, name the proposed labels.

18. Investigate which initially opened facets from the facet view are contained in most documents in the information landscape.

19. Select all search results positioned around the label *Vermessung, wm-fußballs* using the brush functionality and name the most important terms from the tag cloud.

Finally, the last task group handles interactions using the portfolio and search result view:

20. Register yourself and log in.

21. Rate the second search result in the result list and search for similar results.

22. Create a portfolio called *Fußball* and add a relevant search result using the search query *Fußball*.

23. Select the Portfolio *Fußball* and annotate it with the term *soccer*.

24. Share your portfolio with the user *maxmustermann* by providing read and write access.

10.3.2.3 Participants

The evaluation was performed by 17 participants; 13 males and four females with an average age of 32.8 years. Their age ranged from 29 to 48 years. The evaluation was conducted in German as all participants were native speakers. All, except one participant who reported to have medium computer experience, have much or very much experience in using computers. Additionally, only one participant stated to have little experience in using search interfaces, four stated to have medium experience, 11 have much experience and only one participant had very much experience in using search interfaces. Additionally, participants were divided into experts in using the proposed search application (5) and non-experts (12).

10.3.2.4 Measurements

In order to investigate the subjective demand for each participant, they had to fill in questions on their personal assessment using a five-point Likert scale, where one is the lowest and five the highest value respectively. The questions for each group are the following:

- How mentally demanding was it for you to execute the tasks?
- How temporally demanding was it for you to execute the task?
- How successful in performing the tasks have you been?
- How much effort did it cost you to perform the tasks?
- How frustrated have you been when performing the tasks?

Additionally, a questionnaire was filled out to gain further insights. Some questions proposed a five-point Likert scale for selection, others allowed to provided free answer possibilities. This includes questions about the main dashboard views, the provided information in the dashboard, details about which visualisations they would prefer for analysis and which are suitable for performing distinct tasks. Details about the results of this questionnaire are provided in the following.

Other than this, it was also analysed

- whether participants managed to perform the task successfully.
- whether participants performed tasks by using the expected optimal procedure or a different approach.
- whether participants required further assistance in executing the task.

10.3.3 Evaluation Results

10.3.3.1 Quantitative Results

This section discusses the quantitative results. Table 10.2 shows the task completion rate for each task dividing users by their experience in using the proposed application. The results show that the task completion rate for expert users is comparably higher than those for non-expert users. There are only two exceptions to this. In task eight, the task completion rate of experts was 40 percent and for task 18 the task completion rate was only 20 percent, even lower than for non-expert users. Other than these two tasks, the task completion was mainly at hundred percent for expert users. This might be due to being familiar with facets, knowing how they are extracted from the underlying data and therefore mean in this context. Additionally, expert users performed all tasks using the optimal procedure, which was defined in advance, not wasting time with other interactions. For non-expert users, the task completion rate ranged from 50 to 100 percent except for all tasks using the geographical visualisation (task seven to nine with task completion rates between 25 and 33 percent) and for task 18 with a task completion rate of 25 percent, which had an overall poor task completion rate. Analysing the procedure non-expert users took to receive the results, shows that non-expert users did not always take the optimal procedure for task completion. However, the majority who successfully completed the tasks did take the expected procedure.

If participants did not understand how to perform a task, they had the option to receive additional support for performing the task. The support was mainly required by non-expert users. Most participants, namely seven, required support for task nine. Additionally, four participants required support for task eight and 15, three for task seven, 13 and 16. Two participants equally required support for task 11, 14, 17, 18 and 20 to 24. Only one participant required support for task 19. For all other tasks no support was required. In comparison with the overall completion rate, this also shows that participants mainly had problems in using the geographical visualisation.

Table 10.3 shows the results of personal assessment using the NASA Task Load Index on a five-point Likert scale. The results also show that, overall, tasks for the geographical visualisations were comparably more demanding, required most effort and increased the frustration level of participants significantly. This might be due to the large number of interaction possibilities and information provided by the geographical visualisation. Results for the other visualisations are slightly better. The information landscape received best results for the personal assessment compared to other visualisations. This might be due to only providing few interaction possibilities compared to other visualisations. The first task group received best results for both expert and non-expert users. This might be due to participants being familiar in using faceted search interfaces. Table 10.3 also show that results for expert users are better than those of non-expert users. Considering task completion and personal assessment, results for both measurements are better if participants have detailed knowledge about search applications.

Additional insights were received as participants had to fill in a questionnaire. Results from those using a five-point Likert scale are provided in the following, whereas higher values are better. The proposed search dashboard uses a star rating for search results as well-known search interfaces on the web do. Therefore, the rating schema received an average rating of 3.94. Participants rated the snippet information for search results helpful to receive additional information with an average rating of 4.53, which was the highest within this survey. The boosting functionality was rated with an average value of 3.53. The facet view for search refinement also received a high rating of 4.29, including the bar charts in their background with an average rating of 4.12. Only few participants stated that bar charts in the background are disturbing. The average rating therefore was 1.29. Furthermore, the portfolio view was considered to be suitable for collecting search results with an average rating of 4.18, which is the second highest value. The proposed features for the portfolio view including annotations, sharing and visualising received an average rating of 3.68.

	Non-Experts		Experts		All	
	Completion	**Procedure**	**Completion**	**Procedure**	**Completion**	**Procedure**
Search Query, Facets and Search Results						
Task 1	100	100	100	100	100	100
Task 2	100	91.66	100	100	100	94.12
Task 3	66.67	58.33	80	80	70.59	64.71
Task 4	100	100	100	100	100	100
Task 5	100	100	100	100	100	100
Task 6	100	100	100	100	100	100
Geographical Visualisation						
Task 7	33.33	25	80	40	47.06	29.41
Task 8	25	8.33	40	40	29.41	17.65
Task 9	33.33	16.67	100	40	52.94	23.53
Stream Graph						
Task 10	91.67	83.33	100	100	94.12	88.24
Task 11	83.33	41.67	100	100	88.24	58.82
Task 12	75	58.33	100	100	82.35	70.59
Task 13	50	8.33	100	100	64.71	35.29
Graph Visualisation						
Task 14	91.67	66.67	100	100	94.12	76.47
Task 15	66.67	58.33	100	100	76.47	70.59
Task 16	75	58.33	100	100	82.35	64.71
Information Landscape						
Task 17	100	75	100	100	100	82.35
Task 18	25	8.33	20	20	23.53	11.76
Task 19	91.67	91.67	100	100	94.12	94.12
User and Portfolio Management						
Task 20	91.67	83.33	100	100	94.12	88.24
Task 21	75	66.67	80	80	76.47	70.59
Task 22	83.33	75	100	80	88.24	76.47
Task 23	100	75	100	100	100	82.35
Task 24	83.33	83.33	100	100	88.24	88.24

Table 10.2: Percentage for successfully performing tasks (Completion) by following optimal expected procedures (Procedure) divided by experience in using the proposed search interface.

10.3.3.2 Qualitative Results

The questionnaire contained several questions providing free answers. The results for these questions are discussed in the following.

When asking whether the information provided for a search result is sufficient, too much or something important is missing, the majority stated the information is appropriate. Some participants would like to receive information about the actual relevance values for each search result or providing longer snippet information, for some participants the

	mental demand	temporal demand	performance	effort	frustration
Facet tree tasks					
Non-Experts	1.5	1.7	4.2	1.8	1.2
Experts	1.2	1.0	4.6	1.0	1.0
Geographical visualisation tasks					
Non-Experts	2.9	2.1	3.1	3.2	2.5
Experts	2.0	1.6	3.6	1.6	2.2
Stream graph tasks					
Non-Experts	2.8	2.4	3.0	2.6	2.1
Experts	1.4	1.6	4.4	1.4	1.6
Graph view tasks					
Non-Experts	2.7	2.0	3.3	2.7	2.2
Experts	1.8	1.2	4.6	1.4	1.0
Information landscape tasks					
Non-Experts	2.5	2.0	3.7	2.4	2.0
Experts	1.6	1.4	4.4	1.4	1.2
Portfolio tasks					
Non-Experts	2.7	2.0	3.8	2.5	2.1
Experts	1.6	1.4	4.4	1.6	1.6

Table 10.3: Average results on personal assessment subdivided by experience in using search interfaces concerning mental demand, temporal demand, performance, effort and frustration on a five-point Likert scale, where low values indicate low demand, performance, effort and frustration, and high values indicate high demand, performance, effort and frustration.

information was already abundant. Those who stated that it was too much did not provide information on what can be removed.

Considering the proposed visualisations, participants provided information on which visualisations in the dashboard they would use for different task.

Five participants stated that they would not use the geographical visualisation for different reasons. Some of them stated that it was too complex, others could not imagine any use case related to their demand. All others stated that the geographical visualisation would be interesting for them and they would therefore use it for getting an overview of the underlying dataset and the distribution of locations as well as for search refinement.

The stream graph was considered to be not useful for four participants. Those who could identity benefits would use it for identifying trends and correlations, monitoring opinions, for getting an overview or for temporal search refinement.

Seven participants stated that they would not use the graph visualisation at all. Other participants reported to use it for navigating along similar documents, for investigating document and their content, for understanding relationships between documents, for investigating facets from the facet view using the drag and drop functionality, and for visual navigation along the dataset.

The information landscape could not convince ten participants. Others stated to use it for receiving a thematic overview, identifying thematic accumulations, for interactively describing large result sets and for similarity search for example literature review.

When investigating, which visualisations are suitable for performing different tasks the following answers were provided:

- Search refinement for geographical regions: geographical visualisation (16), stream graph (3), graph visualisation (3), facets (1)
- Overview of extracted meta data: stream graph (9), graph visualisation (9), information landscape (3), facets (1)
- Overview of topical affiliation: information landscape (12), graph visualisation (7), stream graph (2), facets (1)
- Temporal assignment of search results: stream graph (16), facets (1)
- Accumulation of meta data for certain time points: stream graph (14), graph visualisation (1)

Furthermore, participants were requested whether it is an advantage or disadvantage in having several visualisations in a search dashboard. Eleven participants could identify advantages as they provide additional methods for investigating the data receiving different insights. They make it possible to receive a better overview of the result set, allow recognising correlations, grouping results and therefore allowing better search refinement. Five participants stated that there are also disadvantages. Their main arguments were related to the large number of interaction possibilities, which can be overwhelming.

Participants also provided information for which tasks they would use the search dashboard. Most participants stated that they would use the search dashboard for their purposes if the underlying dataset has a high quality and contains relevant data. They would use it for searching news articles, literature research or for finding similar documents correlated to a certain topic. Those who would not use it stated that some features are not interactive or it is not suitable for their purposes.

The participants also provided information about the advantages and disadvantages of the proposed facets, visualisations and portfolios. The results for the facets were overall positive as they allow refining the search results, providing an overview of the search results and providing information about the frequency of each extracted facet. Disadvantages were only identified for wrongly extracted terms as well as the fact that bar charts sizes are only related to the corresponding facet group. Therefore, facets from different groups cannot be compared when investigating the bars.

Investigating the results for the visualisations showed that participants appreciated that they provide an overview, different insights, offer interaction possibilities and support users in identifying relations. Disadvantages included the difficulty of usage and information interpretation, they are often not self-explaining, and require more time for analysis.

The portfolio view also received positive and negative comments. The participants would use it for saving good search results, for personalised sharing or reusing of results or for fast access of important documents. Negative comments mainly included missing portfolio functionalities for search queries but also the request for increasing intuitiveness for certain functionalities.

Requests for additional functionalities within the search dashboard included extending the portfolio view for storing queries, direct access for example to local file storage, a specific search for scientific papers, more interaction possibilities with other users and guidance by proposing next interaction possibilities including alerts or search queries.

Finally, participants stated that they prefer Google for searching since the underlying data set is very large, it is simple to use and they are familiar with the search interface.

Chapter 11

Conclusion

11.1 Lessons Learnt

In the following insights gained during different dashboard design processes are summarised:

- In general, the acceptance in using a dashboard always depends on the underlying data. If provided information is not sufficient or is not relevant at all, users will not use a dashboard for data analysis.
- When creating station or symbol designs, it is recommended to keep the amount of information conveyed within these symbols as small as possible. Detailed information should be provided on demand. This enables users to investigate the displayed information within a short period of time and thus, increases acceptance and readability. Complex symbols or visualisations, require too much time for analysis and interpretation and decrease the acceptance dramatically.
- When using certain visualisations, the amount of information should be limited to the most relevant, details again should only be provided on demand.
- When investigating symbols or visualisations within a dashboard containing multiple elements, drawbacks on the readability can be identified more easily. Thus, investigating proposed symbols and visualisations in a larger context is essential.
- Using company specific language including colours, shapes and icons, but also using visualisation which users are already familiar with and are suitable for visualising information, increases the readability of the dashboard and raises acceptance.
- Providing a configurable dashboard, which enables users to visualise the information they are interested in, enables users with different requirements to analyse the data they are interested in without overloading the dashboard with less relevant information.
- Providing a dashboard also for mobile devices mainly depends on the company's

requirements. Since optimising a dashboard to serve requirements of mobile devices can be quite time consuming and many visualisations are not suitable for analysis using mobile devices, this should always be investigated carefully. Often, it makes more sense to provide a simplified version of the dashboard with only few interaction possibilities suitable for mobile devices.

- Besides the design process, the most important part is the evaluation at the end or for an intermediate version. Evaluation results outlined the weaknesses of the proposed dashboards and enable the author of this thesis to further improve the dashboard to meet all requirements and handle potential problems.

In the following, the most important insights gained from the evaluation results and additional feedback are discussed:

- The evaluations discussed in this thesis revealed that visualisations containing many interaction possibilities as well as many details were rated comparable poor and require much more effort when analysing data. This was partially due to participants usually do not use visual representations for analysing their data, but mainly due to the overwhelming amount of analysis opportunities and information displayed within visualisations. Example therefore are the geographical visualisations proposed in chapter 8 and 9.

- If users are not familiar with certain visualisations and the underlying calculations, visualisations are considered more critical. Thus, when providing a dashboard, the information on how visualised data is gathered or calculated should be outlined. This enables users to draw the right conclusions and raises acceptance. Additionally, enabling users to configure certain data fields and thus, influencing representations, further increases acceptance. This was confirmed by the evaluation results of the dashboard discussed in chapter 3, where users were not familiar with the similarity calculation.

- Although users are more critical if they have the impression that certain visualisations do not reveal all available information, finding a trade-off between the number of displayed elements and thereof increasing readability and providing all available information is essential. Providing dashboards where users can select the amount of information for being visualised enables them to receive an overview of a larger set or select only few details for investigation. An example therefore was the graph visualisation discussed in chapter 9, where users complained about not being able to explore more than 10 connected elements.

- The evaluation results for the dashboard in chapter 9 revealed, that even if some visualisations provide more complex interaction possibilities, learning curves might be steep if users see any benefit in using dashboards containing more complex visualisation and interaction possibilities. This is also outlined when comparing expert users with less experienced users. Task completion rate is significantly higher and results for personal assessment are better if users have more experience and training in using dashboards for analysing their data.

- The dashboards proposed in this thesis were generally considered to be suitable for their specific purposes by providing new insights and analysis method. However, as users are often not familiar with visual analysis methods, thus, training is required.

11.2 Summary

This thesis discusses dashboard designs for different industry scenarios implemented to overcome specific problems due to missing monitoring and analysis methods. In order to provide a dashboard suitable for certain purposes, it is essential to investigate the problem in detail, analyse tools and methods, which are already available and proof whether they can be extended to overcome the arising problem. If not, it is essential to identify main requirements on the dashboard to handle the problem and investigate the data, which is used for being visualised. If all these aspects have been investigated sufficiently, a dashboard considering all insights can be designed in several iterations. Thus, the main success factor for creating dashboards suitable for handling certain problems is to gain full knowledge about the problem statement and the requirements on the new approach overcoming existing problems. Furthermore, to increase acceptance, it is important to involve the project partner in the design process by considering their ideas when providing different design proposals and refine these during several iterations. Providing alternative representation possibilities enables the project partner to identify those representations, which better match their expectations for analysing their data. Nevertheless, the underlying data visualised in the dashboard requires serving the right information. If essential data is missing, this decreases acceptance and analysts will not use the dashboard for data analysis.

Evaluations have been performed for three dashboards to investigate whether the proposed solutions are suitable for performing tasks related to the problem statement and are therefore suitable for their purposes. Results were mainly positive and revealed that the proposed dashboards supported analysts in executing their tasks if guidelines are considered.

Additional feedback from project partner revealed that they appreciated the provided dashboards. In contrast to previously used tools the new dashboards supported them in monitoring and analysing data and take decisions based on the information visualised in the dashboards. Thus, the provided dashboards enabled project partners to take the right decisions and therefore having significant advantages over their competitors.

Although all dashboards discussed in this thesis handle specific scenarios, most of them can be also applied to other domains easily.

Some of the dashboards and their evaluation results have been published before. The dashboard for project portfolios analysis, described in chapter 3, has been published in *A Visual approach to Project and Portfolio Monitoring* [98]. The search dashboard, discussed in chapter 9, has been published in *Knowminer Search - A Multi-visualisation Collaborative Approach to Search Result Analysis* [99]. The graph approach provided within the search dashboard, described in section 9.5.5.3, has been discussed in *Semantic Blossom Graph: A New Approach for Visual Graph Exploration* [100]. Finally, the evaluation results for the dashboard discussing assembling and testing data, described in chapter 7, have been published, together with three other use cases, in the poster *Evaluation of Visual Decision Support Systems used in Semiconductor Industry* [97] and are described in the deliverable *Evaluation report on the effectiveness of the dashboard components* [26] in detail. The author of this thesis has been the first author of all publications and one of the main contributors to the deliverable.

11.3 Future Work

The implementations of the discussed dashboards have mainly been finished and partially implemented within companies supporting monitoring, analysis and decision processes. Some of the discussed dashboards have been considered as being suitable for a more general usage, thus the idea for developing a product, which can be applied to several domains has been further refined. Some companies appreciated the resulting dashboard but decided to continue developing the approach by themselves for strategic reasons. Others have been more on the research side investigating whether approaches can serve requirements in general and thus, require further investigation and refinement before being applied for monitoring and analysis.

Bibliography

[1] F. Alleva, R. Brennan, H. wen Hon, R. Mosur, E. Thayer, K. Lenzo, A. W. Black, E. Gouvea, D. Huggins-Daines, A. Solovets, and V. Klimkov. PocketSphinx. https://github.com/cmusphinx/pocketsphinx.

[2] Alteryx, Inc. Alteryx - Selfe-Service Data Analytics Platform. https://www.alteryx.com.

[3] A. A. Angehrn. Designing humanized systems for mulitiple criteria decision making. *Human Systems Management, 10 (3)*, pages 221–231, 1991.

[4] Apache Software Foundation. Apache Tomcat. http://tomcat.apache.org/.

[5] B. Arnaud, G. Artignan, J. Cance, G. Delmas, M. Hascoët, and N. Rodriguez. CoViz: cooperative visualization to facilitate sense making by groups of users. *In Proceedings of the 15th International Conference on Information Visualisation (IV)*, pages 255–260, 2011.

[6] Attivio. Natural Language Processing, AI-Powered Search & Text Analytics. http://www.attivio.com.

[7] M. Auer, B. Graser, and S. Biffl. An Approach to Visualizing Empirical Software Project Portfolio Data Using Multidimensional Scaling. *International Conference on Information Reuse and Integration, 2003 (IRI 2003)*, pages 504–512, 2003.

[8] Ben Sullins. Add togetherjs to your Tableau Server Implementation for real-time collaboration on data visualizations. https://gist.github.com/bsullins/6923034.

[9] J. H. Cable, J. F. Ordonez, G. Chintalapani, and C. Plaisant. Project Portfolio Earned Value Management Using Treemaps. *Project Management Program; HCI Laboratory, University of Maryland*, 2004.

[10] E. Clarkson, K. Desai, and J. Foley. ResultMaps: Visualization for Search Interfaces. *IEEE Transactions on Visualization and Computer Graphics. Los Alamitos, CA, USA*, 2009.

[11] Coveo Solutions Inc. Coveo. https://coveo.com.

[12] Dan Vanderkam. Dygraphs. http://dygraphs.com.

[13] Dassault Systemes. Netvibes Dashboard Intelligence. https://www.3ds.com/de/produkte-und-services/netvibes.

[14] Dassault Systems. 3D Design & Engineering Software. https://www.3ds.com.

[15] Dassault Systems. Manufacturing Operations Management - Delmia Apriso. https://www.3ds.com/products-services/delmia/products/delmia-apriso.

[16] Dassault Systems. Product Overview - Apriso. http://www.apriso.com/products/overview.

[17] Datapine. 10 Dashboard Desgin Principles & Best Practices To Convey Your Data. https://www.datapine.com/blog/dashboard-design-principles-and-best-practices.

[18] DigitalOcean. The MIT License. https://opensource.org/licenses/MIT.

[19] Y. Ding, Y. Sun, B. Chen, K. Borner, L. Ding, D. Wild, M. Wu, D. DiFranzo, A. G. Fuenzalida, D. Li, S. Milojevic, S. Chen, M. Sankaranarayanan, and I. Toma. Semantic Web Portal: A Platform for Better Browsing and Visualizing Semantic Data. *In Proceedings of the 2010 International Conference on Active Media Technology. Toronto, Canada*, 2010.

[20] Docker Inc. Docker - Enterprise Container Platform. https://www.docker.com.

[21] Eclipse Foundation, Inc. Jetty Runner. https://www.eclipse.org/jetty/documentation/9.4.x/runner.html.

[22] Elasticsearch B.V. Elasticsearch: RESTful, Distributed Search & Analytics. https://www.elastic.co/products/elasticsearch.

[23] S. Few. Information Dashboard Design - The Effective Visual Communication of Data. *O'Reilly*, 2016.

[24] S. Foo and D. Hendry. Desktop Search Engine Visualisation and Evaluation. *Proceedings of the 10th International Conference on Asian Digital Libraries, ICADL, Hanoi, Vietnam*, 2007.

[25] T. Fruchtermann and E. Reingold. Graph drawing by force-directed placement. *Practice & Experience (Wiley)*, pages 1129–1164, 1991.

