

Bachelorarbeit am Institut für Informatik der Freien Universität Berlin

Human-Centered Computing (HCC), AG NBI

Untersuchung von Auswirkungen der Adaptierbarkeit einer Visualisierung auf die Benutzerzufriedenheit

– Exposé –

Stefan Rolfs

Matrikelnummer: 4287451

email@stefan-rolfs.de

Betreuerin: Prof. Dr. C. Müller-Birn

Berlin, 11. März 2013

1 Motivation der Arbeit

Mit Hilfe computerunterstützter Methoden zur grafischen Repräsentation großer Mengen von Daten (Informationsvisualisierung) wird es dem Menschen möglich, die zentralen Informationen dieser Daten effektiv zu erfassen.

Solche Informationsbedürfnisse entstehen in der Forschung zum Beispiel hinsichtlich einer möglichen finanziellen Unterstützung - etwa durch Kooperationen mit Institutionen oder durch Förderprogramme. Auch die Erleichterung der Suche nach potentiellen Projektmitarbeitern wäre wünschenswert. Von allgemeinem Interesse ist ferner die Veränderung der Förderlandschaft Deutschlands - insbesondere die Frage, in welchen Zeiträumen und Regionen welche wissenschaftlichen Disziplinen besonders im Fokus der Förderung standen. In dieser Arbeit finden für die Visualisierung daher die Daten der Deutschen Forschungsgemeinschaft Verwendung.

2 Thematische Einordnung der Arbeit

Die Menge an Daten wächst laut Gantz und Reinsel, 2012 [3] exponentiell an. Aus diesen Informationen können wertvolle Erkenntnisse gewonnen werden. Damit es dem Menschen möglich wird, die enormen Datenmengen zu erfassen, müssen geeignete Visualisierungsformen gefunden werden.

Solche Visualisierungsformen sind unterschiedlich gut für verschiedene zu visualisierende Daten geeignet. Im Rahmen dieser Arbeit wird eine interaktive Visualisierung entworfen und implementiert, welche sich für die Visualisierung von Daten, wie sie die Deutsche Forschungsgemeinschaft bereitstellt, als geeignet erweist.

Hierfür ist es zunächst erforderlich, die für die Visualisierung notwendigen Daten zu beschaffen. Die Deutsche Forschungsgemeinschaft bietet keine Schnittstelle für das Abrufen der dort gespeicherten Daten an. Dieser Umstand macht das Extrahieren der Daten aus den Webseiten des Informationssystems GEPRIS¹ erforderlich.

Möglicherweise wurden von der Deutschen Forschungsgemeinschaft Vorkehrungen getroffen, die das automatisierte Auslesen der Webseiten unterbinden bzw. erschweren sollen.

Das Aufrufen einer GEPRIS-Seite ist nur mit einem Cookie möglich, der die Session-Variable enthält. Vor dem Erreichen der gewünschten Seite muss ein Request abgesetzt werden, der zunächst alle Variablen in die entsprechende Session schreibt. Diese Besonderheit wird von den zu implementierenden Scrapern beachtet.

¹<http://gepris.dfg.de/>

Die extrahierten Daten werden in eine MySQL-Datenbank überführt. Für den Entwurf des dafür notwendigen Datenbank-Schemas werden die Seiten von GEPRIS hinsichtlich ihrer Struktur untersucht.

Der konzeptuell „saubere“ Entwurf sollte die Voraussetzung aller Datenbank-anwendungen sein.

Kemper und Eickler, 2006, S. 31 [4]

Die schließlich vorliegenden Daten werden im Rahmen dieser Bachelorarbeit interaktiv visualisiert. Das Ziel von Visualisierungen ist es, den Benutzern zu neuen Erkenntnissen zu verhelfen, indem Sie die vorliegenden Daten in geeigneter Weise darstellen.