[26] A. Gaal, M. Rauch, M. Steinlechner, M. Cordeiro, A. Zernig, A. Felsberger, and M. Rosenberger. Evaluation report on the effectiveness of the dashboard components. *Deliverable, SemI40 Power Semiconductor and Electronics Manufacturing 4.0*, 2019.

[27] Gartner Inc. Gartner Magic Quadrant of Insight Engines 2018. https://www.gartner.com/doc/3881780/magic-quadrant-insight-engines.

[28] Gartner, Inc. Magic Quadrant for Analytics and Business Intelligence Platforms. https://www.gartner.com/doc/3900992/magic-quadrant-analytics-business-intelligence.

[29] Gartner, Inc. Magic Quadrant for Data Science and Machine-Learning Platforms. https://www.gartner.com/doc/3860063/magic-quadrant-data-science-machinelearning.

[30] Gartner Inc. Magic Quadrant for Manufacturing Execution Systems. https://www.gartner.com/doc/3892763/magic-quadrant-manufacturing-execution-systems.

[31] Google Limited Liability Company. Angular. https://angular.io/.

[32] Google Limited Liability Company. AngularJS - Superheroic JavaScript MVW Framework. https://angularjs.org.

[33] Google Limited Liability Company. Google Analytics. https://analytics.google.com/analytics/web.

[34] Google Limited Liability Company. Google Web Toolkit Incubator. http://code.google.com/p/google-web-toolkit-incubator.

[35] Google Limited Liability Company. The Go Programming Language. https://golang.org.

[36] Grafana Labs. Grafana - Data Sources. http://docs.grafana.org/features/datasources.

[37] Grafana Labs. Grafana - The open platform for beautiful analytics and monitoring. https://grafana.com.

[38] Grafana Labs. Grafana Alerting. http://docs.grafana.org/http_api/alerting.

[39] Grafana Labs. Grafana Annotations. http://docs.grafana.org/reference/annotations.

[40] Grafana Labs. Grafana Plugins - extend and customize your Grafana. https://grafana.com/plugins.

[41] Grafana Labs. Grafana Sharing Features. http://docs.grafana.org/reference/sharing.

[42] H2O.ai. H2O.ai - Open Source Leader in AI and ML. https://www.h2o.ai.

[43] Hadley Wickham. ggplot2. http://had.co.nz/ggplot2.

[44] M. A. Hearst. Modern Information Retrieval: The Concepts and Technology behind Search Engines. *Addison Wesley, 2nd Edition*, 2011.

[45] Honeywell. Industrial Automation and Control Solutions from Honeywell. https://www.honeywellprocess.com.

[46] Honeywell. The Honeywell Connected Plant. https://www.honeywellprocess.com/library/news-and-events/presentations/hug-america-2018-honeywell-connected-plant-introduction.pdf.

[47] IBM Corporation. IBM Streaming Analytics. `https://www.ibm.com/cloud/streaming-analytics`.

[48] IBM Corporation. IBM Watson. `https://www.ibm.com/watson`.

[49] IBM Corporation. IBM Watson Analytics. `https://developer.ibm.com/watson-analytics`.

[50] IBM Corporation. IBM Watson Explorer. `https://www.ibm.com/support/knowledgecenter/en/SS8NLW_11.0.1/com.ibm.swg.im.infosphere.dataexpl.welcome.doc/doc/watsonexplorer_11.0.1.html`.

[51] InfluxData Inc. InfluxDB - Time Series Database Monitoring & Analytics. `https://www.influxdata.com`.

[52] InfluxData Inc. Kapacitor - A Real-Time Streaming Data Processing Engine. `https://www.influxdata.com/time-series-platform/kapacitor`.

[53] A. Inselberg. The Plane with Parallel Coordinates. *Visual Computer, Vol. 1 (2)*, pages 69–91, 1985.

[54] James Burke. RequireJS - A JavaScript file and module loader. `https://requirejs.org`.

[55] Joanneum Research Forschungsgesellschaft mbH. Joanneum Research. `https://www.joanneum.at`.

[56] P. M. Johnson, H. Kou, J. M. Agustin, Q. Zhang, A. Kagawa, and T. Yamashita. Practical automated process and product metric collection and analysis in a classroom setting: lessons learned from Hackystat-UH. *Proceedings of the 2004 International Symposium on Empirical Software Engineering, IEEE Computer Society*, page 136144, 2004.

[57] Joyent Inc. Node.js. `https://nodejs.org`.

[58] Justin Mifsud. Dashboard Design User Experience Guidelines. `https://usabilitygeek.com/dashboard-design-user-experience-guidelines`.

[59] W. Kienreich, E. Lex, and C. Seifert. APA Labs: An experimental web-based platform for the retrieval and analysis of news articles. *Proceedings of the First International Conference on the Applications of Digital Information and Web Technologies, ICADIWT*, pages 58–62, 2008.

[60] P. Kilponen and J. Vanhanen. ViCA - A WWW based tool for visualizing project status. *Proceedings of the 33rd Hawaii International Conference on System Sciences*, 2000.

[61] KNIME AG. KNIME: Konstanz Information Miner. `https://www.knime.com`.

[62] T. Kohonen. Self-Organizing Maps. *Springer, 3rd edition*, page 521, 2000.

[63] F. B. Losa, R. van den Honert, and A. Joubert. The multivariate analysis biplot as a tool for conflict analysis in MCDA. *Journal of Multi-Criteria Decision Analysis*, pages 273–284, 2001.

[64] Lucidworks, Inc. Lucidworks: Display Panels. https://doc.lucidworks.com/fusion-server/latest/system-administration/dashboards/display-panels.html#graphical-visualization-panels.

[65] Lucidworks, Inc. Lucidworks: Smart data Experiences for the enterprise. https://lucidworks.com.

[66] G. Marchionini. Exploratory Search: From Finding to Understanding. *Communications of the ACM*, 2006.

[67] Microsoft Corporation. Microsoft Azure. https://azure.microsoft.com/de-de.

[68] Microsoft Corporation. Microsoft Dynamics 365. https://dynamics.microsoft.com.

[69] Microsoft Corporation. Microsoft Excel. https://products.office.com/de-at/excel.

[70] Microsoft Corporation. Microsoft Exchange. https://products.office.com/de-at/exchange.

[71] Microsoft Corporation. Microsoft Power BI. https://powerbi.microsoft.com.

[72] Microsoft Corporation. Search in SharePoint. https://docs.microsoft.com/en-us/sharepoint/dev/general-development/search-in-sharepoint.

[73] Microsoft Corporation. Time Series Insights - Microsoft Azure. https://azure.microsoft.com/de-de/services/time-series-insights.

[74] Microsoft Corporation. TypeScript - JavaScript that scales. https://www.typescriptlang.org.

[75] Mike Bostock. Data-Driven Documents. https://d3js.org.

[76] Mindbreeze GmbH. Mindbreeze Illuminating Information. https://www.mindbreeze.com.

[77] MongoDB, Inc. MongoDB - Open Source Document Database. https://www.mongodb.com/de.

[78] Mozilla Corporation. TogetherJS. https://togetherjs.com.

[79] M. Muhr, V. Sabol, and M. Granitzer. Scalable Recursive Top-Down Hierarchical Clustering Approach with Implicit Model Selection for Textual Data Sets. *In Proceedings of the 2010 Workshop on Database and Expert System Applications, DEXA*, pages 15–19, 2010.

[80] T. Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press, 2015.

[81] NASA. NASA Task Load Index. `https://humansystems.arc.nasa.gov/groups/tlx/downloads/TLXScale.pdf`.

[82] K. Nazemi, M. Breyer, D. Burkhardt, C. Stab, and J. Kohlhammer. SemaVis: A New Approach for Visualizing Semantic Information. *Towards the Internet of Services: The THESEUS Research Program. Cognitive Technologies*, pages 191–202, 2014.

[83] NE GmbH. Brockhaus. `https://brockhaus.de/info`.

[84] Neil Turner. 10 guidelines for great Dashboard design. `http://www.uxforthemasses.com/dashboard-design`.

[85] T. N. Nguyen and J. Zhang. A Novel Visualization Model for Web Search Results. *IEEE Transactions on Visualization and Computer Graphics*, 2006.

[86] Novus Partners. NVD3 Re-usable charts for d3.js. `http://nvd3.org`.

[87] Object Refinery Limited. JFreeChart. `http://www.jfree.org/jfreechart`.

[88] Oracle Corporation. MySQL. `https://www.mysql.com`.

[89] Oracle Corporation. Oracle Database. `https://www.oracle.com/de/database`.

[90] I. Paparrizos, H. Jeung, and K. Aberer. Advanced Search, Visualization and Tagging of Sensor Metadata. *Proceedings of the 2011 IEEE 27th International Conference on Data Engineering*, 2011.

[91] PhantomJS contributors. PhantomJS - Scriptable Headless Browser. `http://phantomjs.org`.

[92] Project Jupyter. Jupyter Notebook. `https://jupyter.org`.

[93] Python Software Foundation. Python. `https://www.python.org`.

[94] QlikTech International. Analyze Data from Databases, Spreadsheets, Websites and Custom Resources. `https://www.qlik.com/us/products/qlik-connectors`.

[95] QlikTech International. Qlik: Data Analytics for Modern Business Intelligence. `https://www.qlik.com`.

[96] RapidMiner, Inc. RapidMiner: Lightning Fast Data Science for Teams. `https://rapidminer.com`.

[97] M. Rauch, A. Gaal, I. Šimić, and V. Sabol. Evaluation of Visual Decision Support Systems used in Semiconductor Industry. *19th European Advanced Process Control and Manufacturing (apcm) Conference*, 2019.

[98] M. Rauch, W. Kienreich, G. Aquila, and V. Sabol. A Visual Approach to Project and Portfolio Monitoring. *2013 17th International Conference on Information Visualisation*, pages 313–318, 2013.

[99] M. Rauch, W. Klieber, R. Wozelka, S. Singh, and V. Sabol. Knowminer Search - A Multi-visualisation Collaborative Approach to Search Result Analysis. *2015 19th International Conference on Information Visualisation*, pages 379–385, 2015.

[100] M. Rauch, R. Wozelka, E. E. Veas, and V. Sabol. Semantic Blossom Graph: A New Approach for Visual Graph Exploration. *2014 18th International Conference on Information Visualisation*, pages 234–240, 2014.

[101] S. Ribecca. Stream Graph. `https://datavizcatalogue.com/methods/stream_graph.html`.

[102] Rockwell Automation Inc. FactoryTalk Analytics Platform. `https://www.rockwellautomation.com/global/products/analytics/overview.page?pagetitle=FactoryTalk-Analytics-Platform&docid=c76b03a6a882d1802d961e37d1f00bb4`.

[103] Rockwell Automation Inc. Safety, Food Manufactury, Connected Enterprise. `https://www.rockwellautomation.com/global/overview.page`.

[104] RStudio, Inc. Shiny. `https://shiny.rstudio.com`.

[105] V. Sabol, W. Kienreich, M. Muhr, W. Klieber, and M. Granitzer. Visual Knowledge Discovery in Dynamic Enterprise Text Repositories. *In Proceedings of the International Conference on Information Visualisation (IV)*, pages 361–368, 2009.

[106] T. Saito, H. N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya, and T. Kaseda. Two-tone pseudo coloring: compact visualization for one-dimensional data. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 173–180, Oct 2005.

[107] Salesforce Corporation. Online CRM Software Systems and Solutions. `https://www.salesforce.com`.

[108] SAP. SAP Manufactoring Execution System. `https://www.sap.com/austria/products/digital-supply-chain/manufacturing.html`.

[109] SAS Institute, Inc. SAS - Analytics, Business Intelligence and Data Management. `https://www.sas.com/en_us/home.html`.

[110] ServiceNow, Inc. ServiceNow - Digital Workflows for Enterprise - Make work, work better. `https://www.servicenow.com`.

[111] V. S. Sharma and V. Kaulgud. PIVoT: Project Insights and Visualization Toolkit. *3rd International Workshop on Emerging Trends in Software Metrics (WETSoM)*, pages 63–69, 2012.

[112] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *In Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[113] Siemens AG. Siemens Fleet Manager. https://www.industry.siemens.com/services/global/en/portfolio/plant-data-services/condition-monitoring/fleet-management/pages/index.aspx.

[114] Siemens AG. Siemens MES SIMATIC IT Components. https://w3.siemens.com/mcms/mes/de/mes_komponenten/pages/default.aspx.

[115] Siemens AG. Siemens MindSphere. https://siemens.mindsphere.io/en.

[116] Siemens AG. Siemens PLM Software. https://www.plm.automation.siemens.com/global/de.

[117] Siemens AG. Siemens Visual Analyzer. https://documentation.mindsphere.io/resources/html/visual-analyzer/en-US/index.html.

[118] Siemens AG. Siemens Visual Analyzer Features. https://www.plm.automation.siemens.com/media/store/en_us/Siemens-PLM-MindSphere-Analytics-Services-fs-68317-A7_tcm29-19033.pdf.

[119] Siemens AG. Siemens Visual Explorer. https://documentation.mindsphere.io/resources/html/visual-explorer/en-US/index.html.

[120] Sinequa Corporation. Sinequa Insight Platform. https://www.sinequa.com.

[121] Sitecore Corporation. Sitecore: Integrated .NET CMS Platform & E-Commerce Solution. https://www.sitecore.com.

[122] Sonar. Sonar open platform to manage code quality. http://www.sonarsource.org, 2012.

[123] S. Steinarsson. Downsampling Time Series for Visual Representation. *Master Thesis, University of Iceland, Reykjavik*, 2013.

[124] Tableau Corporation. Tableau Software. http://www.tableau.com/public.

[125] Taras Bakusevych. 10 rules for better dashboard design. https://uxplanet.org/10-rules-for-better-dashboard-design-ef68189d734c.

[126] The Apache Software Foundation. Apache Cordova. https://cordova.apache.org.

[127] The Apache Software Foundation. Apache Lucene. http://lucene.apache.org/core.

[128] The Apache Software Foundation. Apache Zeppelin. https://zeppelin.apache.org.

[129] The jQuery Foundation. jQuery. https://jquery.com.

[130] The MathWorks, Inc. MathWorks - Matlab and Simulink Conferences. https://mathworks.com.

[131] The PostgreSQL Global Development Group. PostgresSQL: The world's most advanced open source database. https://www.postgresql.org.

[132] The R Foundation. The R Project for Statistical Computing. https://www.r-project.org.

[133] ThoughtSpot, Inc. ThoughtSpot: Search and AI-Driven Analytics. https://www.thoughtspot.com.

[134] Vladimir Agafonkin. Leaflet - a JavaScript library for interactive maps. https://leafletjs.com.

[135] I. Šimić. VISUALIZER - An Extensible Dashboard for Personalized Visual Data Exploration. *Master Thesis, Graz University of Technology*, 2018.

[136] W3C. Resource Description Framework. https://www.w3.org/RDF.

[137] Wikipedia. Dashboard (business). https://en.wikipedia.org/wiki/Dashboard_(business).

[138] Wikipedia. JavaScript. https://en.wikipedia.org/wiki/JavaScript.

[139] G. Zheng. A Multidimensional and Visual Exploration Approach to Project Portfolio Management. *Doctoral Dissertation, Georgia State University*, 2009.

[140] Zoho Projects. Online Project Management Software & Tool. http://www.zoho.com/projects.

Appendix A

Software Project Portfolio Monitoring

A.1 Dashboard Description

Erklärung der Evaluierung
Mit dieser Evaluierung wollen wir herausfinden, wie gut unser Projektsymbol und unsere Portfolio Visualisierungen dazu geeignet sind um sich einen Überblick über die Projekte zu verschaffen und diese auch analysieren kann. Weiters stellen wir auch drei Excel Grafiken zur Verfügung, damit die ProbandInnen einen Eindruck dafür bekommen, wie eine etwaige Visualisierung in Excel aussehen würde.

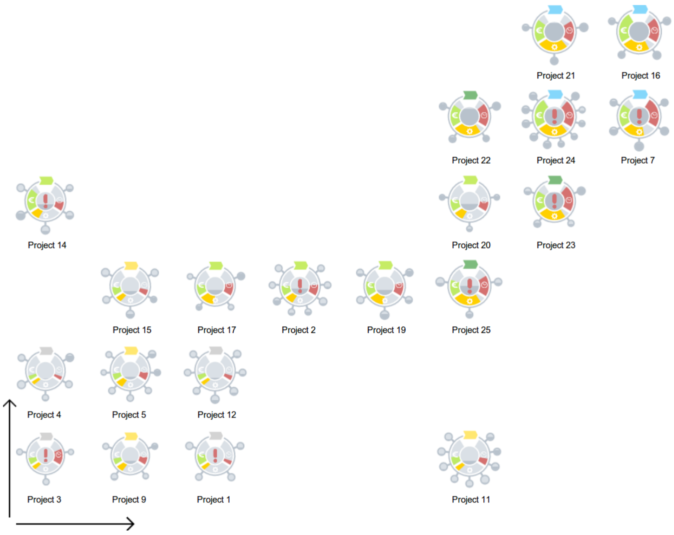
Beschreibung Symbol

Für unsere Visualisierung haben wir das oberhalb abgebildete Symbol entwickelt. Es beinhaltet mehrere Komponenten, die wir in Folge beschreiben:

- Pfeil (hier grün, oben): bezeichnet die Projektphase, in welcher sich das Projekt aktuell befindet. Es gibt 5 verschiedene Projektphase nämlich die Idee (grau), die Entscheidung (gelb), die Planung (hellgrün), die Implementierung (dunkelgrün) und der Abschluss (blau)
- In der Mitte des Symbols befindet sich ein rotes Rufzeichen, welches ein vorhandenes Risiko symbolisiert. Beinhaltet das Projekt kein Risiko, so ist auch kein Rufzeichen vorhanden.
- Der Kreis in der Mitte bezeichnet den Gesamtfortschritt und füllt sich von unten nach oben. Ist das Projekt ganz am Anfang, so ist der gesamte Kreis hellgrau, ist es abgeschlossen, dann ist der gesamte Kreis dunkelgrau.
- Um den inneren Kreis, gibt es drei Teilkreise. Diese zeigen den Fortschritt in finanziellen Mitteln, Ressourcen und Zeit. Der Fortschritt wird in den entsprechenden Farben (finanzielle Mittel: grün, Ressourcen: gelb, Zeit: rot) angezeigt. Das Symbol oben zeigt also, dass finanzielle Mittel und Ressourcen fast aufgebraucht sind, es aber noch verhältnismäßig viel Zeit für die Fertigstellung gibt.
- Die Kreise um den äußersten Kreis symbolisieren die Arbeitspakete. Diese werden im Uhrzeigersinn angeordnet und beginnen rechts oben (Pfeilrichtung). Wie man in der Grafik sehen kann, sind nicht alle Arbeitspakete gleich groß, was deren Anteil am Gesamtvolumen also an finanzielle Mittel, Ressourcen und Zeit darstellt. Die Arbeitspakete beinhalten ebenfalls eine Fortschrittsanzeige, wodurch man diese ebenfalls analysieren kann.

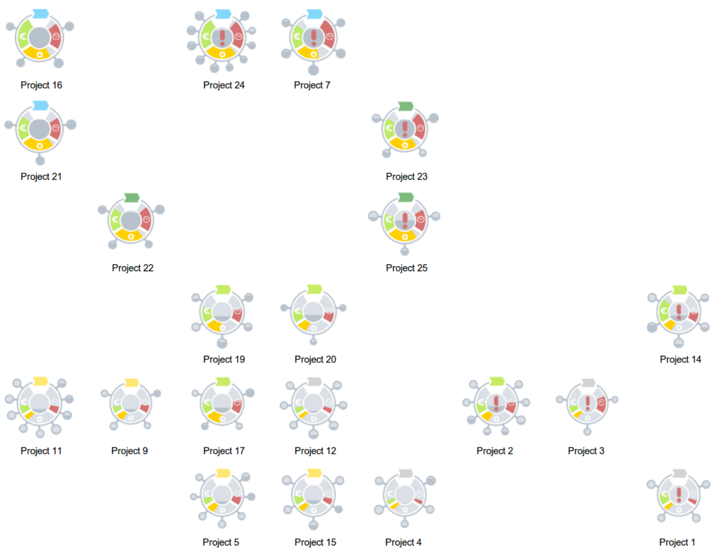
Beschreibung Portfolio Visualisierung
Um nun ein Portfolio an Projekten zu visualisieren, haben wir die Projekte, so wie die Arbeitspakete, nach deren Volumen an Zeit, finanziellen Mittel und Ressourcen skaliert, so dass größere Projekte auch größer dargestellt werden.
Für unsere Portfolio Visualisierung haben wir zwei verschiedene Arten gewählt, nämlich eine Matrixvisualisierung und eine Similarity Visualisierung.

Beschreibung Matrixvisualisierung



Bei der Matrixvisualisierung werden zwei Parameter, jeweils einer für x- und y-Achse verwendet um die Projekte entsprechend dieser beiden Werte positioniert. Für die obere Darstellung verwenden wir Zeit (x-Achse) und Ressourcen (y-Achse). Wobei Die Projekte ganz links am wenigsten Zeit haben und ganz recht am meisten. Das selbe gilt für Projekte ganz unten, welche die geringsten Ressourcen haben wohingegen die Projekte ganz oben die meisten Ressourcen beinhalten.

Beschreibung Similarity Visualisierung



Bei der Similarity Visualisierung verwenden wir Projekteigenschaften um eine Ähnlichkeit zwischen den einzelnen Projekten zu berechnen. Dazu verwenden wir die Projektphase, das Risiko, der Gesamtfortschritt, die finanzielle Mittel, Ressourcen, die Zeit, und deren Fortschritte und die Anzahl der Arbeitspakete. Berechnet man aus diesen Werten die Ähnlichkeiten zwischen Projekten, so erhält man eine Visualisierung wie oben. Daraus sollte ersichtlich sein, warum bestimmte Projekte nahe bei einander liegen, wohingegen andere weiter entfernt sind.

Beschreibung Excel Visualisierung

Die oberste der drei Abbildungen zeigt den Anteil der Arbeitspakete am Gesamtprojekt. Diese entsprechen der Größe der Arbeitspakte in unserer Visualisierung.
Die zweite Abbildung zeigt die Arbeitspakete und deren Fortschritt, entsprechend der Fortschrittsanzeige der Arbeitspakete in unserer Visualisierung.
Abbildung drei zeigt den Gesamtfortschritt, die verbrauchten finanzielle Mitteln, Zeit und Ressourcen entsprechend der Fortschrittsanzeigen des Inneren Teils unseres Projektsymbols.

A.2 Evaluation Questionnaire

ProbandInnen-Nummer:

Allgemeine Angaben

Alter: ______
Geschlecht: □ männlich □ weiblich □ intersex
Beruf: __________________

Erfahrung mit Computern:
□ keine □ wenig □ mittel □ viel □ sehr viel

Erfahrung mit Projektvisualisierungen:
□ keine □ wenig □ mittel □ viel □ sehr viel

Erfahrung mit Excel-Visualisierungen:
□ keine □ wenig □ mittel □ viel □ sehr viel

Fragen

Symbolfragen:

1) Wie gut wird die Semantik von Projekteigenschaften durch das Gesamtsymbol und durch die einzelnen Teilsymbole transportiert?

   1. Gesamtsymbol □ gar nicht □ wenig □ mittel □ gut □ sehr gut
   2. Risiko □ gar nicht □ wenig □ mittel □ gut □ sehr gut
   3. Gesamtfortschritt □ gar nicht □ wenig □ mittel □ gut □ sehr gut
   4. Fortschritt für Zeit, Geld, Ressourcen
      □ gar nicht □ wenig □ mittel □ gut □ sehr gut
   5. Projektphase □ gar nicht □ wenig □ mittel □ gut □ sehr gut
   6. Arbeitspakete (mit Fortschritt) □ gar nicht □ wenig □ mittel □ gut □ sehr gut

2) Ist es von Vor- oder Nachteil, wenn viel Projekteigenschaften auf einen Blick ersichtlich sind? Wie viele Projekteigenschaften kann man deiner Meinung nach in einem Symbol zugleich verstehen und analysieren?

   ____

   ____

   ____

   ____

   ____

ProbandInnen-Nummer:

3) Ist die Informationsmenge die durch ein komplexes Projektsymbol dargestellt wird angemessen?
□ sehr überladen □ etwas überladen □ mittel □ angemessen □ sehr angemessen

4) Würdest du eine andere Ansicht wählen um ein Projekt zu analysieren? Wenn ja, welche?

____

____

____

____

5) Warum glaubst du beinhaltet Projekt 3 (oder 24) ein Risiko?

____

____

Fragen zur Similarity Visualisierung:

1) Wenn du diese Projekt Portfolio Ansicht siehst, was fällt dir auf? In welche Bereiche würdest du die Visualisierung dieser Projekte unterteilen?

____

____

____

____

____

2) Hilft eine Anordnung der Projekte nach ähnlichen Kriterien bei einer Visualisierung für die Analyse von Projekten?
□ gar nicht □ wenig □ mittel □ gut □ sehr gut

3) Wie leicht ist es für dich Gruppen im Portfolio zu identifizieren?
□ gar nicht □ nicht leicht □ mittel □ gut □ sehr gut

ProbandInnen-Nummer:

4) Würdest du eine andere Ansicht (anstelle der Similarity Visualisierung) wählen um ein Projekt zu analysieren? Wenn ja, welche?

____

____

____

____

5) Warum glaubst du sind sich Projekt 23 und 25 ähnlicher als Projekt 22 und 23?

____

____

Fragen zur Matrixvisualisierung (Zeit (x-Achse) vs. Ressourcen (y-Achse) :

1) Wenn du die Portfolio Visualisierung betrachtest, fällt dir grundsätzlich irgendetwas auf? Welche Gruppen kannst du finden? Kannst du Verteilungen identifizieren?

____

____

____

____

____

2) Wie leicht ist es für dich mit der Matrix Darstellung Gruppen/Verteilungen im Portfolio zu identifizieren?
□ gar nicht □ wenig □ mittel □ leicht □ sehr leicht

3) Würdest du eine andere Ansicht (als die Matrixvisualisierung) wählen um ein Projekt zu analysieren? Wenn ja, welche?

____

____

____

ProbandInnen-Nummer:

4) Finde ein Projekt mit vielen Ressourcen aber kurzer Dauer oder umgekehrt, also ein Projekt mit inversen Eigenschaften!

____

Allgemeine Fragen:

1) Wo siehst du die Vor- und Nachteile der Visualisierung mittels Symbole und mittels Charts?

____

____

____

____

2) Welche der beiden Symbolanordnungen (Matrix oder Similarity) ist deiner Meinung nach besser geeignet um Projekt Portfolios darzustellen und warum?

____

____

____

____

____

3) Wofür würdest du, wenn eine Projektsymbolvisualisierung zur Verfügung steht, diese verwenden, wofür nicht und warum?

____

____

____

____

____

ProbandInnen-Nummer:

4) Gibt es weitere Projekteigenschaften, die du in einem Projektsymbol zusätzlich anzeigen würdest? Wenn ja, welche?

____

____

____

____

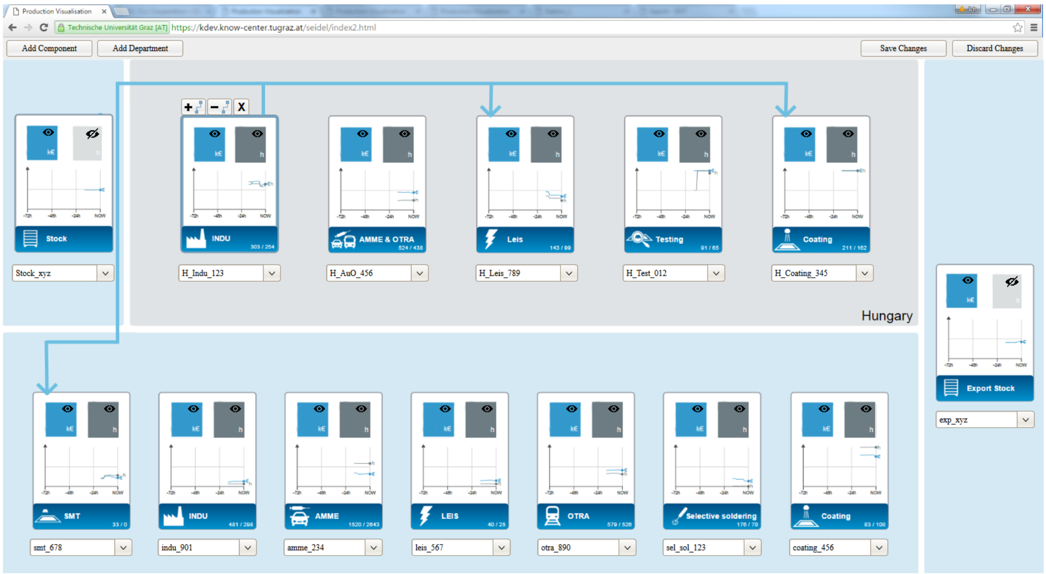
Appendix B

Monitoring Production of Electronic Components

B.1 Mock-ups for a Dashboard Editor

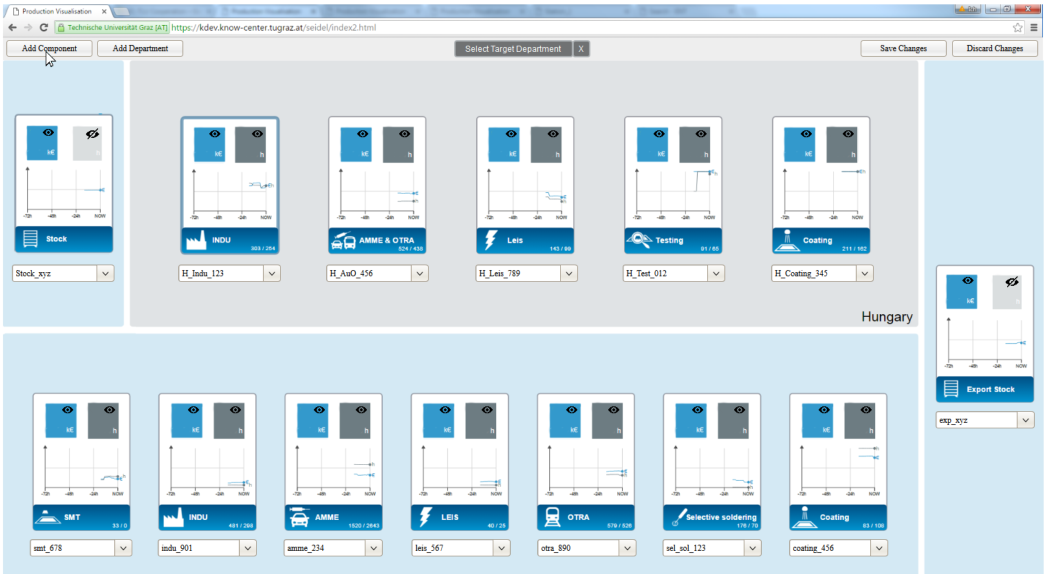
Select Component

Once a component is selected, the following buttons for editing it are displayed: Add connection, Remove connection and Remove component. Also its outgoing connections are displayed (it would be too confusing to show all existing connections between all components at once).

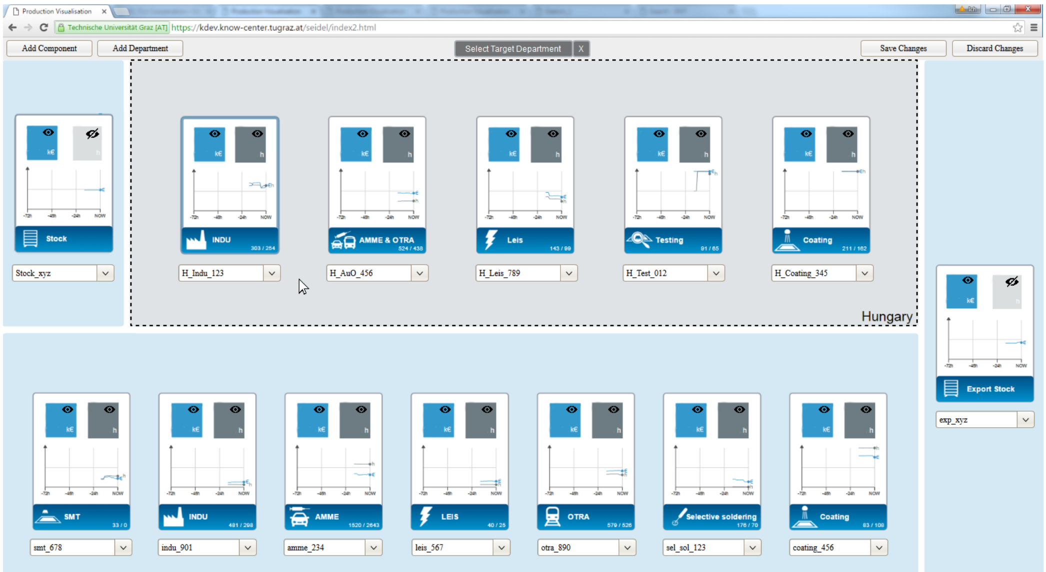


Add Component

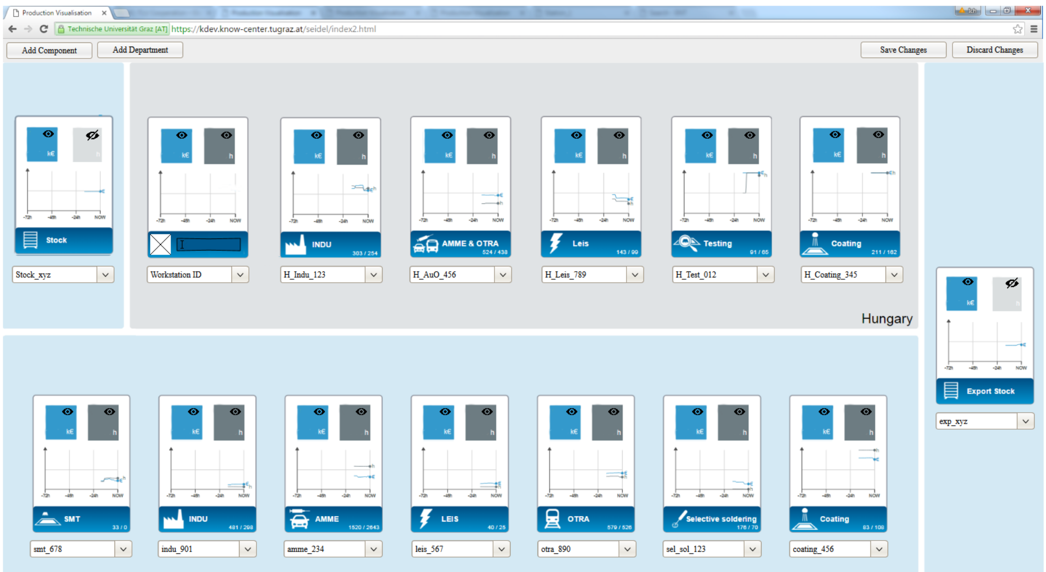
Click on the Add Component-button and a hint appears telling the user to select the department where to place the new component – this process can be canceled by clicking the X-button next to the hint.



While selecting a target department for the new component, departments are highlighted on mouseover.

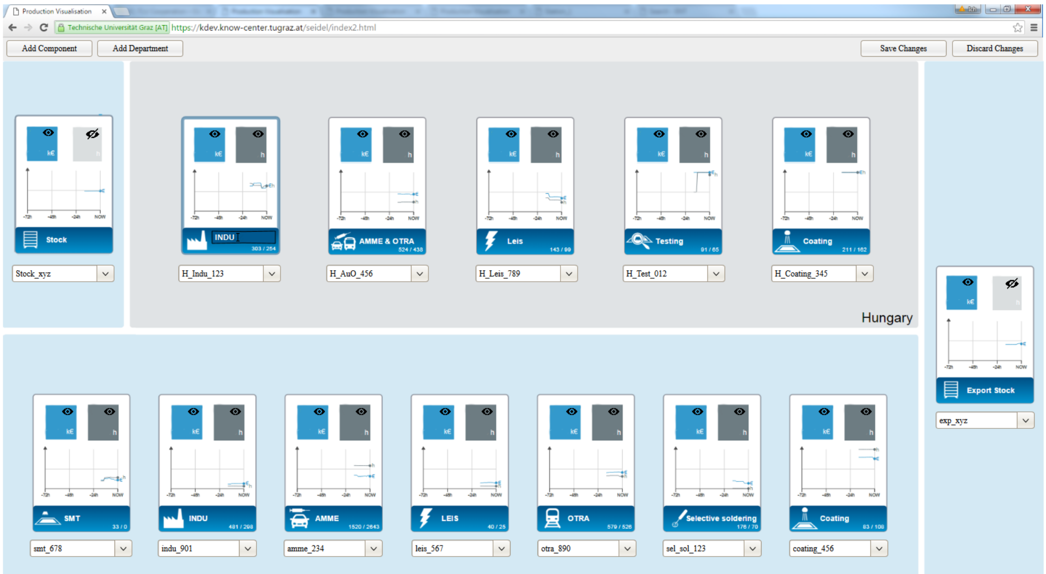


By clicking into a department, the new, empty component is added to this department and its data can be entered. The values a component displays are bound to a data source by selecting a workstation ID from the combo box below the component.



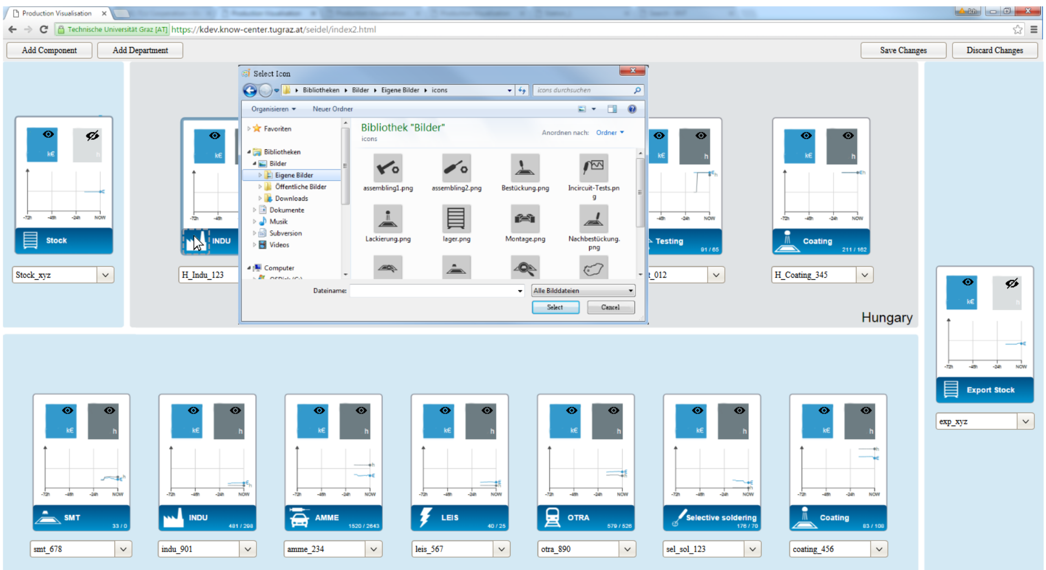
Edit Component Name

A component's name can be edited by simply clicking into the name text – the text input field activates and the new name can be confirmed by pressing Return or clicking somewhere else than the text input field.



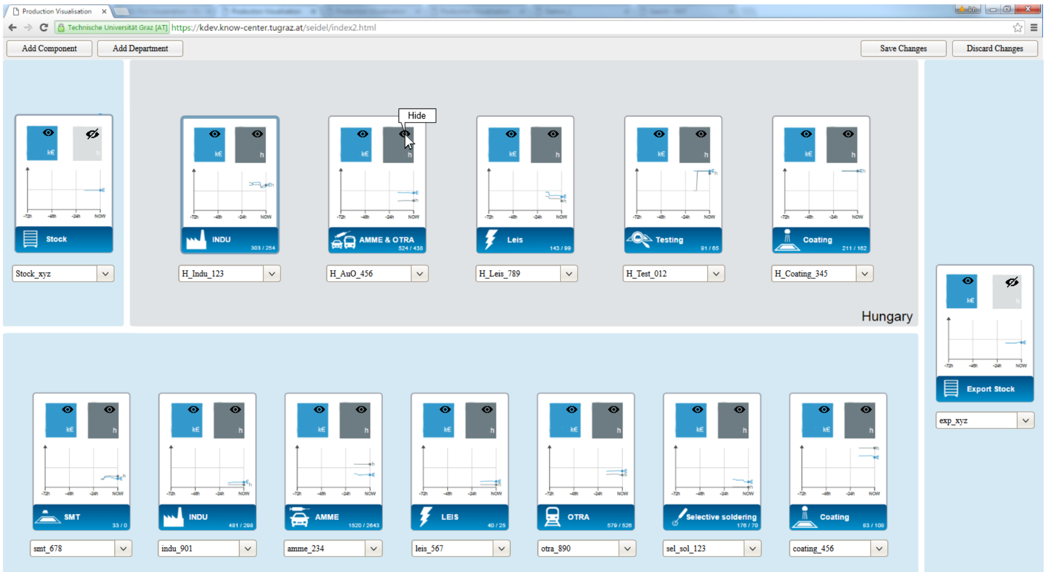
Edit Icon

Clicking onto a component's icon opens a file chooser, where a new icon can be selected.



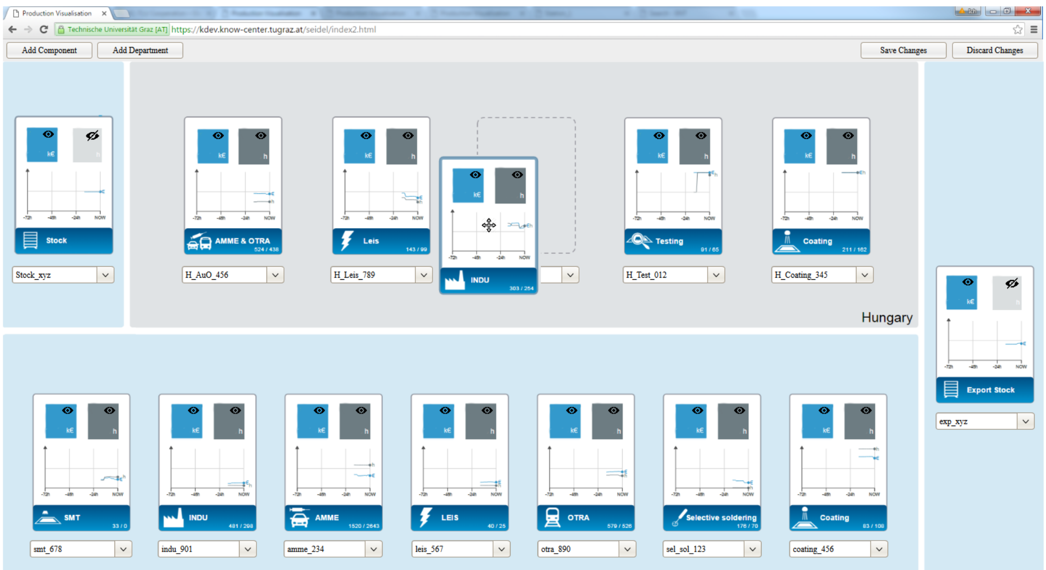
Show/hide Values

The eye in the top right corner of each value box shows whether the value is displayed or hidden (simple eye = displayed, crossed out eye = hidden) – this state can be changed by clicking onto the (crossed out) eye. When hidden, also the colour of the value box fades.