There are many visual design guidelines but the basic principle might be: summarized as the Visual Information Seeking Mantra:

1. Overview first
2. zoom and filter
3. then details-on-demand

Ben Shneiderman, 1996, S. 337 [9]

Eine gute Visualisierung sollte dem Benutzer demnach anfangs einen groben Überblick anbieten und dabei Detailinformationen zusammenfassen. Die Möglichkeit, die Darstellung anzupassen (z.B. durch Zoomen oder Verschieben des Kartenausschnittes) erlaubt es den Benutzern, Details zu untersuchen und unnötige Daten auszublenden. Schließlich wird der Benutzer die gesuchten Informationen finden und/oder neue Erkenntnisse - z.B. über das Auftreten von Anomalien oder über Trends - erlangen. Dabei soll die Visualisierung helfen, die Daten näher mit ihrer eigentlichen Bedeutung für den Benutzer zu verknüpfen.

Während die Benutzer stetig mehr Einblick in die Daten erhalten und zu Erkenntnissen gelangen, könnte sich der Wunsch nach einer anderen Form der Visualisierung manifestieren.

The idea of insight as suddenly reorganizing visual information and insight as the reformulation of a problem seem to be especially relevant for information visualization. In information visualization, the reorganization of visual information can be supported by multiple views, zooming, panning, filtering and similar means.

Pohl et al., 2010, S. 17 [8]

Für verschiedene Aufgaben sind unterschiedliche Visualisierungsformen zudem unterschiedlich gut geeignet. Dies spricht für die Umsetzung von sogenannten Multiple Views.

In einer ersten Version der Anwendung wird nur eine Sicht implementiert werden. Weitere Sichten sind nach Auswertung der Evaluation - welche im Anschluss an die Implementierung erfolgen wird - jedoch denkbar.

Das Ziel dieser Evaluation (vgl. „Methodische Umsetzung“ auf Seite 5) ist es, die bisherige Eignung der Visualisierung festzustellen und Verbesserungspotenziale aufzudecken. Hierzu wird den Testbenutzern die Anwendung vorgestellt, ihnen die Interaktionsmöglichkeiten erläutert sowie die Art der Daten und deren Nutzungsmöglichkeiten vermittelt. Im Anschluss erhalten die Benutzer Aufgaben, die sie mithilfe der Visualisierung lösen sollen. Hierbei sind sie angehalten, ihre Gedanken laut auszusprechen (Thinking Aloud Test). Das Verhalten der Benutzer, ihre Anmerkungen und auftretende Probleme werden protokolliert. Nach der Benutzung der Anwendung werden die Benutzer nach ihren Meinungen und Verbesserungsvorschlägen, sowie zu möglichen Problemen befragt.

Die Ergebnisse der Evaluation werden Einfluss auf die iterative Weiterentwicklung des Interaktionsdesigns nehmen.

Auf eine technische Analyse wie Auswertungen von Log-Dateien und des Klickverhaltens sowie auf das Messen der Dauer von Aktionen wird verzichtet. Zum einen würden die Ergebnisse der Messung aufgrund der gewählten Evaluationsmethode verfälscht.

Furthermore, evidence shows that users who are thinking aloud both change their behaviour and are slower than users who do not think aloud [2, page 105].

Andrews, 2006, S. 2 [1]

Zum anderen ist die Aussagekraft einer Zeitmessung für den Erfolg einer Visualisierung begrenzt. Sind Anwender von einer Visualisierung begeistert, so werden sie die angebotenen Interaktionsmöglichkeiten erforschend ausprobieren in einer Art, in der sie vermutlich nicht auf dem schnellsten Weg die ihnen zugewiesene Aufgabe erfüllen. Doch gerade dieses Erforschen kann zu interessanten neuen Erkenntnissen führen, was für den Erfolg des Experiments spricht.

North [6] emphasizes that it is difficult to evaluate the quality and effectiveness of a complex InfoVis tool with quantitative methods.

Pohl et al., 2010, S. 17 [8]

3 Zielstellung

Ziel dieser Arbeit ist die Entwicklung und iterative Verbesserung einer interaktiven Visualisierung unter Einbeziehung von Benutzern. Auf diesem Weg soll die Visualisierung

so realisiert werden, möglichst umfassend bestehende Informationsbedürfnisse hinsichtlich der Daten der Deutschen Forschungsgemeinschaft zu befriedigen.

- So soll herausgefunden werden, welche Formen der Interaktion vom überwiegenden Teil der Benutzer als intuitiv und nützlich angesehen werden, um diese Erkenntnisse in die Weiterentwicklung der Filtermöglichkeiten einfließen zu lassen.