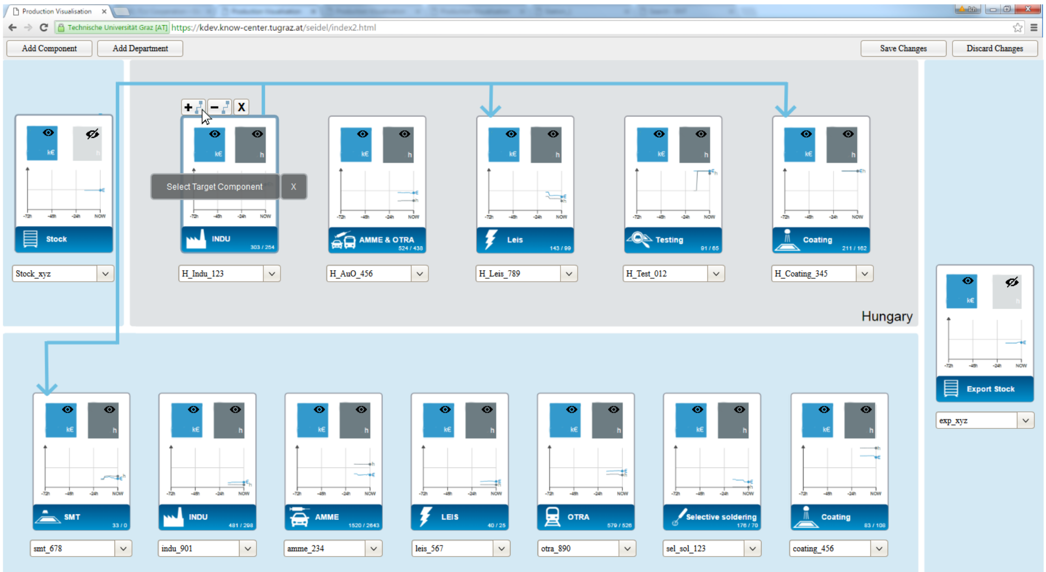
Move Component

Components can be moved by simple drag and drop. On Mouseover possible target positions are highlighted by a dashed frame.

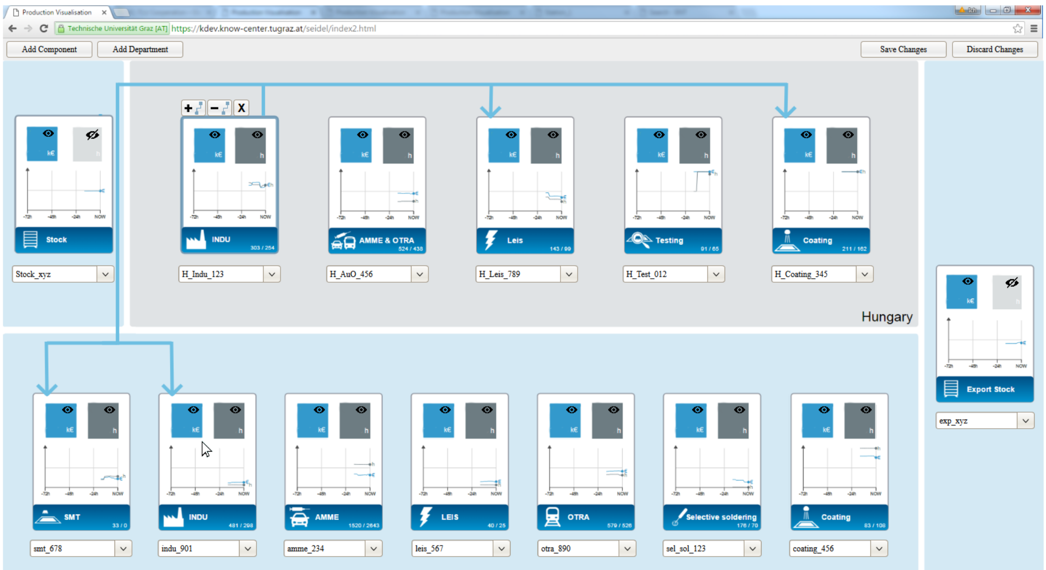


Add Connection

On clicking a component's Add Connection-button a hint is displayed telling the user to select another component as connection target. This process can be cancelled by clicking the X-button next to the hint.

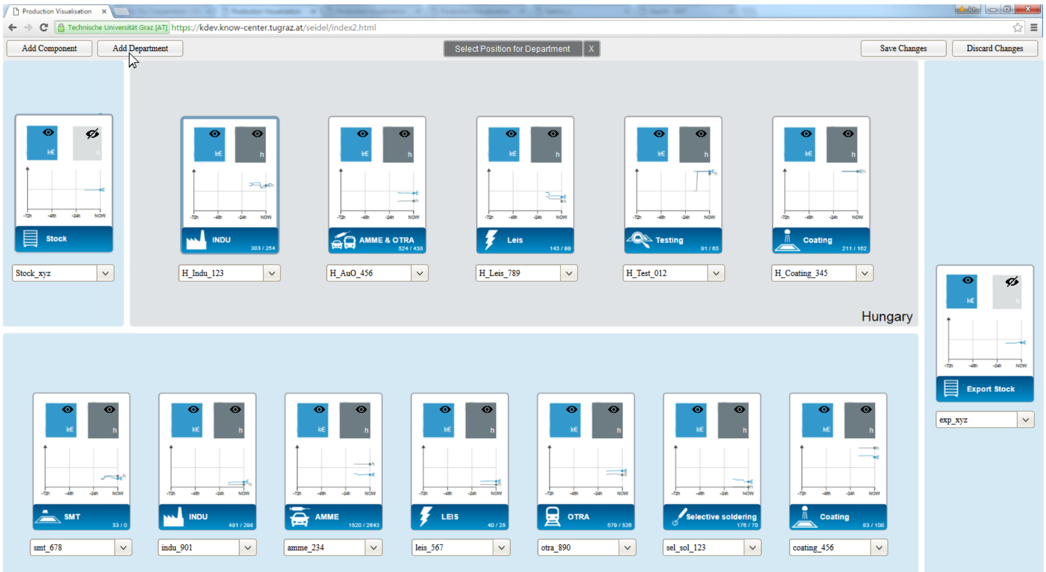


Once the user clicks the target component, source and target component are connected. Removing a connection could follow the same process.

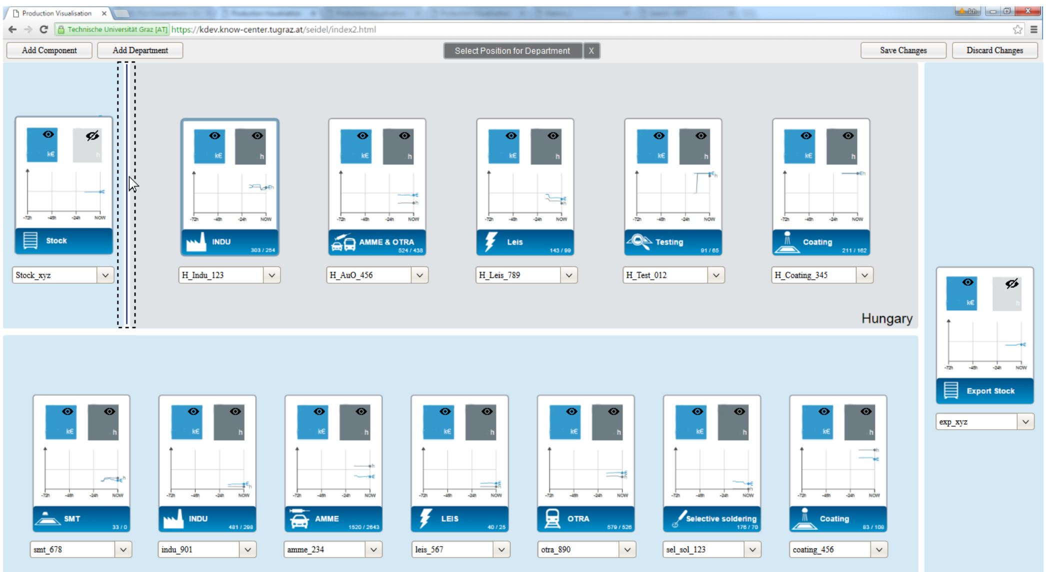


Add Department

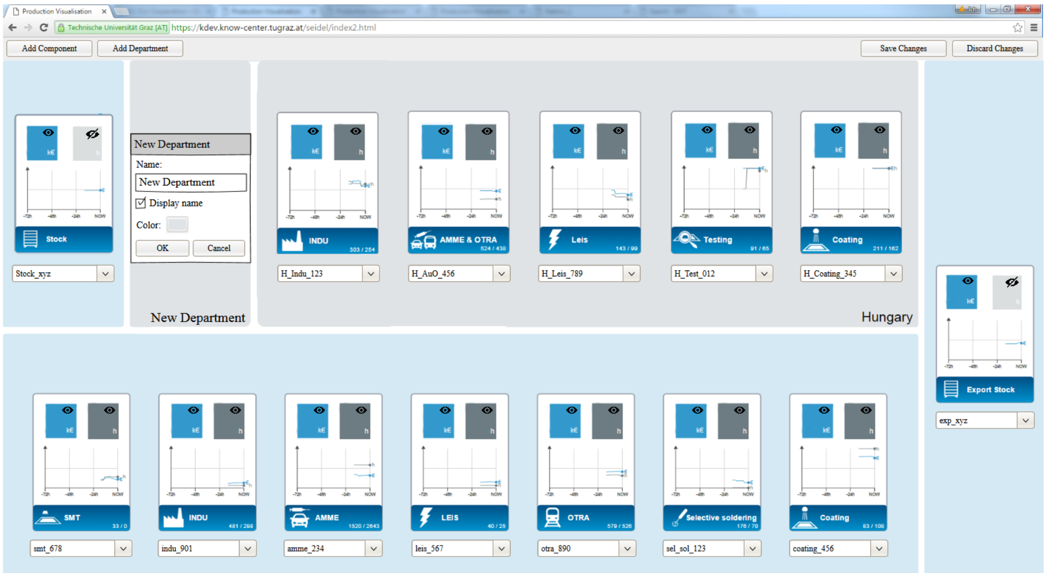
On clicking the Add Department-button a hint is displayed telling the user to select a position on the page for the new department. This process can be cancelled by clicking the X-button next to the hint.



While adding a new department gaps between existing departments are highlighted to show where the new department could be placed.

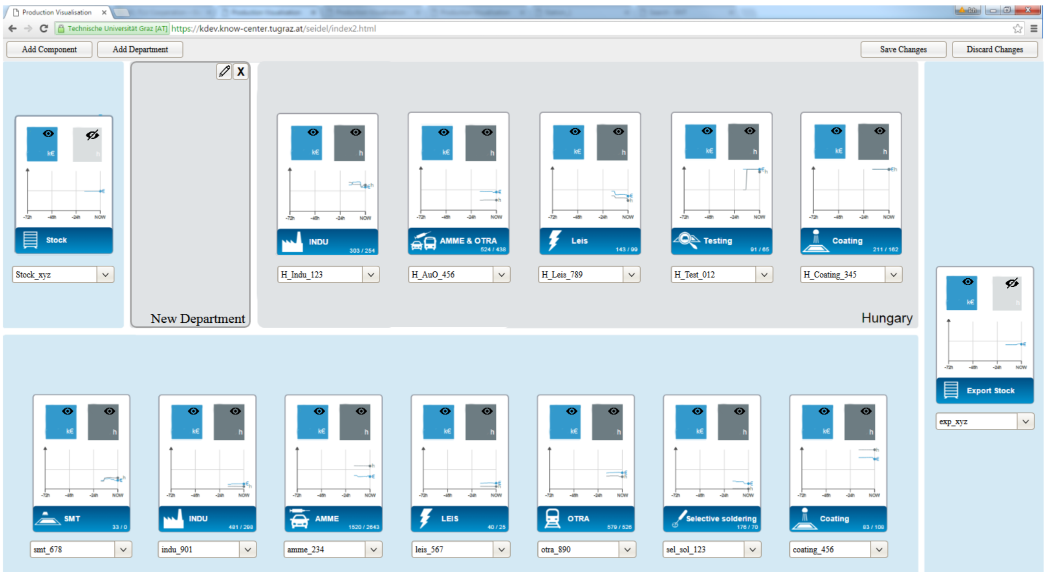


On clicking the target position, a new department is displayed there and its Edit-dialogue is opened automatically.



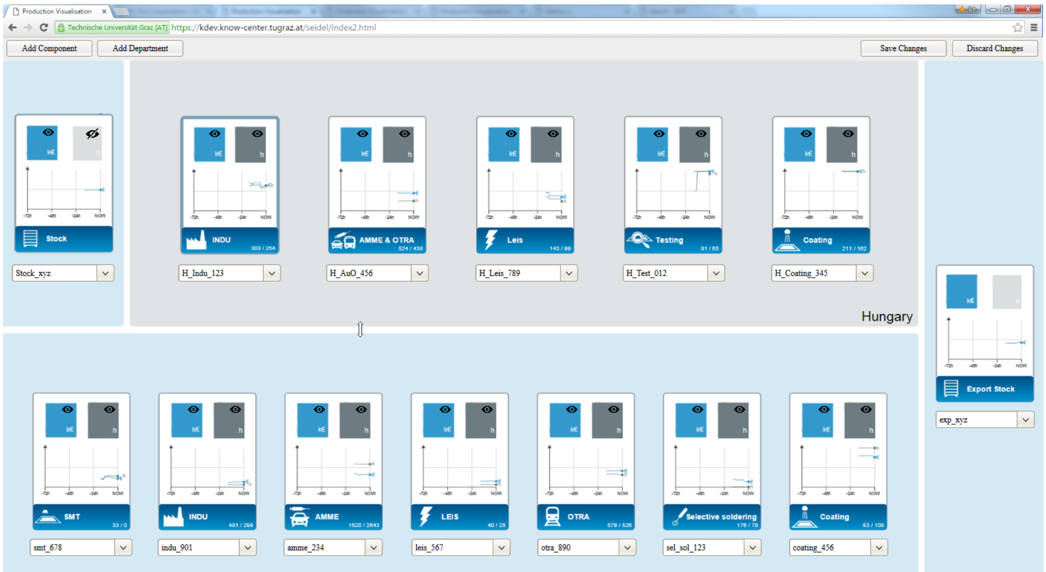
Department Options

When a department is selected by clicking into its background, the following buttons are displayed to edit it: Edit Department (opens the dialogue that also automatically opens after adding a new department) and Remove Department.



Resize Department

On mouseover the cursor indicates that departments can be resized (in dependency of the neighbouring departments) by dragging the splitter that separates departments from each other.



Appendix C

Analysis of Product Assembling and Testing Data

C.1 General Descripton of Visualizer Tool

General Information about the Evaluation

Prerequisites

The Visualizer has been tested on Windows and Linux using Chrome (recommended) and Internet Explorer 11. There are some known issues using Firefox, therefore we ask you not to use it.

In order to be able to conduct a proper evaluation, we kindly ask all use case partners, to ask as many people as possible to participate in the evaluation.

Evaluation Process

This chapter gives an overview about the evaluation process.

In order to be able to evaluate the dashboard, we kindly ask you to participate in the evaluation.

During the evaluation phase, we have 2 steps:

1. Data analysis using the Visualizer
2. Questionnaire to provide feedback (Link will be provided later)

Ad 1:

When opening the Visualizer the first time, a popup appears, which asks you to **share interaction information** with us. All of the shared data is anonymized and does not contain any dataset related information. Meaning, if you change the data type of the second column in the table, for example the following information is sent: Column 2 has changed from integer to string value. On the other hand, generating a visualization sends the following information: a pie chart was generated using the first and second column in the dataset.

The only information we send about your data, while logging your interactions, is:

- the number of columns
- the data type of the columns
- the number of rows

What we do NOT send

- column header names
- data content
- data statistics

Important Information

There are two features in the dashboard, which share more information about your data. Both features are switched off per default and require an explicit activation by the user. When clicking on them, a popup appears informing the user about sharing data. For both of the features we will only ask one use case partner explicitly to give us feedback on these features, the others are free to test them.

The dashboard allows users to share the information with other users on two steps. The first is the sharing of a generated URL more easily, also with mobile devices, the second to collaboratively analyse the data:

- Share Dashboard: When generating a dashboard, the URL constantly updates itself after every user action in order to allow the user to just copy the URL and reapply the current configuration. Since this URLs can become very large and therefore too long for sharing, users can create a short URL. This short URL is stored on the server and will be deleted when not being used for 30 days. Additionally, a QR Code is generated for the short URL to enable sharing with mobile devices more easily. URL shortening is done on the server. If users allow this explicitly, the URL containing the dataset names and data column names are sent to the server. Furthermore, if a brush is active in one of the visualizations, the corresponding subset of the data is contained, therefore we recommend to remove brushes before (click in an empty area of one visualization).
- Share Dashboard + Enable Collaborative Data Analysis: In contrast to the previous one, which only sends the information once and stores it, this feature is constantly sending interactions, and therefore a subset of the data to connected dashboards via the server, unless it is not switched off again. The server forwards data and keeps only the last interaction temporally stored.

A detailed description of both features containing screenshots will be provided later in this document.

In order to be able to distinguish all use cases, all use case partner will receive their own URL. This URL should then be used by all test persons within this company. External partners like Virtual Vehicle or Fraunhofer (if they are related to the use case) are kindly asked to disable logging in order to not contribute to the logging results.

Logging can always be enabled or disabled using the following link:

https://visualizer.know-center.tugraz.at/log_settings

Ad 2:

After testing and using the framework, we will provide a link with an evaluation. This evaluation form should be filled out by every participant within the use case partners company.

General Information about the Dashboard

This chapter gives an overview about the functionality of the dashboard and its parts.

Data Transformation

In order to be able to get all data in the dashboard, we had to do data transformations for some use cases. Those, which are affected, will receive a detailed explanation about this step. All data transformation is done on the client. Users can prove this by disconnecting their network connection after loading the provided URL and perform the data transformation offline.

This step only needs to be done once for a dataset. The transformed data can then be used with the Visualizer.

In order to use the Visualizer, all provided data is already transformed and ready to use.

Dataset Selection

When starting the Visualizer using the following link https://visualizer.know-center.tugraz.at, users can see the start page of the Visualizer to select a dataset.

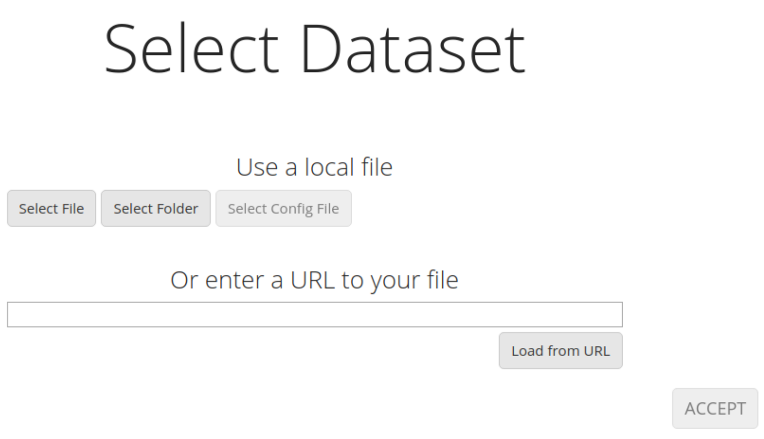
Users can either select a file (or several files), select a folder or select from an existing URL.

Note: When selecting a folder, it says, that the user should "select a folder for uploading". This is a default text from the browsers, which we cannot influence, but in fact, no data is uploaded to the server at all.

Of course, if users have some data online, they can configure an URL with the data. A sample dataset is available here:

https://visualizer.know-center.tugraz.at/sample_data/test/sacramento.csv

Depending on how the data was selected, the button "Accept" or "Load from URL" is activated.



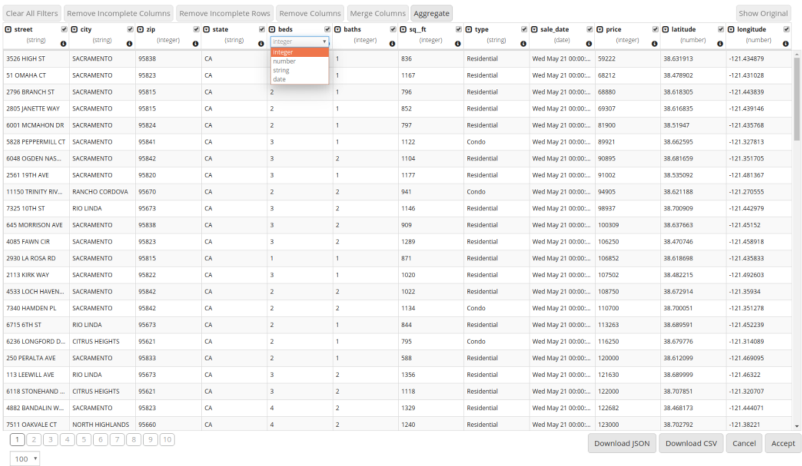
Data Manipulation in Table View

After the dataset has been selected, the next step is the data manipulation. If the data is already in the state the user wants to have it, then the user can simply click on “Accept”, otherwise it is possible to perform various manipulations of the dataset.

In our use cases, we will only use three of these data manipulation features. One is the change of automatically detected data types of columns, the second is the filtering of data within one column and the third is sorting the data.

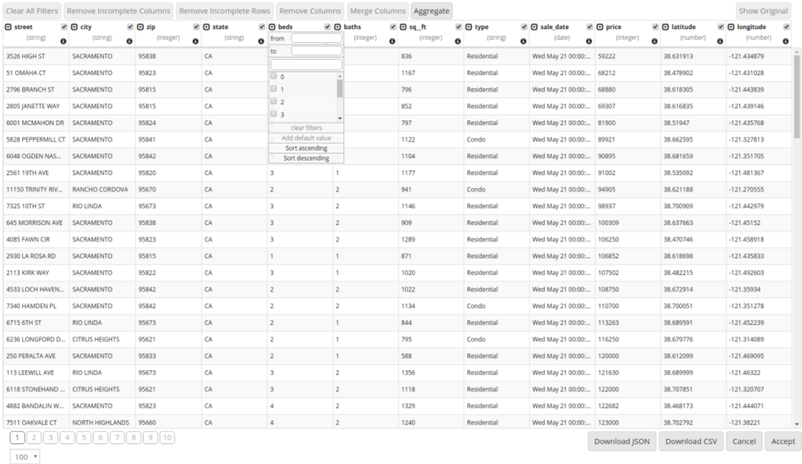
Changing the data type can be easily performed by clicking on the type below each column name. If the type gets detected as a string, the data type cannot be changed due to incompatibilities to other data types.

The screenshot shows the possible data types for one column.



Filtering as well as sorting data in a column is possible by clicking on the arrow left to the column name. The filters are data type dependent meaning a column detected as number additionally has the option to select a value range as well as the actual values.

The screenshot shows possible filtering and sorting methods for a numerical data column.

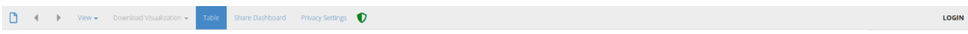


Every change performed on the dataset updates the URL automatically. After reloading the page, the user only needs to select the dataset, all previous data manipulations will be applied automatically.

After data manipulation users can “Accept” their changes and proceed to the dashboard.

Dashboard Generation and Features

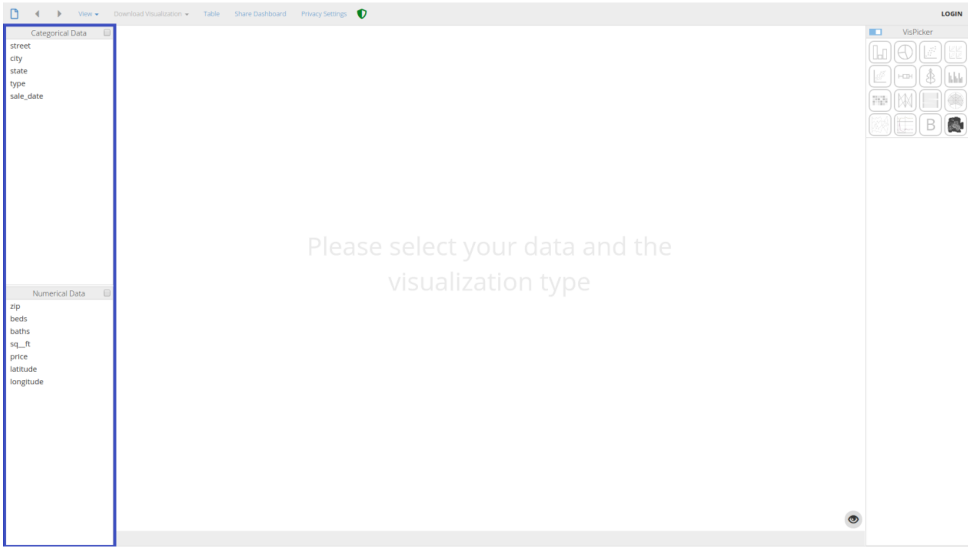
Data can also be manipulated later by clicking on the “Table” in the dashboard header. All previously generated visualizations are not affected by dataset manipulations.



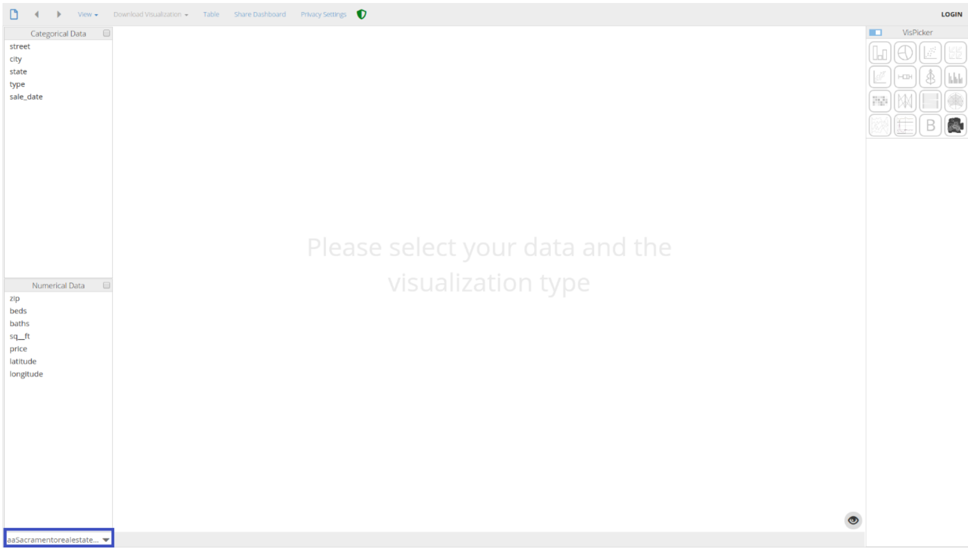
A new dashboard with a different dataset can be created by clicking on the corresponding icon.

Data Selection

On the left side, the users can see two windows containing categorical and numerical values. Categorical values are strings, timestamps and dates, numerical values are integers and numbers. Their assignment depends on the data types defined in the table view.

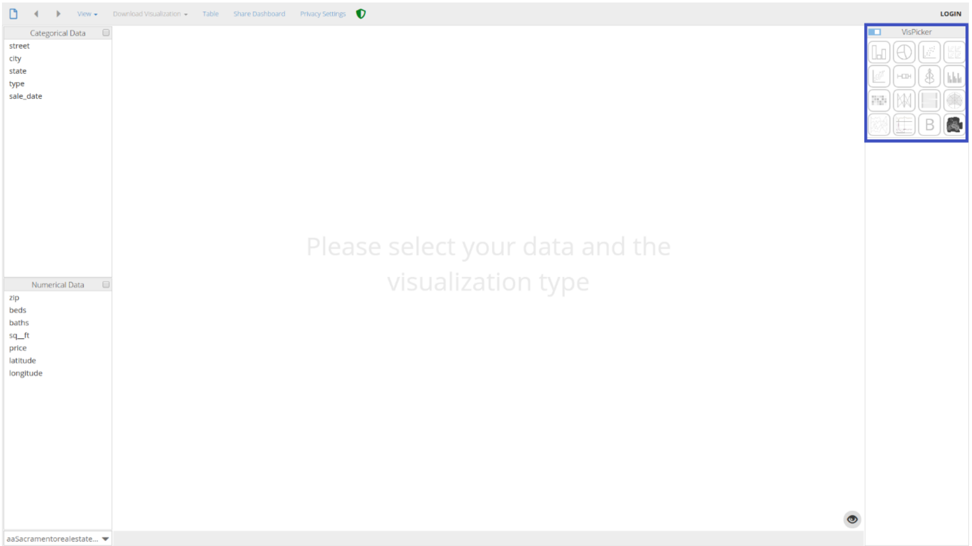


If a user selected multiple datasets or a folder, containing multiple datasets, a selection box is displayed below the data selection area in the left lower part.

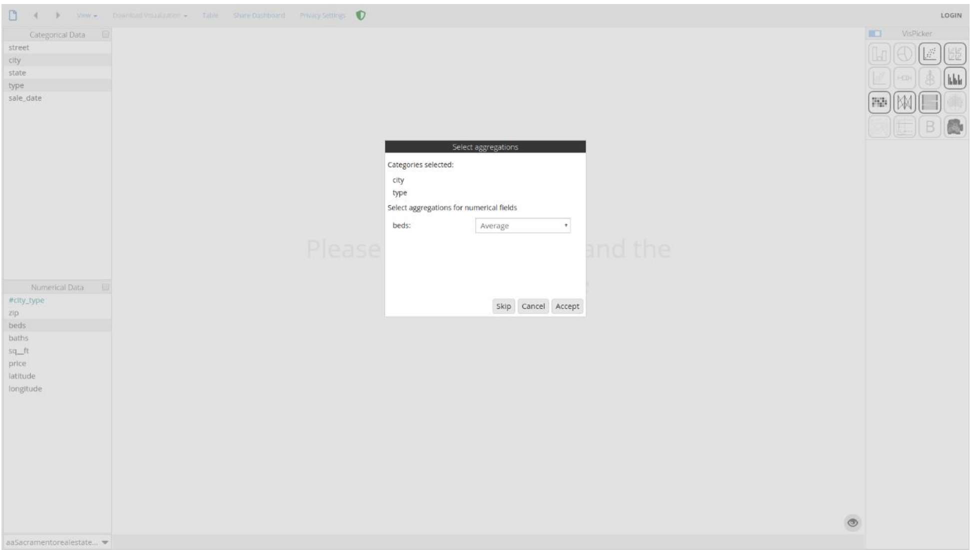


Visualization Selection

On the right side, the user can see the so-called VisPicker.



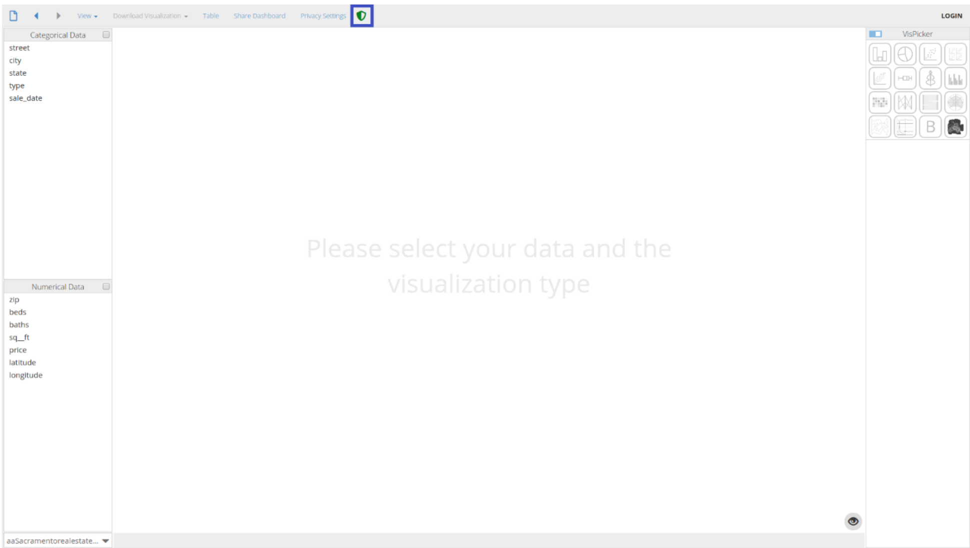
This window allows the selection of visualizations depending on which data is selected on the left side. Tooltips provide information on what kind of data columns need to be selected to generate a visualization. After selecting the fields and clicking on the desired visualization, sometimes (depending on the visualization) a dialog appears asking the user to select how the data should be aggregated, which in our case should be skipped (if available) otherwise accepted using the right aggregation method.



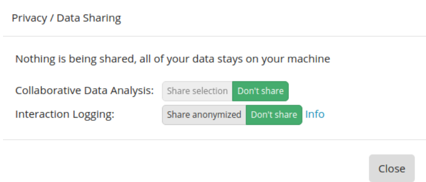
In the upper left area of the VisPicker, there is a switch to enable/disable experimental visualizations. This is required for some use cases and enables more visualizations, which still need improvements.

Data Protection Status

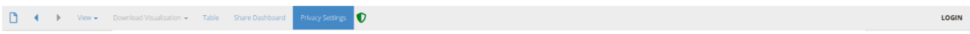
In the dashboard header we can see a shield next to the “Privacy Settings”. This shield indicates the status of data sharing. If the shield is displayed green, nothing is shared with the server. This means, no data leaves the machine.



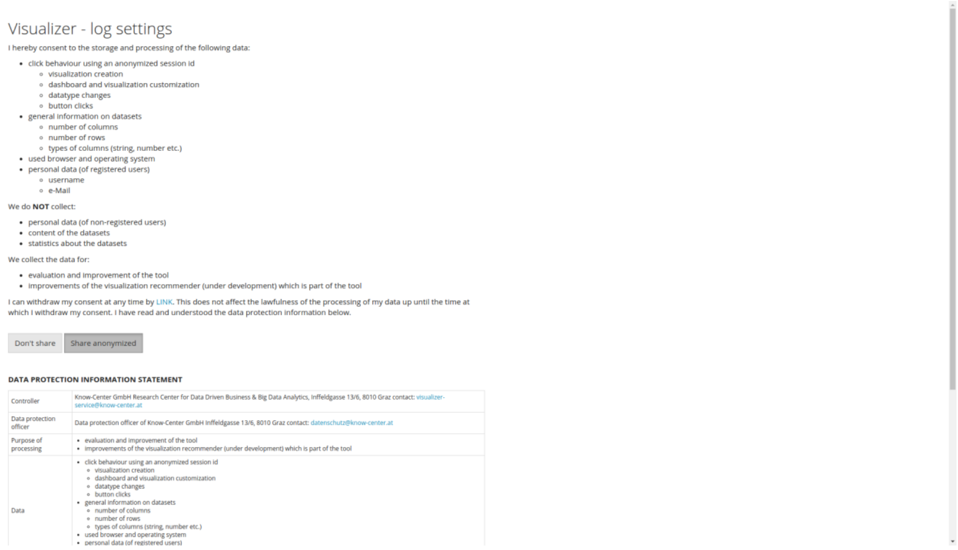
Clicking on the icon, provides detailed information on the status for “Collaborative Data Analysis” and “Interaction Logging”. If one or both of them are activated, users can deactivate them here.



Next to the shield icon there is the link to the “Privacy Settings” where users can activate/deactivate interaction logging. As mentioned above, information is shared anonymously and only as described above.



If a user clicks on “Privacy Settings” a new tab is opened providing all information. If concerns arise, please contact visualizer-service@know-center.at or datenschutz@know-center.at. Logging can always be activated/deactivated using the following link https://visualizer.know-center.tugraz.at/log_settings. **For evaluation purposes, we kindly ask the use case companies to enable this, all others should disable it.**

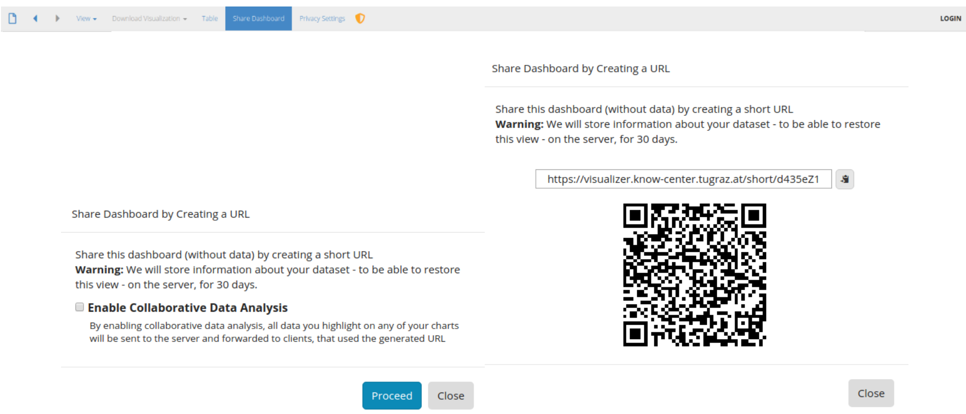


If logging is activated, the shield in the header changes from green to orange.

When sharing the dashboard interaction with other users (Share Dashboard -> “Enable Collaborative Data Analysis”), this shield switches from orange to red, as real data is transmitted via the server to other users. A detailed description of this is provided in the following.

Data Analysis

After configuring a dashboard, the user can start analysing the data using different visualizations. This can be done on one machine, but can also be shared with several other (mobile) devices. On performing a common analysis, users either need to copy the URL from the URL field (which might be very long) or click on “Share Dashboard”, which opens a dialog. After clicking on “Proceed” without further interactions, a short URL is generated as well as a QR Code for mobile devices. Both can be used to access the created dashboard.



URL shortening means, that the whole URL is sent to the server. So be aware of the fact that if performing this step, the dataset name(s), the column names selected for the visualizations, and if a brush is active, a subset of the data, are sent and stored on the server. Since this is a sensitive feature in terms of user data, the users have to confirm their action in a dialog when executing the feature. This feature allows the user to share the link with other users as well as mobile devices more easily. The generated URL combined with the data transmitted to the server will be stored at least for 30 days. After a period of 30 days of not using an URL, the entry will be deleted from the server.

This allows to share the current static status of a dashboard with other users.

Note: When copy/pasting the link to a new browser window or tab, it may happen that the colour mapping is not the same due to deletion of visualizations and the fact that colours are not encoded in the URL.

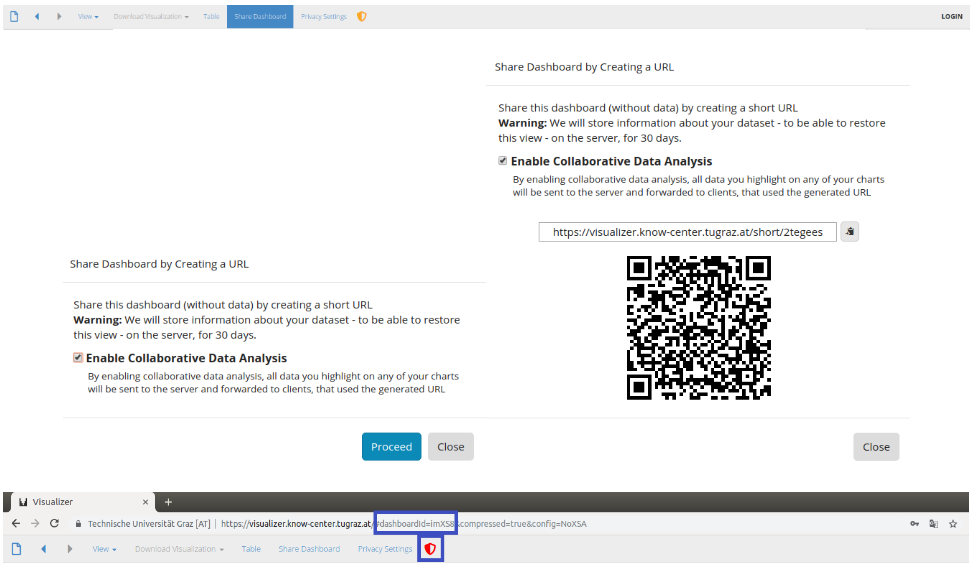
Note: Everything which is behind the hash tag (#) in the URL is in general not sent to the server when applying an URL, only when using the "Share Dashboard" feature this information is sent to the server.

Collaborative Data Analysis

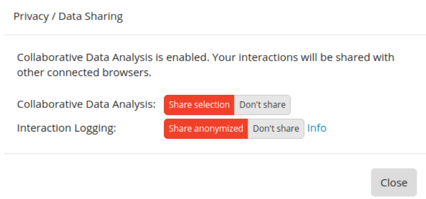
Collaborative Data Analysis allows for a predefined dashboard to share interactions. This is done via websockets, which send the data over the server to connected dashboards.

When clicking on "Share Dashboard" and "Enable Collaborative Data Analysis" the URL gets updated and adds a dashboard id. Since this is a sensitive feature in terms of user data, a warning appears to make the user aware of sharing data. Following that, the user needs to click on "Proceed" which generates a new short URL to enable the user to easily share the dashboard.

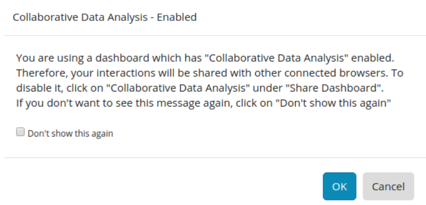
After this, all interactions (and therefor a subset of the data) are sent via the websocket to all remote user/window connected to the same dashboard id. This means, only those users with the same dashboard id in the URL will receive the interactions and therefore the data subset.



In order to raise awareness of the user of sharing interaction data, the shield in the header switches from green/orange to red. The interaction sharing can be easily stopped by clicking on the shield icon and deactivating “Collaborative Data Analysis”. This deletes the dashboard id from the URL and stops sharing the interactions. Re-activating collaborative data analysis will generate a new id and URL, so interactions are not shared any more with those using the previously generated URL.



If the dashboard link is shared with another user, the following message appears as soon as the Visualizer gets loaded, to make the user aware that sharing interactions is enabled. Of course, the user can disable this feature any time in the dashboard by clicking on the shield.



Note: In order to analyse the results mutually, all users need of course access to the same dataset selected in the Visualizer.

Note: Only data selections inside the visualizations are transferred via the websocket, adding/deleting visualizations is not transferred. If new visualizations should be shared, the process of sharing the dashboard with enabled collaborative data analysis has to be performed again and the new link needs to be shared.

C.2 Questionnaire for Analysing Product Assembling and Testing Data

Visualizer Questionnaire

The following questionnaire relates to the evaluation of the Semi40 use cases in work package 3 task 3.3.3.

The descriptions of the use cases have been provided on November, 22nd 2018 via e-mail.

In order to answer the questions, please perform the proposed tasks using the Visualizer first! The task descriptions are available in pdf format and should be provided by the person who is responsible for the use case in your company.

Please fill in the questionnaire until the last page to enable a proper analysis of the collected data! Otherwise results are not stored and cannot be used for evaluation.

If you have any questions or problems in finishing the questionnaire, please contact mrauch@know-center.at.

The following information is provided to abide with the EU General Data Protection Regulation (GDPR):

- The data is collected strictly anonymously, nevertheless personal reference (according to the broad understanding of the GDPR) could possibly be assumed due to limited participation.
- Your data will be used only for the following exactly specified purposes on the basis of our legitimate interests in achieving these purposes:
    - evaluation and improvement of the Visualizer
    - publication of the evaluation results
    - improvements of the visualization recommender (under development) which is part of the tool
- The data will be stored at the Know Center as long as we need them to achieve our purposes and will be accessed only by employees of the Know Center as well as by Fraunhofer Austria Research GmbH (task lead), who are involved in this project.
- The collected data will not be shared or used for any purposes other than stated above.

There are 316 questions in this survey

Test User Information

You will need some basic English skills to complete the evaluation. If you don't understand some words, it's not a problem at all, just ask at any time. *

Please choose **only one** of the following:

- ○ I don't speak English
- ○ Basic
- ○ Okay
- ○ Fluent
- ○ I am a native speaker

How frequent do you use visualisation tools? *

Please choose **only one** of the following:

- ○ Never
- ○ Once a month or less often
- ○ Several times a month
- ○ Several times a week
- ○ Every (work)day

If you use visualization tools: which tools have you been using?

Please choose **all** that apply:

- ☐ Tableau
- ☐ Power BI
- ☐ Spotfire
- ☐ Qlik
- ☐ Other: ______

If you use vizualisation tools: For which tasks have you been using visualization tools?