Die erste Version der Anwendung wird dabei über einen Disziplinen-Filter verfügen, womit die wissenschaftlichen Fachgebiete selektiert werden können. Dieser wird als initial eingeklappter Checkbox-Tree umgesetzt. Bei der Auswahl einer Disziplin werden - sofern vorhanden - auch die untergeordneten Fachbereiche berücksichtigt. Das Aktivieren einer Checkbox sorgt für die Einbeziehung des Fachgebiets in die Visualisierung und markiert den Namen der Disziplin im Baum mit der für die Darstellung verwendeten Farbe.

Ein weiteres Filter-Element stellt ein Schalter dar, mithilfe dessen der Anwender entscheidet, ob in der Visualisierung Personen oder Institutionen dargestellt werden sollen.

Die dritte Interaktionsmöglichkeit bietet ein Range-Slider, mit dem der zu berücksichtigende Zeitraum eingegrenzt werden kann. Auf diese Weise können Veränderungen und Trends untersucht werden.

Interagieren kann der Benutzer auch durch die Bewegung des Mauszeigers über eines der Elemente der Visualisierung. Handelt es sich bei dem Element um ein Segment eines Kreisdiagramms oder um eine Verbindungslinie zwischen zwei Orten, wird ein Tooltip angezeigt. In der ersten Version der Visualisierung wird diesen Namen der Disziplin und die Anzahl der Personen bzw. Institutionen angezeigt, welche in diesem Element zusammengefasst wurden.

Durch die Wahl einer Karten-Darstellung (vgl. „Geplante Vorgehensweise“ auf Seite 5) sind implizit weitere Interaktionsmöglichkeiten vorgegeben. Der jeweils sichtbare Kartenausschnitt ist entscheidend für die Berücksichtigung von Institutionen oder Personen je nach Längen- und Breitengrad. Verändern kann der Benutzer diesen Ausschnitt durch Verschieben mithilfe der Maus (Drag & Drop) oder durch Vergrößern bzw. Verkleinern mithilfe des Mauseisens.

- Es wird im Rahmen dieser Arbeit untersucht werden, wie die Funktionen verbessert werden können, um eine effektivere Arbeit mit der Anwendung zu ermöglichen.

4 Geplante Vorgehensweise

Der Verlauf der Arbeit gliedert sich in drei wesentliche Phasen:

1. Da das „GEPRIS“ - das Informationssystem der DFG - keine API bereitstellt, ist zunächst das maschinelle Auswerten der Katalogseiten notwendig. Mittels *Web Scraping* werden dabei die auf den Seiten dargestellten Informationen extrahiert und in eine Datenbank geschrieben.
2. Die gesammelten Daten werden im Anschluss durch eine Web-Anwendung visualisiert. Hierbei werden Personen bzw. Institutionen sowie deren Kooperationen in gemeinsamen Produkten als Netzwerk in einer Karte dargestellt. Die Knoten des Netzwerkes stellen hierbei Kreisdiagramme dar, deren Radius von der Anzahl der am Ort befindlichen Personen / Institutionen abhängt. Die Kanten und deren Breite stellen die Anzahl derjenigen Projekte dar, bei denen Personen bzw. Institutionen jeweils aus beiden von der Kante verbundenen Orten beteiligt waren.
3. Nach Fertigstellung einer ersten Version der Web-Anwendung wird eine Evaluation durchgeführt (vgl. „Thematische Einordnung“ auf Seite 3 und „Methodische Umsetzung“ auf Seite 5). Hierfür wird den Nutzern die Anwendung und das Ziel der Visualisierung vermittelt. Anschließend werden ihnen Aufgaben gestellt, die sie mithilfe der Visualisierung beantworten sollen. Während der Benutzung der Anwendung sollen die Benutzer ihr Vorgehen kommentieren. Schlussendlich können die Benutzer sich dazu äußern, wie gut die Visualisierung ihnen bei der Beantwortung der Fragen geholfen hat, wo Probleme auftraten und welche zusätzlichen Funktionen oder Änderungen sie sich wünschen würden. Die Ergebnisse der Benutzerbefragung fließen in die Weiterentwicklung der Web-Anwendung ein.