Please write your answer here:

______

How frequent do you use data analysis tools? *

Please choose **only one** of the following:

- ○ Never
- ○ Once a month or less often
- ○ Several times a month
- ○ Several times a week
- ○ Every (work)day

If you use data analysis tools: which tools have you been using?

Please choose **all** that apply:

- ☐ Power BI
- ☐ OpenRefine
- ☐ Excel
- ☐ Weka
- ☐ KNIME
- ☐ Other: ______

If you use analysis tools: For which tasks have you been using analysis tools?

Please write your answer here:

______

What's your age?

Only numbers may be entered in this field.

Please write your answer here:

Alternatively select one, if you prefer not to enter your real age:

Please choose **only one** of the following:

- 18-27
- 28-37
- 38-47
- 48-57
- 58-67
- 68-77

What is your gender? *

Please choose **only one** of the following:

- Female
- Male

What is your nationality?

*****

Please choose **only one** of the following:

- Austria
- Germany
- France
- Portugal
- Other

What is your profession? *

Please choose **only one** of the following:

- Data Analyst
- Engineer
- Manager
- Production worker
- Researcher
- Student
- Other

What is your educational background? *

Please choose **only one** of the following:

- No formal education
- Primary education
- Secondary education or high school
- Work-related qualifications
- Bachelor's degree
- Master's degree
- Doctorate or higher

What are your primary interests when using a visualization tool?

Please write your answer here:

Have you been using the Visualizer before? *

Please choose **only one** of the following:

- Yes
- No

Questions on tools in your company

For which process steps do you usually use visualization tools?

Please choose **all** that apply:

- ☐ Quality assessment/assurance
- ☐ Investigation of production problems
- ☐ Predictive analytics on production data
- ☐ Configuration of production setups
- ☐ Gaining understanding of new data sets you are unacquainted with
- ☐ Management presentation
- ☐ Other: [ ]

Considering the visualisation possiblities your current system provides: Please provide an importance rating for you or your organisation?

*

Please choose the appropriate response for each item:

	very important				not important at all
Bar Chart	○	○	○	○	○
Stacked Bar Chart	○	○	○	○	○
Grouped Bar Chart	○	○	○	○	○
Bullet Chart	○	○	○	○	○
Bubble Chart	○	○	○	○	○
Heatmap	○	○	○	○	○
Histogram	○	○	○	○	○
Line Chat	○	○	○	○	○
Area Chart	○	○	○	○	○
Stacked Area Chart	○	○	○	○	○
Stream Graph	○	○	○	○	○
Timeline	○	○	○	○	○
Pie Chart	○	○	○	○	○
Sunburst Chart	○	○	○	○	○
Box Plot	○	○	○	○	○
Violin Plot	○	○	○	○	○
Candlestick Chart	○	○	○	○	○
Scatterplot Matrix	○	○	○	○	○
Wafer Map	○	○	○	○	○
Geographic Map	○	○	○	○	○
Radar (Spider) Chart	○	○	○	○	○
Parallel Coordinates	○	○	○	○	○
Tree Map	○	○	○	○	○
Gantt Chart	○	○	○	○	○
Flow Chart	○	○	○	○	○
Gauge Chart (e.g. Speedometer)	○	○	○	○	○
Sankey Diagram	○	○	○	○	○
Node-Link Diagram	○	○	○	○	○

What are the limitations of your current system in terms of data processing and analysis?

Please write your answer here:

What are the limitations of your current system in terms of visual data representation (visualizations)?

Please write your answer here:

What do you like about your current system?

Please write your answer here:

What do you dislike about your current system?

Please write your answer here:

What should be improved on your current system?

Please write your answer here:

Which analysis steps (tasks) would you additionally like to perform in your current system?

Please write your answer here:

Which visualizations (visual data representation) would you additionally like to have in your current system?

Please write your answer here:

What are essential requirements for you to use a tool?

Please write your answer here:

What are current restrictions within your company for using a tool? (E.g. run it locally/company internal with no communication to the outside)

Please write your answer here:

Questions on using the Visualizer

In the following sections, we will ask you questions regarding different aspects of the Visualizer. There will be questions in the following categories:

- **User Expectations**
- **Data Analysis**
- **Dashboard Sharing and Collaborative Data Analysis**
- **Data Security**
- **Instructions for using the Visualizer**
- **General subjective questions**

Please take your time and answer diligently.

User's Expectations

What did you expect when getting the Visualizer introduced?

Please write your answer here:

Did the Visualizer match your expectations?

*

Please choose the appropriate response for each item:

more than expected				not at all
○	○	○	○	○

Why did the Visualizer (not) match your expectations?

Please write your answer here:

Visualizer Improvement Opportunities

What needs further improvements in the Visualizer?

Please write your answer here:

What is missing in the Visualizer?

Please write your answer here:

Could the Visualizer be an improvement/extension to your current system?

*

Please choose **only one** of the following:

- Yes
- Maybe
- No

Why could/why couldn't be the visualizer be an improvement/extension to your current system?

Please write your answer here:

Why could the Visualizer be an improvement/extension to your current system?

Please write your answer here:

Why couldn't the Visualizer be an improvement/extension to your current system?

Please write your answer here:

For which analysis steps would you like to use the Visualizer?

Please choose **all** that apply:

- ☐ Quality assessment/assurance
- ☐ Investigation of production problems
- ☐ Predictive analytics on production data
- ☐ Configuration of production setups
- ☐ Gaining understanding of new data sets you are unacquainted with
- ☐ Management presentation
- ☐ Other:

For which analysis steps would you not use the Visualizer?

Please choose **all** that apply:

- ☐ Quality assessment/assurance
- ☐ Investigation of production problems
- ☐ Predictive analytics on production data
- ☐ Configuration of production setups
- ☐ Gaining understanding of new data sets you are unacquainted with
- ☐ Management presentation
- ☐ Other: ____________

Data Analysis

Did the Visualizer help you in gaining new insights in your data?

*****

Please choose the appropriate response for each item:

yes, discovered meaningful insights				no new insights at all
○	○	○	○	○

Which functionalities did you try in the Visualizer?

*****

Please choose the appropriate response for each item:

	used very often				didn't use at all
Data manipulation in table view: datatype change	○	○	○	○	○
Data manipulation in table view: filtering	○	○	○	○	○
Data manipulation in table view: sorting	○	○	○	○	○
Data manipulation in table view: merging columns	○	○	○	○	○
Data manipulation in table view: removing columns	○	○	○	○	○
Data manipulation in table view: data corrections	○	○	○	○	○
Dashboard creation: Creating multiple visualizations	○	○	○	○	○
Dashboard creation: Manually re-positioning visualization windows	○	○	○	○	○
Dashboard creation: Automatic window positioning	○	○	○	○	○
Interactions in visualizations: Zooming (magnifying)	○	○	○	○	○
Interactions in visualizations: Panning (moving)	○	○	○	○	○
Interactions in visualizations: Changing the visualization title	○	○	○	○	○
Interactions in visualizations: Data filtering	○	○	○	○	○
Interactions in visualizations: Data aggregation	○	○	○	○	○
Interactions in visualizations: Sorting	○	○	○	○	○
Interactions in visualizations: Re-configuring data-to-visualization mapping	○	○	○	○	○
Interactions in visualizations: Brushing (data selection)	○	○	○	○	○
Interactions in visualizations: Coordinated brushing over multiple visualizations	○	○	○	○	○

Which OTHER functionalities did you try in the Visualizer? Please rate them from 1 (used very often) to 5 (didn't use at all)

(for example: filtering 4 --> filtering not used often)

Please write your answer here:

Which functionalities are useful for you?

*

Please choose the appropriate response for each item:

	very useful				not useful
Data manipulation in table view: datatype change	○	○	○	○	○
Data manipulation in table view: filtering	○	○	○	○	○
Data manipulation in table view: sorting	○	○	○	○	○
Data manipulation in table view: merging columns	○	○	○	○	○
Data manipulation in table view: removing columns	○	○	○	○	○
Data manipulation in table view: data corrections	○	○	○	○	○
Dashboard creation: Creating multiple visualizations	○	○	○	○	○
Dashboard creation: Manually re-positioning visualization windows	○	○	○	○	○
Dashboard creation: Automatic window positioning	○	○	○	○	○
Interactions in visualizations: Zooming (magnifying)	○	○	○	○	○
Interactions in visualizations: Panning (moving)	○	○	○	○	○
Interactions in visualizations: Changing the visualization title	○	○	○	○	○
Interactions in visualizations: Data filtering	○	○	○	○	○
Interactions in visualizations: Data aggregation	○	○	○	○	○
Interactions in visualizations: Sorting	○	○	○	○	○
Interactions in visualizations: Re-configuring data-to-visualization mapping	○	○	○	○	○
Interactions in visualizations: Brushing (data selection)	○	○	○	○	○
Interactions in visualizations: Coordinated brushing over multiple visualizations	○	○	○	○	○

Which OTHER functionalities are useful for you? Please rate them from 1 (very useful) to 5 (not useful)

(for example: filtering 4 --> filtering is not useful)

Please write your answer here:

Which functionalities could be removed as they are useless?

*

Please choose the appropriate response for each item:

	must be available				can be removed
Data manipulation in table view: datatype change	◯	◯	◯	◯	◯
Data manipulation in table view: filtering	◯	◯	◯	◯	◯
Data manipulation in table view: sorting	◯	◯	◯	◯	◯
Data manipulation in table view: merging columns	◯	◯	◯	◯	◯
Data manipulation in table view: removing columns	◯	◯	◯	◯	◯
Data manipulation in table view: data corrections	◯	◯	◯	◯	◯
Dashboard creation: Creating multiple visualizations	◯	◯	◯	◯	◯
Dashboard creation: Manually re-positioning visualization windows	◯	◯	◯	◯	◯
Dashboard creation: Automatic window positioning	◯	◯	◯	◯	◯
Interactions in visualizations: Zooming (magnifying)	◯	◯	◯	◯	◯
Interactions in visualizations: Panning (moving)	◯	◯	◯	◯	◯
Interactions in visualizations: Changing the visualization title	◯	◯	◯	◯	◯
Interactions in visualizations: Data filtering	◯	◯	◯	◯	◯
Interactions in visualizations: Data aggregation	◯	◯	◯	◯	◯
Interactions in visualizations: Sorting	◯	◯	◯	◯	◯
Interactions in visualizations: Re-configuring data-to-visualization mapping	◯	◯	◯	◯	◯
Interactions in visualizations: Brushing (data selection)	◯	◯	◯	◯	◯
Interactions in visualizations: Coordinated brushing over multiple visualizations	◯	◯	◯	◯	◯

Which OTHER functionalities could be removed as they are useless? Please rate them from 1 (very useful) to 5 (not useful)

(for example: filtering 4 --> filtering is not useful)

Please write your answer here:

Which functionalities should be extended or modified? Please specify how and why?

Comment only when you choose an answer.

Please choose all that apply and provide a comment:

- ☐ Data manipulation in table view: datatype change
- ☐ Data manipulation in table view: filtering
- ☐ Data manipulation in table view: sorting

☐ Data manipulation in table view: merging columns

☐ Data manipulation in table view: removing columns

☐ Data manipulation in table view: data corrections

☐ Dashboard creation: Creating multiple visualizations

☐ Dashboard creation: Manually re-positioning visualization windows

☐ Dashboard creation: Automatic window positioning

☐ Interactions in visualizations: Zooming (magnifying)

☐ Interactions in visualizations: Panning (moving)

☐ Interactions in visualizations: Changing the visualization title

☐ Interactions in visualizations: Data filtering

☐ Interactions in visualizations: Data aggregation

☐ Interactions in visualizations: Sorting

☐ Interactions in visualizations: Re-configuring data-to-visualization mapping

☐ Interactions in visualizations: Brushing (data selection)

☐ Interactions in visualizations: Coordinated brushing over multiple visualizations

Other functionalities that should be extended or modified:

Please write your answer here:

Which functionalities did you especially like in the Visualizer? Please specify why?

Comment only when you choose an answer.

Please choose all that apply and provide a comment:

☐ Data manipulation in table view: datatype change

☐ Data manipulation in table view: filtering

☐ Data manipulation in table view: sorting

☐ Data manipulation in table view: merging columns

☐ Data manipulation in table view: removing columns

☐ Data manipulation in table view: data corrections

☐ Dashboard creation: Creating multiple visualizations

☐ Dashboard creation: Manually re-positioning visualization windows

☐ Dashboard creation: Automatic window positioning

☐ Interactions in visualizations: Zooming (magnifying)

☐ Interactions in visualizations: Panning (moving)

☐ Interactions in visualizations: Changing the visualization title

☐ Interactions in visualizations: Data filtering

☐ Interactions in visualizations: Data aggregation

☐ Interactions in visualizations: Sorting

☐ Interactions in visualizations: Re-configuring data-to-visualization mapping

☐ Interactions in visualizations: Brushing (data selection)

☐ Interactions in visualizations: Coordinated brushing over multiple visualizations

Other functionalities you especially liked in the Visualizer:

Please write your answer here:

Which functionalities did you not like at all? Please specify why?

Comment only when you choose an answer.

Please choose all that apply and provide a comment:

☐ Data manipulation in table view: datatype change

☐ Data manipulation in table view: filtering

☐ Data manipulation in table view: sorting

☐ Data manipulation in table view: merging columns

☐ Data manipulation in table view: removing columns

☐ Data manipulation in table view: data corrections

☐ Dashboard creation: Creating multiple visualizations

☐ Dashboard creation: Manually re-positioning visualization windows

☐ Dashboard creation: Automatic window positioning

☐ Interactions in visualizations: Zooming (magnifying)

☐ Interactions in visualizations: Panning (moving)

☐ Interactions in visualizations: Changing the visualization title

☐ Interactions in visualizations: Data filtering

☐ Interactions in visualizations: Data aggregation

☐ Interactions in visualizations: Sorting

☐ Interactions in visualizations: Re-configuring data-to-visualization mapping

☐ Interactions in visualizations: Brushing (data selection)

☐ Interactions in visualizations: Coordinated brushing over multiple visualizations

Other functionalities you did not like:

Please write your answer here:

Which visualizations did you use in analysing your data?

Please choose **all** that apply:

☐ Bar Chart
☐ Pie Chart
☐ Scatterplot
☐ Scatterplot Matrix
☐ Bubble Chart
☐ Box Plot
☐ Violin Plot
☐ Grouped Bar Chart
☐ Heatmap
☐ Parallel Coordinates

- ☐ Sankey Diagram
- ☐ Radar Chart
- ☐ Stream Graph
- ☐ Line Chart
- ☐ Bullet Chart
- ☐ Landscape

Regarding Bar Chart: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Bar Chart: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Bar Chart: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Bar Chart: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Bar Chart: Is the visualization new to you? *

Please choose **only one** of the following:

- ○ Yes
- ○ No

Regarding Bar Chart: Do you have any comments?

Please write your answer here:

Regarding Pie Chart: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Pie Chart: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Pie Chart: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Pie Chart: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Pie Chart: Is the visualization new to you? *

Please choose **only one** of the following:

- ○ Yes
- ○ No

Regarding Pie Chart: Do you have any comments?

Please write your answer here:

Regarding Scatterplot: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Scatterplot: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Scatterplot: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Scatterplot: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Scatterplot: Is the visualization new to you? *

Please choose **only one** of the following:

○ Yes

○ No

Regarding Scatterplot: Do you have any comments?

Please write your answer here:

Regarding Scatterplot Matrix: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Scatterplot Matrix: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Scatterplot Matrix: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Scatterplot Matrix: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Scatterplot Matrix: Is the visualization new to you? *

Please choose **only one** of the following:

- ○ Yes
- ○ No

Regarding Scatterplot Matrix: Do you have any comments?

Please write your answer here:

Regarding Bubble Chart: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Bubble Chart: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Bubble Chart: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Bubble Chart: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Bubble Chart: Is the visualization new to you? *

Please choose **only one** of the following:

○ Yes

○ No

Regarding Bubble Chart: Do you have any comments?

Please write your answer here:

<u>Regarding Box Plot</u>: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

<u>Regarding Box Plot</u>: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

<u>Regarding Box Plot</u>: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

<u>Regarding Box Plot</u>: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

<u>Regarding Box Plot</u>: Is the visualization new to you? *

Please choose **only one** of the following:

○ Yes

○ No

<u>Regarding Box Plot</u>: Do you have any comments?

Please write your answer here:

Regarding Violin Plot: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Violin Plot: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Violin Plot: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Violin Plot: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Violin Plot: Is the visualization new to you? *

Please choose **only one** of the following:

○ Yes

○ No

Regarding Violin Plot: Do you have any comments?

Please write your answer here:

Regarding Crouped Bar Chart: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Grouped Bar Chart: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Grouped Bar Chart: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Grouped Bar Chart: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Grouped Bar Chart: Is the visualization new to you? *

Please choose **only one** of the following:

- ○ Yes
- ○ No

Regarding Grouped Bar Chart: Do you have any comments?

Please write your answer here:

<u>Regarding Heatmap</u>: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
◯	◯	◯	◯	◯

<u>Regarding Heatmap</u>: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
◯	◯	◯	◯	◯

<u>Regarding Heatmap</u>: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
◯	◯	◯	◯	◯

<u>Regarding Heatmap</u>: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
◯	◯	◯	◯	◯

<u>Regarding Heatmap</u>: Is the visualization new to you? *

Please choose **only one** of the following:

◯ Yes

◯ No

<u>Regarding Heatmap</u>: Do you have any comments?

Please write your answer here:

Regarding Parallel Coordinates: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Parallel Coordinates: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Parallel Coordinates: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Parallel Coordinates: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Parallel Coordinates: Is the visualization new to you? *

Please choose **only one** of the following:

○ Yes

○ No

Regarding Parallel Coordinates: Do you have any comments?

Please write your answer here:

Regarding Sankey Diagramm: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Sankey Diagramm: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Sankey Diagramm: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Sankey Diagramm: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Sankey Diagramm: Is the visualization new to you? *

Please choose **only one** of the following:

○ Yes

○ No

Regarding Sankey Diagramm: Do you have any comments?

Please write your answer here:

Regarding Radar Chart: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Radar Chart: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Radar Chart: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Radar Chart: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Radar Chart: Is the visualization new to you? *

Please choose **only one** of the following:

- ○ Yes
- ○ No

Regarding Radar Chart: Do you have any comments?

Please write your answer here:

<u>Regarding Stream Graph</u>: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

<u>Regarding Stream Graph</u>: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

<u>Regarding Stream Graph</u>: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

<u>Regarding Stream Graph</u>: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

<u>Regarding Stream Graph</u>: Is the visualization new to you? *

Please choose **only one** of the following:

○ Yes

○ No

<u>Regarding Stream Graph</u>: Do you have any comments?

Please write your answer here:

Regarding Line Chart: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Line Chart: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Line Chart: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Line Chart: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Line Chart: Is the visualization new to you? *

Please choose **only one** of the following:

○ Yes

○ No

Regarding Line Chart: Do you have any comments?

Please write your answer here:

<u>Regarding Bullet Chart</u>: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

<u>Regarding Bullet Chart</u>: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

<u>Regarding Bullet Chart</u>: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

<u>Regarding Bullet Chart</u>: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

<u>Regarding Bullet Chart</u>: Is the visualization new to you? *

Please choose **only one** of the following:

○ Yes

○ No

<u>Regarding Bullet Chart</u>: Do you have any comments?

Please write your answer here:

Regarding Landscape: Was the visualization easy to understand? *

Please choose the appropriate response for each item:

yes, very easy to understand				no, very difficult to understand
○	○	○	○	○

Regarding Landscape: Were the interactions intuitive? *

Please choose the appropriate response for each item:

yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Regarding Landscape: Did the visualization help in analysing your data? *

Please choose the appropriate response for each item:

yes, gained all insights				no, not at all
○	○	○	○	○

Regarding Landscape: Does this visualization provide all necessary features? *

Please choose the appropriate response for each item:

yes, it's all there				no, there's a lot missing
○	○	○	○	○

Regarding Landscape: Is the visualization new to you? *

Please choose **only one** of the following:

○ Yes

○ No

Regarding Landscape: Do you have any comments?

Please write your answer here:

Which visualizations are missing in the Visualizer for you to analyse your data?

Please write your answer here:

How mentally demanding was it for you to analyse the data? *

Please choose the appropriate response for each item:

very low ○ ○ ○ ○ ○ very high

How physically demanding was it to analyse the data? *

Please choose the appropriate response for each item:

very low ○ ○ ○ ○ ○ very high

How successful in terms of analysing your data were you? *

Please choose the appropriate response for each item:

very successful ○ ○ ○ ○ ○ not successful

Have you been frustrated when trying to analyse your data? *

Please choose the appropriate response for each item:

not frustrated ○ ○ ○ ○ ○ very frustrated

How insecure, discouraged, irritated, stressed, or annoyed were you when analysing your data? *

Please choose the appropriate response for each item:

very low ○ ○ ○ ○ ○ very high

How difficult was it for you to use the Visualizer analysing your data? *

Please choose the appropriate response for each item:

not difficult at all				very difficult
○	○	○	○	○

Further comments on the data analysis using the Visualizer:

Please write your answer here:

Dashboard Sharing and Collaborative Data Analysis

Did you try the “Share Dashboard” functionality?

*****

Please choose **only one** of the following:

- ◯ Yes
- ◯ No

Did you share a dashboard by

Please choose **all** that apply:

- ☐ sending a link
- ☐ sharing a QR code

Did it support you in performing your tasks?

Please write your answer here:

Why did you not use the "Share Dashboard" functionality? Any concerns?

Please write your answer here:

Did you try the feature “Collaborative Data Analysis”?

*

Please choose **only one** of the following:

- Yes
- No

Did the feature “Collaborative Data Analysis” help you in analysing your data?

*

Please choose **only one** of the following:

- Yes
- Maybe
- No

Did the feature "Collaborative Data Analysis" perform as expected? Why/Why not?

Please write your answer here:

Why didn't you use the feature “Collaborative Data Analysis”? Any concerns?

Please write your answer here:

Do you think “Collaborative Data Analysis” is intuitive to enable?

Please choose the appropriate response for each item:

Yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Do you think "Collaborative Data Analysis" is intuitive to disable?