5 Methodische Umsetzung

Bei der Bearbeitung dieser Bachelorarbeit werde ich nach der Design Science Research Methodology (DSRM) vorgehen, wie sie von Peffers et al. (2008) [7] beschrieben wurde.

The DS process includes six steps: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication.

Pfeffers et al., 2008, S. 4[7]

In diesem Prozess-Modell beginnt der Prozess als „problem centered“ in der Phase „Problem Identification and Motivation“. Es soll herausgefunden werden, welche Interakti-

onsformen von den Benutzern präferiert werden.

Ziel der zu entwickelnden Visualisierung ist es, den Benutzern zu ermöglichen, relevante Informationen schnell und zuverlässig zu erfassen. Dabei wirken die Benutzer durch Anbringen von Kritik und Verbesserungsvorschlägen im Rahmen der Evaluationsphase selbst mit.

Als Evaluationsmethode eignet sich einerseits die Sichtung der Implementierung durch Domänen-Experten.

[...] specialist evaluators inspect an interface and use their experience and judgement to assess it [...]

Andrews, 2006, S. 1 [1]

Von Heidi Lam et al. (S. 1530) [5] wird diese Methode der Evaluation „Informal Evaluation“ genannt.

Außerdem werden Thinking Aloud Tests mit einer kleinen Gruppe von Testbenutzern durchgeführt, bei denen diese während der Interaktion mit der Anwendung Ihre Gedanken verbalisieren.

A thinking aloud test is a valuable technique during the development of an information visualisation to gain design feedback and fix bugs.

Andrews, 2006, S. 1 [1]

Heidi Lam et al. spricht hierbei von „Usability Tests“ und nennt als Hauptziel die Optimierung des Entwurfs. Es wird im Rahmen dieser Arbeit hauptsächlich darum gehen, die Visualisierung hinsichtlich der Interaktionsmöglichkeiten zu verbessern bzw. zu erweitern, wenngleich auch die Benutzerfreundlichkeit der implementierten Anwendung ausschlaggebend für die Bewertung der Anwender sein kann.

Im Anschluss an die Evaluationsphase werden deren Ergebnisse in die iterative Weiterentwicklung der Visualisierung einfließen, wobei erneut die DSRM-Phase „Design and Development“ durchlaufen wird. Die Kommunikation der Ergebnisse erfolgt durch die Veröffentlichung dieser Arbeit.

6 Technische Umsetzung

Um die benötigten Daten von den Seiten der DFG zu extrahieren, wird die Plattform ScraperWiki² genutzt. Für die unterschiedlichen Entitäten *person*, *institution*, *project*