Please choose the appropriate response for each item:

Yes, very intuitive				no, not intuitive at all
○	○	○	○	○

Do you think "Collaborative Data Analysis" is easy to use?

Please choose the appropriate response for each item:

Yes, very easy				no, very difficult
○	○	○	○	○

What additional features would be useful for Collaborative Data Analysis and Sharing?

Please choose **all** that apply:

- ☐ Annotating findings in the visualizations
- ☐ Collaborative report/summary writing
- ☐ Built-in chat
- ☐ Publishing visualizations and dashboards (incl. annotations)
- ☐ Publishing reports
- ☐ Commenting on published visualizations, dashboards and reports
- ☐ Other: [ ]

Do you already have a tool which enables you to perform visual data analysis collaboratively?

*****

Please choose **only one** of the following:

- ○ Yes
- ○ No

Do you often use this tool? *

Please choose **only one** of the following:

- ○ Never
- ○ Once a month or less often
- ○ Several times a month
- ○ Several times a week
- ○ Every (work)day

Would you like to have such a tool?

*****

Please choose **only one** of the following:

- Yes
- Maybe
- No

Do you see any benefits in having a tool to share dashboards with colleagues?

*

Please choose **only one** of the following:

- Yes
- Maybe
- No

Which benefits can you identify?

Please write your answer here:

Which benefits/concerns can you identify?

Please write your answer here:

What are your concerns?

Only answer this question if the following conditions are met:
Answer was 'No' at question '173 [An36]' (Do you see any benefits in having a tool to share dashboards with colleagues?)

Please write your answer here:

Does the functionality “Collaborative Data Analysis” encourage you to analyse your data together with a colleague?

*****

Please choose **only one** of the following:

- Yes, strongly
- To a certain degree
- Maybe
- Not much
- No, not at all

Do you think using a short URL for sharing the dashboard is more suitable than sharing the automatically generated long URLs? *

Please choose **only one** of the following:

- Yes
- Maybe
- No

Would you use collaborative data analysis when analysing the data with a remote colleague? *

Please choose **only one** of the following:

- Yes
- Maybe
- No

Can you identify any benefits in using collaborative data analysis? *

Please choose **only one** of the following:

- Yes
- No

Which benefits can you identify in using collaborative data analysis?

Please write your answer here:

Why can't you identify any benefits in using collaborative data analysis?

Please write your answer here:

Think about the following situation: You are travelling, sitting in a train, and your colleague wants to show you some insights he found in a specific dataset. He sends you the link to the visualizer dashboard and uses "Collaborative Data Analysis" to present his results.

Please rate the beneficing of “Collaborative Data Analysis” in this situation.

*****

Please choose the appropriate response for each item:

very beneficing				not beneficing
○	○	○	○	○

Data Security

Is it self-explanatory that the shield icon is used to symbolize whether the Visualizer is or isn’t sharing data with the outside world? *

Please choose **only one** of the following:

- ○ Yes
- ○ No

Do you think the concept of using a shield within the Visualizer user interface to show the current status of data sharing is sufficient?

*

Please choose the appropriate response for each item:

yes, absolutely sufficient				too much/too little
○	○	○	○	○

What else could be displayed to raise awareness of the data sharing status?

Please write your answer here:

Is the provided colour encoding self-explaining?

- red: sharing (parts) of user’s data set is enabled for collaborative analysis
- orange: only anonymous, uncritical user interaction log data is shared (used for improving the Visualizer)
- green: nothing is shared (no data ever leaves user’s device)

*

Please choose the appropriate response for each item:

yes, absolutely self explaining				no, not at all
○	○	○	○	○

Would you use a tool which always enables collaborative data analysis?

*

Please choose the appropriate response for each item:

yes, of course				no, not at all
○	○	○	○	○

Comments on previous question welcome:

Please write your answer here:

Would you use a tool, where you can enable/disable collaborative data analysis as you prefer?

*

Please choose the appropriate response for each item:

yes, of course				no, not at all
○	○	○	○	○

Comments on previous question welcome:

Please write your answer here:

Would you use a tool which always enables anonymous logging of uncritical user interactions (no information about the dataset)?

*

Please choose the appropriate response for each item:

yes, of course				no, not at all
○	○	○	○	○

Comments on previous question welcome:

Please write your answer here:

Would you use a tool, where you can enable/disable logging of user interactions as you prefer?

*

Please choose the appropriate response for each item:

yes, of course				no, not at all
O	O	O	O	O

Comments on previous question welcome:

Please write your answer here:

Instructions for Using the Visualizer (provided via E-Mail)

Please rate the following statements *

Please choose the appropriate response for each item:

	very low				very high
How mentally demanding was it to follow the instructions?	○	○	○	○	○
How physically demanding was it to follow the instructions?	○	○	○	○	○
How insecure, discouraged, irritated, stressed, and annoyed were you when performing the tasks?	○	○	○	○	○

How successful in terms of following the analysis steps were you? *

Please choose the appropriate response for each item:

very successful				not successful
○	○	○	○	○

Have you been frustrated when trying to follow the instruction? *

Please choose the appropriate response for each item:

not frustrated				very frustrated
○	○	○	○	○

How difficult was it for you to use the Visualizer following the instructions? *

Please choose the appropriate response for each item:

not difficult at all				very difficult
○	○	○	○	○

After following the instructions once, are you able to perform analysis without instructions?

*****

Please choose **only one** of the following:

- ○ Of course, easily in all cases
- ○ Yes, in majority of cases
- ○ So-so, sometimes yes, sometimes no
- ○ Only in a few simple cases
- ○ No, not at all

Did the instructions help you in analysing your data?

*****

Please choose the appropriate response for each item:

yes, very helpful				no, not helpful at all
○	○	○	○	○

Further comments on the provided instructions:

Please write your answer here:

General subjective questions

Please rate the following statements:

*

Please choose the appropriate response for each item:

	strongly agree				strongly disagree
I would like to use the Visualizer frequently.	○	○	○	○	○
I think the Visualizer is easy to use.	○	○	○	○	○
I think I can use the Visualizer without support of an expert or further instructions.	○	○	○	○	○
I would imagine that most people would learn using the Visualizer very quickly.	○	○	○	○	○
I think the Table View of the Visualizer is well structured and intuitive to use.	○	○	○	○	○
The Table View provides a good overview of the provided dataset.	○	○	○	○	○
Performing data type changes in the Table View is easy to use.	○	○	○	○	○
Filtering of data in the Table View is easy to use.	○	○	○	○	○
Sorting of data in the Table View is easy to use.	○	○	○	○	○
I think the Visualizer area containing the visualizations is well structured and easy to use.	○	○	○	○	○
Generating a visualization by first selecting the required fields is very intuitive.	○	○	○	○	○
The interactions within the visualizations are in general intuitive.	○	○	○	○	○
The visualizations helped me in analysing and understanding the data.	○	○	○	○	○
Brushing (data selection) interactions within the visualizations are helpful in obtaining insights.	○	○	○	○	○
Coordination of multiple visualizations during brushing (“multiple coordinated views” feature) was helpful for data analysis.	○	○	○	○	○

Which of the provided visualizations have you been using before testing the Visualizer?

*

Please choose the appropriate response for each item:

	used very often				didn’t use at all
Bullet Chart	○	○	○	○	○
Stream Graph	○	○	○	○	○
Line Chart	○	○	○	○	○

Which of the provided visualizations helps in discovering new insights?

*

Please choose the appropriate response for each item:

	yes, discovered meaningful insights				no useful insights at all
Bullet Chart	◯	◯	◯	◯	◯
Stream Graph	◯	◯	◯	◯	◯
Line Chart	◯	◯	◯	◯	◯

Which of the provided visualizations help you in getting an overview of the data?

*

Please choose the appropriate response for each item:

	very helpful				not helpful at all
Bullet Chart	◯	◯	◯	◯	◯
Stream Graph	◯	◯	◯	◯	◯
Line Chart	◯	◯	◯	◯	◯

Which of the provided visualizations help you in identifying anomalies within the data?

*

Please choose the appropriate response for each item:

	very helpful				not helpful at all
Bullet Chart	◯	◯	◯	◯	◯
Stream Graph	◯	◯	◯	◯	◯
Line Chart	◯	◯	◯	◯	◯

Which of the provided visualizations help you in discovering and monitoring trends?

*

Please choose the appropriate response for each item:

	very helpful				not helpful at all
Bullet Chart	◯	◯	◯	◯	◯
Stream Graph	◯	◯	◯	◯	◯
Line Chart	◯	◯	◯	◯	◯

Which of the provided visualizations help you in unveiling data distributions and correlations?

*

Please choose the appropriate response for each item:

	very helpful				not helpful at all
Bullet Chart	○	○	○	○	○
Stream Graph	○	○	○	○	○
Line Chart	○	○	○	○	○

Which of the provided visualizations did you use for further restricting the data? (setting a brush/selecting elements)?

*****

Please choose the appropriate response for each item:

	used very often				didn't use at all
Bullet Chart	○	○	○	○	○
Stream Graph	○	○	○	○	○
Line Chart	○	○	○	○	○

Which further interactions did you perform for analysing your data? Select interactions and mention visualizations where you used them.

Comment only when you choose an answer.

Please choose all that apply and provide a comment:

☐ Zooming (magnifying):

☐ Panning (moving):

☐ Changing the visualization title:

☐ Data filtering:

☐ Data aggregation:

☐ Sorting:

☐ Re-configuring data-to-visualization mapping:

☐ Other interactions:

Did the provided visualizations and instruction in generating a dashboard in general help you in analysing your data?

*****

Please choose the appropriate response for each item:

yes, absolutely				no, not at all
○	○	○	○	○

Why did the provided visualizations and descriptions help you in analysing your data?

Please write your answer here:

Why did/why didn't the provided visualizations and descriptions help you in analysing your data?

Please write your answer here:

Why didn't the provided visualizations and descriptions help you in analysing your data?

Please write your answer here:

Which further visualizations did you try when using the Visualizer? Select those you have tried and rate their usefulness for the use case

Please choose the appropriate response for each item:

	very useful				not useful at all
Bar Chart	○	○	○	○	○
Pie Chart	○	○	○	○	○
Scatterplot	○	○	○	○	○
Scatterplot Matrix	○	○	○	○	○
Heatmap	○	○	○	○	○
Parallel Coordinates	○	○	○	○	○
Bubble Chart	○	○	○	○	○
Box Plot	○	○	○	○	○
Violin Plot	○	○	○	○	○
Grouped Bar Chart	○	○	○	○	○
Sankey Diagram	○	○	○	○	○
Radar Chart	○	○	○	○	○
Landscape	○	○	○	○	○

Final Thoughts

Please add any further comments here:

Please write your answer here:

Appendix D

Competitor Monitoring and Analysis

D.1 Dashboard Description

Search Interface Description

The user interface can be split up into five main parts.

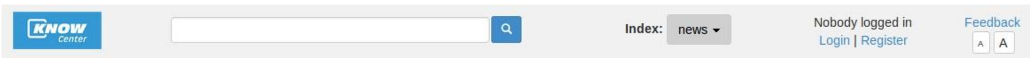
1) Search Form and Settings
2) Search Results
3) Facets
4) Visualizations
5) Portfolios

These parts are described in the following.

1) Search Form and Settings

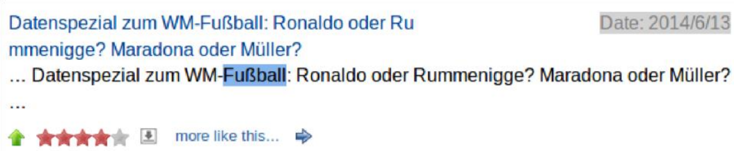
In the upper part of the Search-UI the user has to enter their search query as well as configuring several settings.

Query Input: User can put their search query receiving suggestions after entering several characters
Index-select: Allows the user to select the dataset to search in. Different search indices allow the user to make different sources searchable, e.g. web pages, shared or private devices.
Register/Login: Create account in order to receive further features on the search page
Feedback: Link to provide feedback as well as report bugs
Font-Size: Change font-size if text is displayed too small



2) Search Results

After searching for a query, the user receives the 10 most relevant search results, paging is provided to receive further results.
The search results are sorted by their relevance concerning the search query. They can also be sorted by their rating or modification date.
Furthermore, search results can be saved on the device or added to portfolios.



One search result consists of the Title which is displayed blue and is linked to the result page. Next to the title on the right side, the date is displayed. The date indicates the last modification of documents or extracted date information on web pages.
Below the title, the snippet is displayed. The snippet contains a small part of the content which may refer to the search query and helps in deciding whether the result is relevant for the selected query. The selected search query is highlighted in the snippet.
The last line of the search result contains several icons and links.

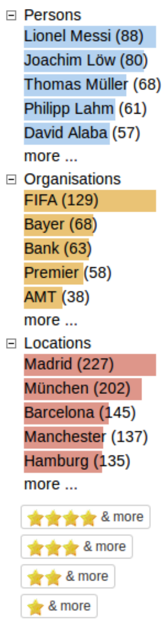
possible for a maximum number of ten results. Boosting can be removed by the red X-Icon which is replaced by the green arrow when boosted. Boosting is only allowed if a user is logged in.
Next to the boost-Icon there are the rating stars. These stars show how this result is rated by all users in average. Rating is only allowed if a user is logged in.
Next to the rating there is a download icon displayed. This icon indicates the type of result (web page, PDF, …) and allows to download or open the result.
Last but not least, the blue arrow to the right (only displayed when a user is logged in) allows the user to add a result to a portfolio. Clicking on the icon opens an dialogue to select an existing portfolio to add the result to.
Adding a result to a portfolio is also possible by dragging the result and drop it to the selected portfolio.

3) Facets

On the left side of the web page the facets are displayed. Facets are extracted meta data from the search results. Facets are grouped respectively to their type. Bar charts in the background of each facet shows as well as the corresponding number in brackets how many results contain the specific meta data. Clicking on selected meta data refines the search and narrows down the number of results. Afterwards this facet is highlighted and added to the breadcrumb bar below the search form.

Depending on the displayed visualization, further interaction possibilities are provided. These interaction possibilities are explained in Section visualizations.

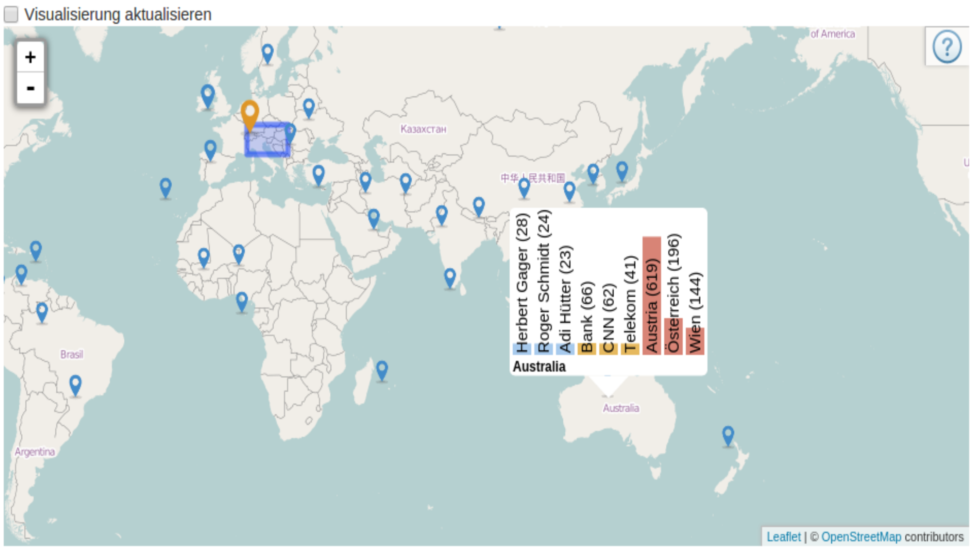
Below the facets we provide further refinement of the search results by their rating. Clicking on one button narrows down the amount of results by the the number of minimum selected stars.



4) Visualizations

Initially the visualizations are not displayed but only buttons for all visualizations. Clicking on one of the buttons, a tab view is opened with the selected visualization.
We provide four different visualizations, a geo-visualization, a theme river, a graph and a landscape visualization.
All visualizations (except the graph visualization) provide the possibility to freeze the visualization on refinement of the search by facet selection.

Geo-Visualization



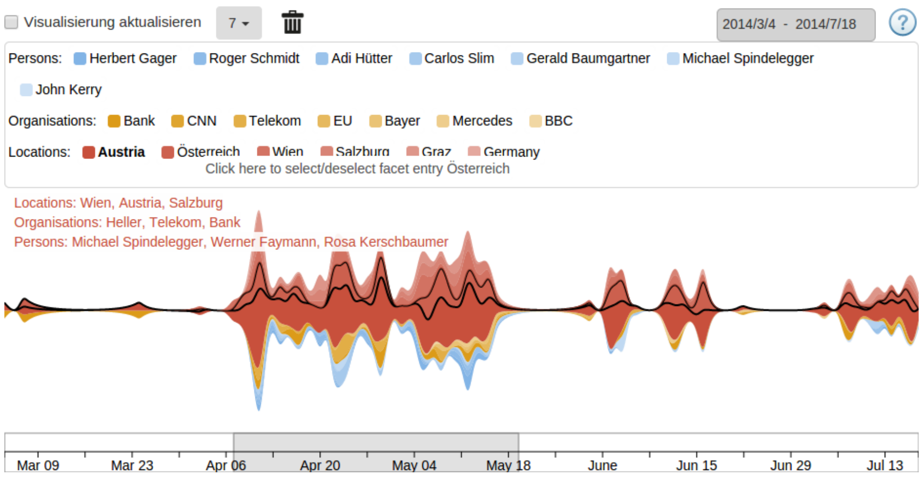
The geo visualization initially shows the whole map. Places are displayed as location icons. As many places can be displayed, close places are aggregated and displayed in one icon. The icon size depends on the number of places aggregated.

<u>Interaction possibilities:</u>
Zoom: Zooming in the geovis resolves aggregated locations and displays underlying places. Zooming in allows users to get a detailed view of several regions and getting an overview when zooming out.
Pan: Additionally to zooming, the geo visualization also allows to pan which allows better exploration of all locations without excessive zooming
Tooltip: On an aggregated icon displays the amount of places which are hidden by this icon. On a single place displays the most relevant facets of each facet type as bar charts in combination with the selected search query. E.g. search query “test” on mouseover of an location e.g. Australia the query “test” AND locations:”Australia” is used for getting the most relevant facets of each type.
Click: Clicking on an icon or bar chart view of a place selects the distinct place and in case of an aggregated place all places represented by this icon are used for refining the search. Selected places are displayed orange instead of light blue.
Area Select: Besides selecting single places, areas can be selected by CTRL + mouse-down, mouse-

move, mouse-up.

Interaction via facets:
Mouse-over of facets highlights those places which in the geo visualization which are results of the current search query in combination with current tooltip. E.g. search query “test” on mouseover of an organization e.g. CNN the query “test” AND organisations:”CNN” is used for highlighting locations.

Theme river



The theme river displays extracted information in the context of time. Time information are either extracted from the content e.g. web pages, the modification date of a document is used or the time of crawling the content.

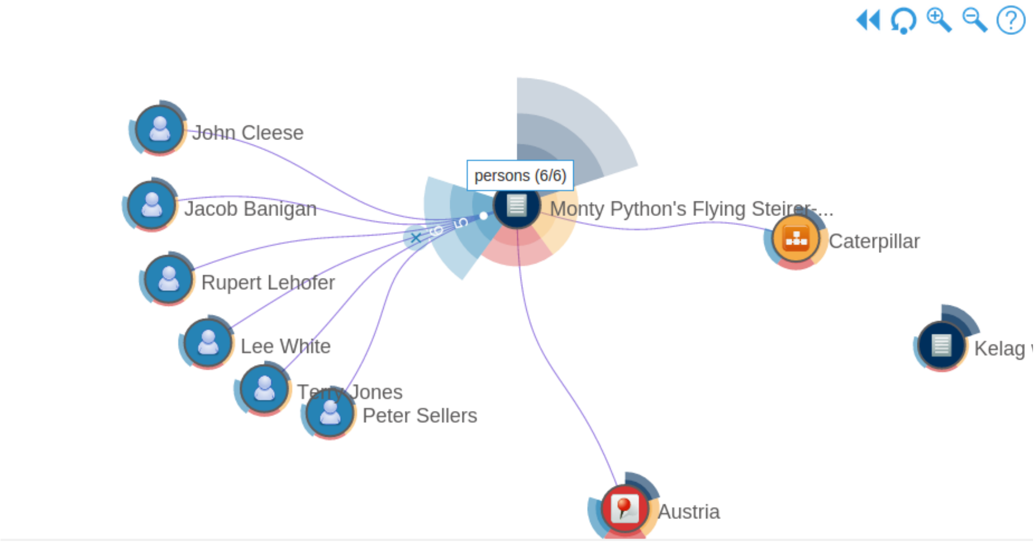
The uppermost area provides several configuration possibilities, the refreshment of the vis, the amount of facets for each type displayed as well as the possibility to remove facets from the visualization.
Below the legend for the theme river is displayed. Facets are sorted by their type and by the the amount of results.
Right at the bottom of the time range is displayed. This area allows the user to select a time range for search refinement.

Interaction possibilities:
Time range selection: Allows the user to refine the search in a temopral context as well as receive further insights concerning the results.
Tooltip: Tooltips on legend elements as well as on paths in the visualization highlights the corresponding path in the visualization as well as searches for the most important facets from the association index and displays them.
Select amount of facets: The dropdown menu in the upper part of this visualization view allows the user to define how many results from each type are displayed in the visualization.
Removing Facets: If specific facets are not interesting for the user, the user can remove these facets by dragging and drop it over the garbage icon. Besides single facets, all facets of a specific type can be removed by dragging the facet type to the garbage icon.

Click: Clicking on a facet ion the theme river legend or on the path in the theme river allows the user to refine the search.

Interaction via facets:
Add/remove facets: On mouse-over in the facet view, facets can be added or removed to the theme river. This can be done by the icon displayed next to the facet or by drag & drop.