²<https://scraperwiki.com/>

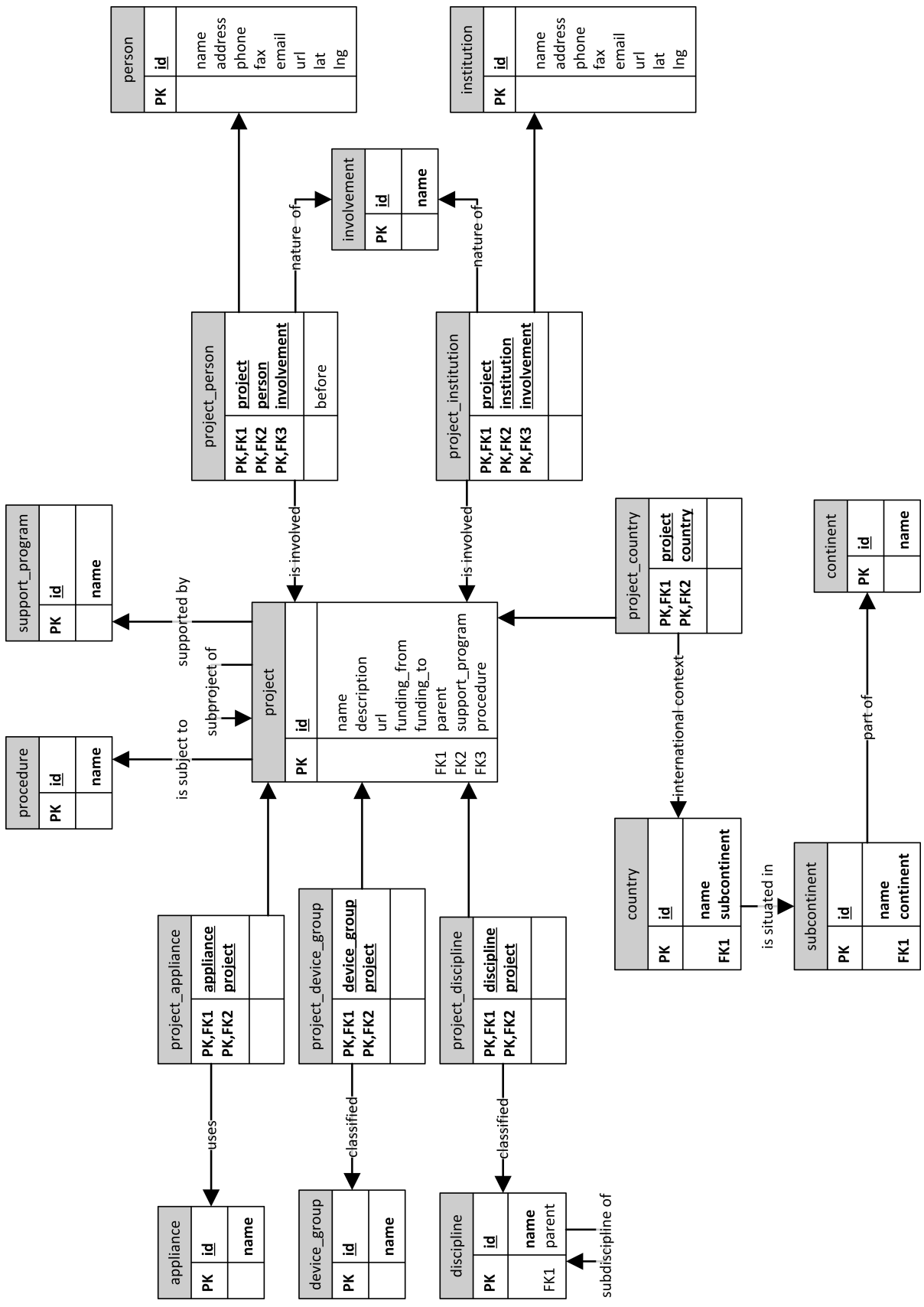


Abbildung 1: Datenbankstruktur

etc. (vgl. Datenbankstruktur in Abbildung 1) sammeln kleinere PHP-Skripte die Daten in der ScraperWiki-Datenbank, von wo aus diese über das CSV-Format importiert werden können.

Projekte und Institutionen verfügen zu einem großen Teil über Adressen. Für die Darstellung in einer Karte werden mithilfe der Google Geocoding API³ die Koordinaten ermittelt. Aufgrund der Beschränkung von maximal 2.500 Anfragen pro Tag erfolgt dies sukzessive.

Der Server, auf dem die zu entwickelnde Web-Anwendung laufen wird, erhält ein Linux-Betriebssystem (Cent OS). Die verwendete Programmiersprache PHP liegt für das Cent OS in Version 5.3.3 vor. Als HyperText Transfer Protocol Server wird der Apache HTTP Server httpd (Version 2.2.15) eingesetzt. Das relationale Datenbankverwaltungssystem MySQL wird in Version 5.1.66 installiert. Die im Rahmen der Bachelorarbeit entstehenden Dateien werden mit dem Versionsverwaltungssystem Git⁴ (Version 1.7.1) versioniert.

Als Entwicklungsumgebung wird eclipse eingesetzt. Außerdem wird das PHP Framework Symfony⁵ Verwendung finden. Mit diesem Framework wird ein Model View Controller Ansatz verfolgt. Dabei wird für die View-Komponente auf die Template Engine Twig⁶ gesetzt. Ein Front-Controller nimmt sämtliche Anfragen an den Server entgegen und entscheidet anhand der URL über den im weiteren Verlauf der Bearbeitung zuständigen Controller. In diesen Controllern erfolgt der Zugriff auf die Model-Komponente. Diese wird bei Anfragen zunächst prüfen, ob bereits eine Version der angefragten Daten im Cache vorliegt. Ist dies nicht der Fall, werden über den Objektrelationalen Mapper Doctrine⁷ die benötigten Daten erfragt.