Graph

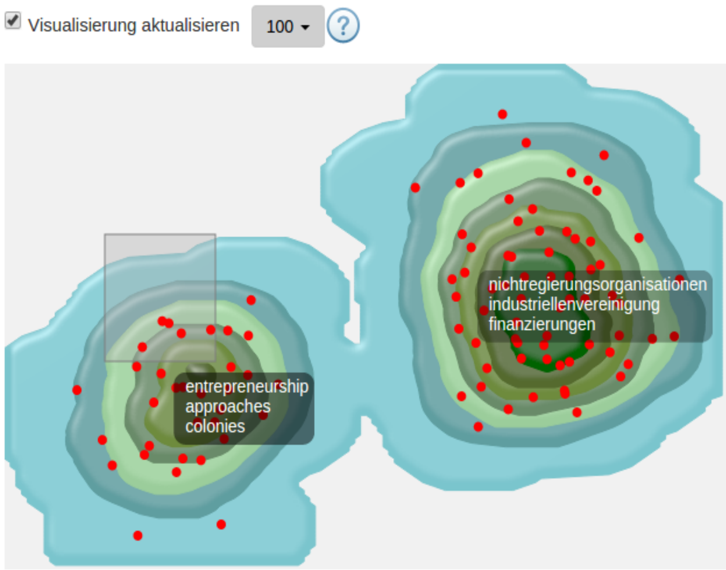


The graph visualization provides a different view of the search results. Initially the most relevant results are displayed as nodes. Meta data and thereof facets contained in these documents are displayed in the surrounding of each node. Clicking on these areas expands the corresponding meta data.

<u>Interaction possibilities:</u>
This visualization provides several possibilities, expansion of nodes, closing of nodes, zooming, panning, and many other methods by using node's context menu.
The search can be refined by clicking on facet nodes, documents can be opened on clicking on a document node.

Interaction via facets:
Add facet: Facets can be added to this visualization by drag & drop.

Landscape

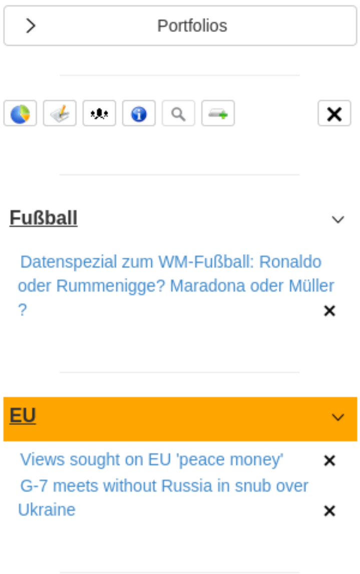


This visualization provides an overview of the content of the search results. Results are positioned according to their content as red dots. Results with similar content are positioned close together, whereas results with different content are in different clusters. The colour of the landscape differs depending on the amount of results positioned in this area. Different area contain their own labels consisting of the most important keywords.

Interaction possibilities:
Zoom: The landscape allows the user to zoom in to allow easier interaction if several results are close together.
Tooltip: Tooltips on the red dots provides information on the corresponding result.
Click: Clicking on the red dot opens the corresponding result. Clicking on several keywords in the labels refines the search by this keyword.
Area select: Selecting one or several results calculates the most important keywords and are displayed right to the visualization as a tag cloud.

Interaction via facets:
Highlight: On mouse-over on the facets in the facet view, the documents containing the specific facet as meta data is highlight.

5) Portfolios



The portfolio view allows user to collect a set of results for different themes in order to access them later fast, annotate them or share them with other users.
Portfolios can be selected by CTRL + Click in order to enable interaction possibilities.
Results can be added to portfolios by the corresponding icon or by drag & drop. Later further interaction possibilities like visualizing selected portfolios as well as searching for similar content should b^^e provided.

D.2 Evaluation Questionnaire

ProbandInnen-Nummer:

Allgemeine Angaben

Alter: _____
Geschlecht: □ männlich □ weiblich □ intersex
Beruf: _______________

Erfahrung mit Computern:
□ keine □ wenig □ mittel □ viel □ sehr viel

Erfahrung mit Suchinterfaces:
□ keine □ wenig □ mittel □ viel □ sehr viel

Fragen

Likert-scale:

- 1: trifft nicht zu
- 2: trifft wenig zu
- 3: vielleicht
- 4: trifft etwas zu
- 5: trifft völlig zu

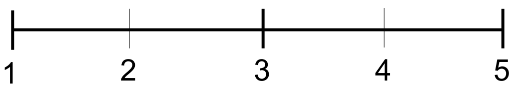
Suchresultate:

1) Sind die Informationen, welche bei einem Suchresultat dargestellt werden, ausreichend, zu wenig oder zu viel? Welche fehlen? Welche sind nicht nötig?

____

____

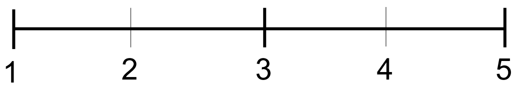
2) Das vorhandene Stern Bewertungsschema für Suchresultate ist angemessen.



Wenn nicht, warum? Was wäre ein besseres Bewertungsschema?

____

3) Snippets sind hilfreich um den Kontext des Dokuments zu erkennen.



ProbandInnen-Nummer:

4) Das Boosten von Dokumenten ist ein nützliches Feature um wichtige Dokumente hoch zu ranken.

1 2 3 4 5

Facetten:

5) Facetten sind eine hilfreiche Unterstützung bei der Verfeinerung der Suche.

1 2 3 4 5

6) Die Anzahl der Resultate in Klammern hilft bei der Verfeinerung der Suche.

1 2 3 4 5

7) Barcharts als visuelle Unterstützung bei den Facetten sind bei der Identifizierung der Häufigkeit der Facetten hilfreich.

1 2 3 4 5

3) Barcharts als visuelle Unterstützung bei den Facetten sind störend.

1 2 3 4 5

Portfolios:

4) Die Portfolio Ansicht eignet sich gut um Dokumente zu sammeln und später schnell darauf zugreifen zu können.

1 2 3 4 5

5) Die vorhandenen Features, Annotieren, Teilen, Visualisierung von Portfolios sind ausreichend.

1 2 3 4 5

ProbandInnen-Nummer:

Wenn nicht, welche Features würdest du dir noch wünschen? Welche Features sind nicht notwendig?

Visualisierungen:

6) Würdest du die vorgestellten Visualisierungen in einem Suchinterface verwenden, wenn ja, für welche Aufgaben? Wenn nicht, warum?
    - Geovis:
    - Themeriver:
    - Graph:
    - Landscape:

7) Welche Visualisierungen ist für folgende Aufgaben geeignet (Mehrfachantworten möglich):
    - Einschränkung der Suche nach regionalen Gebieten:
    - Überblick von extrahierte Metadaten der Suchresultate:
    - Übersicht der Themenzugehörigkeit:
    - Zeitliche Zuordnung von Dokumenten:
    - Häufung von Metadaten zu bestimmten Zeitpunkten:

8) Ist es von Vor- oder Nachteil, wenn mehrere Visualisierungen in einem Suchinterface zur Verfügung stehen und warum?

ProbandInnen-Nummer:

Allgemein:

9) Würdest du dieses Suchinterface verwenden? Wenn ja, wozu? Wenn nicht, warum nicht?

10) Wo siehst du die Vor- und Nachteile dieses Suchinterfaces:
   - Facetten:
     Vorteile:
     Nachteile:
   - Visualisierungen:
     Vorteile:
     Nachteile:
   - Portfolios:
     Vorteile:
     Nachteile:

11) Welche zusätzlichen Funktionalitäten würdest du dir wünschen?

ProbandInnen-Nummer:

12) Gib es eine existierendes Suchinterface, das du bevorzugst? Wenn ja, warum?

13) Zusätzliche Anmerkungen:

D.3 Evaluation Tasks

Suche, Facetten & Resultate:

Task 1:

Suche nach dem Begriff “Fußball” und sage laut, wie viele Ergebnisse es gibt.

Hilfe: Suche nach der Anzahl der Resultate bei der Resultat liste und sage die Anzahl laut.

Task 2:

Welche verschiedenen Facettentypen gibt es?

Hilfe:
1) In der linken Spalte neben den Suchresultaten befinden sich die Facetten
2) Jeder Facettentyp hat darunter liegende Elemente.
3) Nenne die einzelnen Typen

Task 3:

Welche ist die häufigste Facette?

Hilfe:
1) Betrachte nun die Einträge zu den verschiedenen Facettentypen
2) Hinter jeden Facettennamen befindet sich die Anzahl der Resultate in welchen diese Facette auftritt
3) Suche die Facette mit den häufigsten Resultaten und nenne sie laut.

Task 4:

Verfeinere nun deine Suche indem du die Facette “David Alaba” auswählst.

Hilfe:
1) Gehe zum Facettentyp “Persons”
2) Suche nach der Facette “David Alaba”
3) Klicke auf diese Facette

Task 5:

Wie hoch ist das Rating des ersten Resultats?

Hilfe:
1) Betrachte nun das erste Suchresultat
2) Das Rating wird in Form von Sternen dargestellt.
3) Wie viele der 5 Sterne sind in rot dargestellt

Task 6:

Entferne nun die selektierte Facette “David Alaba”

Hilfe:
1) Unterhalb der Suchleiste befindet sich die eine Leiste, wo die selektierten Facetten dargestellt werden.
2) Klicke auf das x-Icon hinter den Namen “David Alaba”

Fragen:
1. gar nicht
2. wenig
3. mittel
4. etwas
5. sehr

Wie groß war die mentale Herausforderung um diese Tasks durchzuführen?

1 2 3 4 5

Wie groß war die zeitliche Herausforderung um diese Tasks durchzuführen?

1 2 3 4 5

Wie gut glaubst du hast du in der Ausführung der Tasks abgeschnitten?

1 2 3 4 5

Wie groß war die Anstrengung um diese Tasks durchzuführen?

1 2 3 4 5

Wie groß war die Frustration bei der Durchführung diese Tasks?

1 2 3 4 5

Geo-Visualisierung:

Task 7:

Öffne die Geo-Visualisierung und finde heraus wie viele Orte in Australien gefunden wurden und wie diese heißen.

Hilfe:
1) Klicke auf den Button für die Geo-Visualisierung (unter der Breadcurmb Leiste und oberhalb der Suchresultate)
2) Suche den Kontinent Australien auf der Landkarte
3) Wie viele Icons siehst du?
4) Erfahre wie viel Städte sich hinter dem Icon befinden indem du mit der Maus darüberfährst und nenne die Zahl
5) Zoome mit dem Maus-Rad hinein bis die Tooltips dir die Namen der Städte liefern
6) Nenne die Städte

Task 8:

Verfeinere deine Suche indem du “Melbourne” auswählst und finde heraus wie viele der Ergebnisse Fußball-relevant sind indem du nur die Suchresultate betrachtest.

Hilfe:
1) Klicke auf das Icon, das für “Melbourne” steht und schränke damit die Suche ein
2) Betrachte die Titel & Snippet der Ergebnisse und sage wie viele der Ergebnisse wirklich relevant sein könnten.

Task 9:

Wähle nun alle Orte in Europa aus indem du die Brush-Funktionalität der Geo-Visualisierung verwendest. Welche Orte sind es?

Hilfe:
1) Führe CTRL+Mousedown – Mause bewegen – Mouseup über Europa aus um die Suche zu verfeinern
2) Die Breadcrumb Leiste zeigt dir an welche zusätzlichen Orte neben “Melbourne” ausgewählt wurden.
3) Nenne sie.

Fragen:

1. gar nicht
2. wenig
3. mittel
4. etwas
5. sehr

Wie groß war die mentale Herausforderung um diese Tasks durchzuführen?

1 2 3 4 5

Wie groß war die zeitliche Herausforderung um diese Tasks durchzuführen?

1 2 3 4 5

Wie gut glaubst du hast du in der Ausführung der Tasks abgeschnitten?

1 2 3 4 5

Wie groß war die Anstrengung um diese Tasks durchzuführen?

1 2 3 4 5

Wie groß war die Frustration bei der Durchführung diese Tasks?

1 2 3 4 5

Theme River:

Task 10:

Führe nun die Suche nach “Fußball” erneut aus und öffne den Themeriver. Finde heraus in welchem Zeitraum die Suchresultate umfassen.

Hilfe:

1) Nach erneuter Suche nach “Fußball” Klicke auf das zweite Symbol bei den Visualisierungen
2) Rechts oben in dem Tab steht die Information um welchen Zeitraum es sich handelt.

Task 11:

Schränke nun die Suche zeitlich ein, beginnend mit dem 25. Mai bis zum 22. Juni um herauszufinden welcher Ort bei am häufigsten vorkommt und wie oft und wann am häufigsten.

Hilfe:

1) Wähle in der Zeitleiste des Themerivers den Zeitraum 25. Mail bis 22. Juni mittels brushing.
2) Nachdem eine neue Suche ausgeführt wurde werden die Facetten links aktualisiert.
3) Der erste Eintrag bei den Locations zeigt den häufigsten Ort mit der Häufigkeit.
4) Wenn man nun diesen Ort im Themeriver betrachtet, so kann man über die größte Ausdehnung auf der Y-Achse herausfinden, wenn dieser Ort am häufigsten erwähnt wurde.

Task 12:

Fahre nun mit der Maus über diesen häufigsten Ort und finde heraus welche Organisationen in Zusammenhang mit diesem Ort gefunden werden.

Hilfe:

1) Fahre in der Legende (oder im Themeriver selbst) über der Ort der am häufigsten vorkommt
2) Unterhalb der Legende werden jene Facetten mit der häufigsten Assoziation zu diesem Ort angezeigt.
3) Nenne die Organisation, die an erster Stelle steht

Task 13:

Interaktionen:

1) Zeige für alle Facettentypen im Themeriver 10 Einträge
2) Entferne alle Orte aus dem Themeriver
3) Füge anschließend den Ort “Österreich” hinzu.

Hilfe:

1) Ändere die Anzahl der angezeigten Facetten im Dropdown Menü oberhalb der Legende im Themeriver auf “10”
2) Ziehe das Label “Locations” in der Themeriver Legende mittels Drag & Drop auf den Mülleimer in der Toolleiste.
3) Gehe mit der Maus links auf die Facetten auf Österreich und füge es über das “+” Symbol oder mittels Drag & Drop (in die Themeriver Legende) ein

Fragen:

1. gar nicht
2. wenig
3. mittel
4. etwas
5. sehr

Wie groß war die mentale Herausforderung um diese Tasks durchzuführen?

1 2 3 4 5

Wie groß war die zeitliche Herausforderung um diese Tasks durchzuführen?

1 2 3 4 5

Wie gut glaubst du hast du in der Ausführung der Tasks abgeschnitten?

1 2 3 4 5

Wie groß war die Anstrengung um diese Tasks durchzuführen?

1 2 3 4 5

Wie groß war die Frustration bei der Durchführung diese Tasks?

1 2 3 4 5

Graph:

Task 14:

Finde nun in Graphen heraus wie viele Orte im Dokument “Manipulationen um Fußball-WM 2022: Auch Australien soll bestochen haben” gefunden wurden öffne diese und nenne die Orte.

Hilfe:
1) Klicke auf den dritten Tab um den Graphen darzustellen
2) Finde das Dokument “Manipulationen um Fußball-WM 2022: Auch Australien soll bestochen haben” und finde mittels Tooltips heraus wie viele Orte darin vorkommen
3) Klicke auf den Bereich der die Zahl angibt um diese Knoten zu öffnen und nenne die Orte.

Task 15:

Schränke nun die aktuelle Suche im Graphen nach dem Ort “Katar” weiter ein und finde heraus wie viele ähnliche Dokumente das Dokument “WM in Katar: Flossen fünf Millionen Dollar Schmiergeld?” beinhaltet und öffne die 10 relevantesten.

Hilfe:
1) Finde den Knoten, der für den Ort “Katar” steht und klicke darauf.
2) Finde den Dokument-Knoten “WM in Katar: Flossen fünf Millionen Dollar Schmiergeld?”
3) Mittels Tooltips findet man heraus wie viele ähnliche Dokumente gefunden wurden.
4) Klicke auf den Bereich mit der “10” um die 10 ähnlichsten Dokumente zu öffnen.

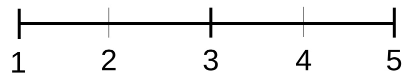
Task 16:

Ziehe die Organisation “FIFA” links bei den Facetten in den Graphen und finde heraus in wie vielen Dokumenten diese vorkommt.

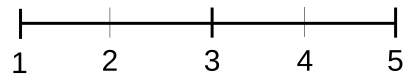
Hilfe:
1) Finde die Facette “FIFA” links im Facettenbaum unter dem Typ “Organisations”
2) Ziehe die Organisation mittels Drag & Drop in den Graphen
3) Finde mittels Tooltips heraus in wie vielen Dokumenten die Organisation FIFA vorkommt indem man auf das Kreissegment für Dokumente geht und die entsprechende Zahl vorliest.

Fragen:
1. gar nicht
2. wenig
3. mittel
4. etwas
5. sehr

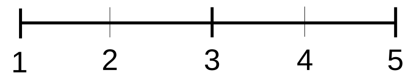
Wie groß war die mentale Herausforderung um diese Tasks durchzuführen?



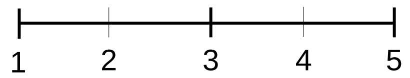
Wie groß war die zeitliche Herausforderung um diese Tasks durchzuführen?



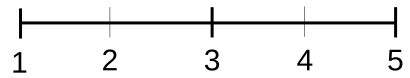
Wie gut glaubst du hast du in der Ausführung der Tasks abgeschnitten?



Wie groß war die Anstrengung um diese Tasks durchzuführen?



Wie groß war die Frustration bei der Durchführung diese Tasks?



Landscape:

Task 17:

Führe nun die Suche Fußball erneut aus, zeige die ersten 200 Resultate in der Landscape an und nenne die Inhalte der Labels.

Hilfe:
1) Führe die Suche nach "Fußball" erneut aus.
2) Klicke auf den vierten Tab um die Landscape anzuzeigen.
3) Wähle im Dropdown-Menü die Zahl 200 aus (Standardmäßig 20)
4) Nachdem die Landscape neu generiert wurde, werden die entsprechenden Labels mit den wichtigsten Begriffen angezeigt. Lies diese vor.

Task 18:

Finde heraus welche der offen dargestellten Facetten links in den meisten Dokumenten vorkommt.

Hilfe:
1) Fahre mit der Maus über die dargestellten Facetten im linken Bereich nacheinander
2) Bei welcher der Facetten werden am meisten Dokumente in der Landscape gehighlightet. Nenne diese Facette

Task 19:

Wähle nun alle Dokumente die sich um das Label "Vermessung, wm-fußballs" befinden mittels brush und nenne die wichtigsten Begriffe aus der Tag-Cloud

Hilfe:
1) Wähle die Dokumente, die sich im linken unteren Bereich finden mittels brush (Mousedown, Maus bewegen, Mouseup)
2) Rechts neben der Landscape werden die wichtigsten Begriffe dargestellt, nenne sie laut.

Fragen:

1. gar nicht
2. wenig
3. mittel
4. etwas
5. sehr

Wie groß war die mentale Herausforderung um diese Tasks durchzuführen?

1 2 3 4 5

Wie groß war die zeitliche Herausforderung um diese Tasks durchzuführen?

1 2 3 4 5

Wie gut glaubst du hast du in der Ausführung der Tasks abgeschnitten?

1 2 3 4 5

Wie groß war die Anstrengung um diese Tasks durchzuführen?

1 2 3 4 5

Wie groß war die Frustration bei der Durchführung diese Tasks?

1 2 3 4 5

User- und Portfolio-Management:

Task 20:

Registriere dich als Benutzer und logge dich ein.

Hilfe:
1) Links oben auf der Seite befinden sich die Links zum Registrieren und einloggen
2) Registriere dich zuerst mit deinem Namen und logge dich ein

Task 21:

Bewerte das zweite Suchergebnis in der Resultatsliste und suche nach ähnlichen Resultaten.

Hilfe:
1) Finde das erste Ergebnis in der Resultatsliste.
2) Gehe mit der Maus über die Sterne, diese werden nun gelb eingefärbt.
3) Je nach gewünschter Anzahl an Sternen geht klickt man auf den entsprechenden Stern um eine Bewertung abzugeben
4) Klicke anschließend auf den Link "more like this..." um nach ähnlichen Dokumenten zu suchen

Task 22:

Erzeuge ein Portfolio "Fußball" und füge ein relevantes Ergebnis hinzu (Suchbegriff: "Fußball").

Hilfe:

1) Erzeuge in der Portfolio-Ansicht ein neues Portfolio "Fußball" indem du auf das Icon mit dem grünen "+" klickst
2) Suche erneut nach dem Begriff "Fußball"
3) Wähle ein Resultat und füge es mittels Drag & Drop im Portfolio hinzu

Task 23:

Selektiere das Portfolio „Fußball" und annotiere es mit "soccer".

Hilfe:
1) Selektiere das zuvor erstelle Portfolio mit CTRL + Klick
2) Klicke auf das blaue "i" Icon in der Icon Liste
3) Schreibe den Begriff "soccer" ins Tag-Input Feld und klicke auf den "Add" Button
4) Klicke anschließend auf "Save" um die Annotation abzuschließen

Task 24:

Teile dein Portfolio mit dem Benutzer "maxmustermann" und gib ihm Lese- und Schreibrechte.

Hilfe:

1) Selektiere dein Portfolio erneut (wenn nötig) mit CTRL + Klick
2) Klicke auf das Icon mit den Personen (3 Icon)
3) Wähle den User "maxmustermann" aus und setze das Zugriffsrecht auf "write"
4) Klicke anschließend auf "Add/Update" um die Änderung zu übernehmen und schließe den Dialog

Fragen:

1. gar nicht
2. wenig
3. mittel
4. etwas
5. sehr

Wie groß war die mentale Herausforderung um diese Tasks durchzuführen?

1 2 3 4 5

Wie groß war die zeitliche Herausforderung um diese Tasks durchzuführen?

1 2 3 4 5

Wie gut glaubst du hast du in der Ausführung der Tasks abgeschnitten?

1 2 3 4 5

Wie groß war die Anstrengung um diese Tasks durchzuführen?

1 2 3 4 5

Wie groß war die Frustration bei der Durchführung diese Tasks?

1 2 3 4 5