Für die Visualisierung wird die JavaScript Bibliothek *Data-Driven Documents* (D3.js) eingebunden.

7 Erster Terminplan

Nach derzeitiger Planung wird die Bachelorarbeit Mitte März angemeldet werden. Für die Bearbeitung der Arbeit stehen 12 Wochen zur Verfügung. Die Fertigstellung ist für Ende Mai geplant.

³<https://developers.google.com/maps/documentation/geocoding/>

⁴<http://git-scm.com>

⁵<http://symfony.com/>

⁶<http://twig.sensiolabs.org/>

⁷<http://doctrine-project.org>

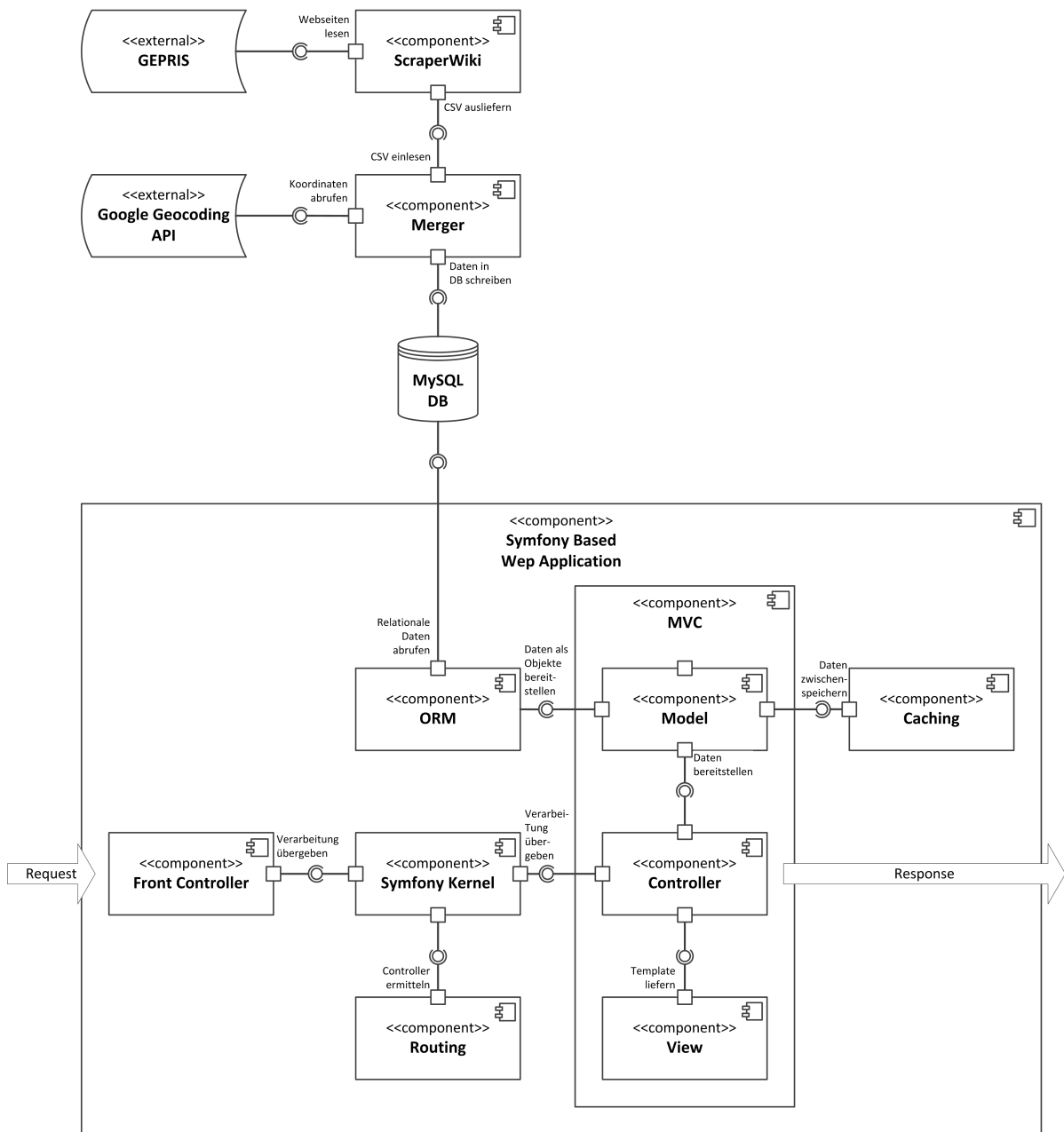


Abbildung 2: Komponenten-Diagramm

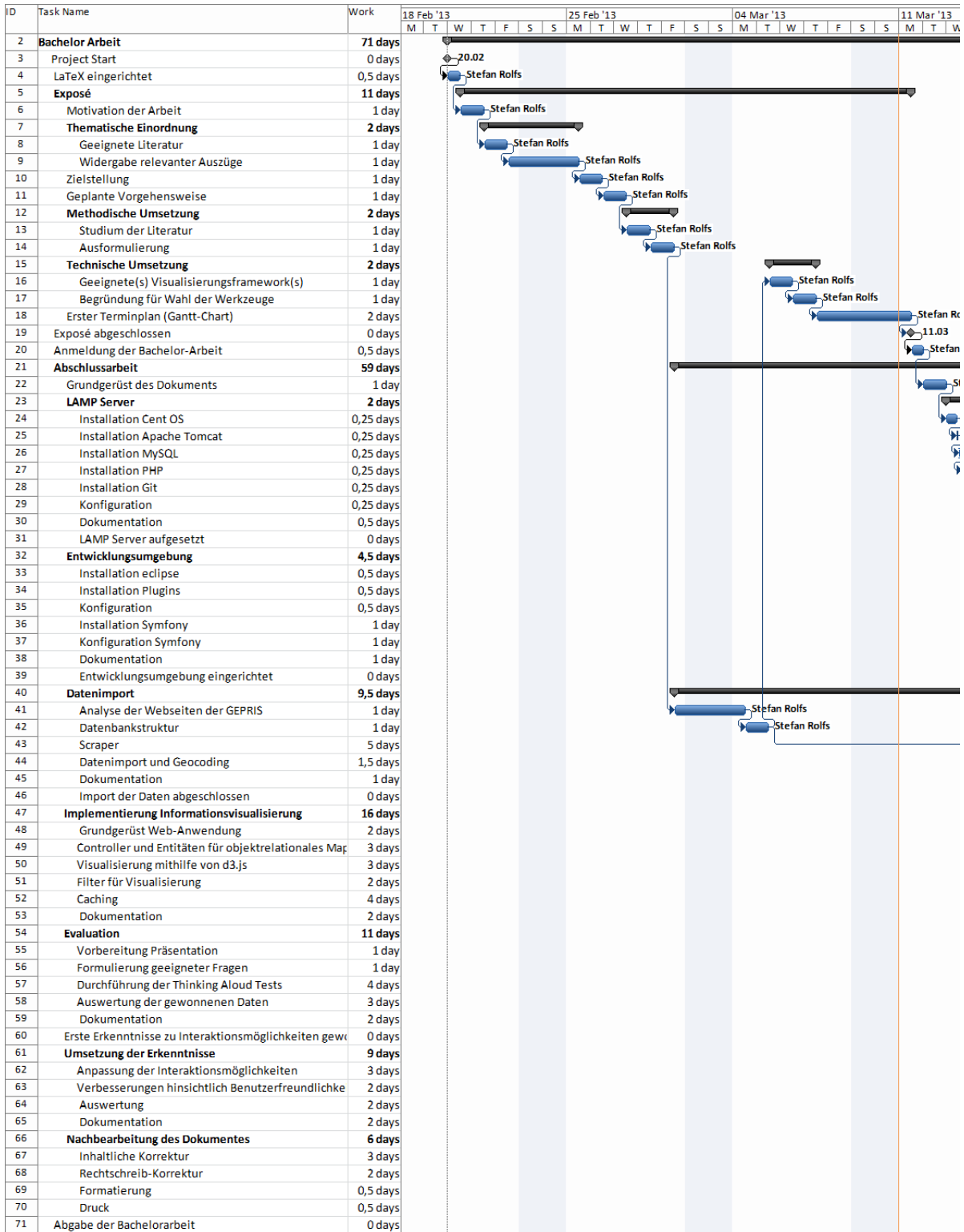


Abbildung 3: Gantt-Diagramm

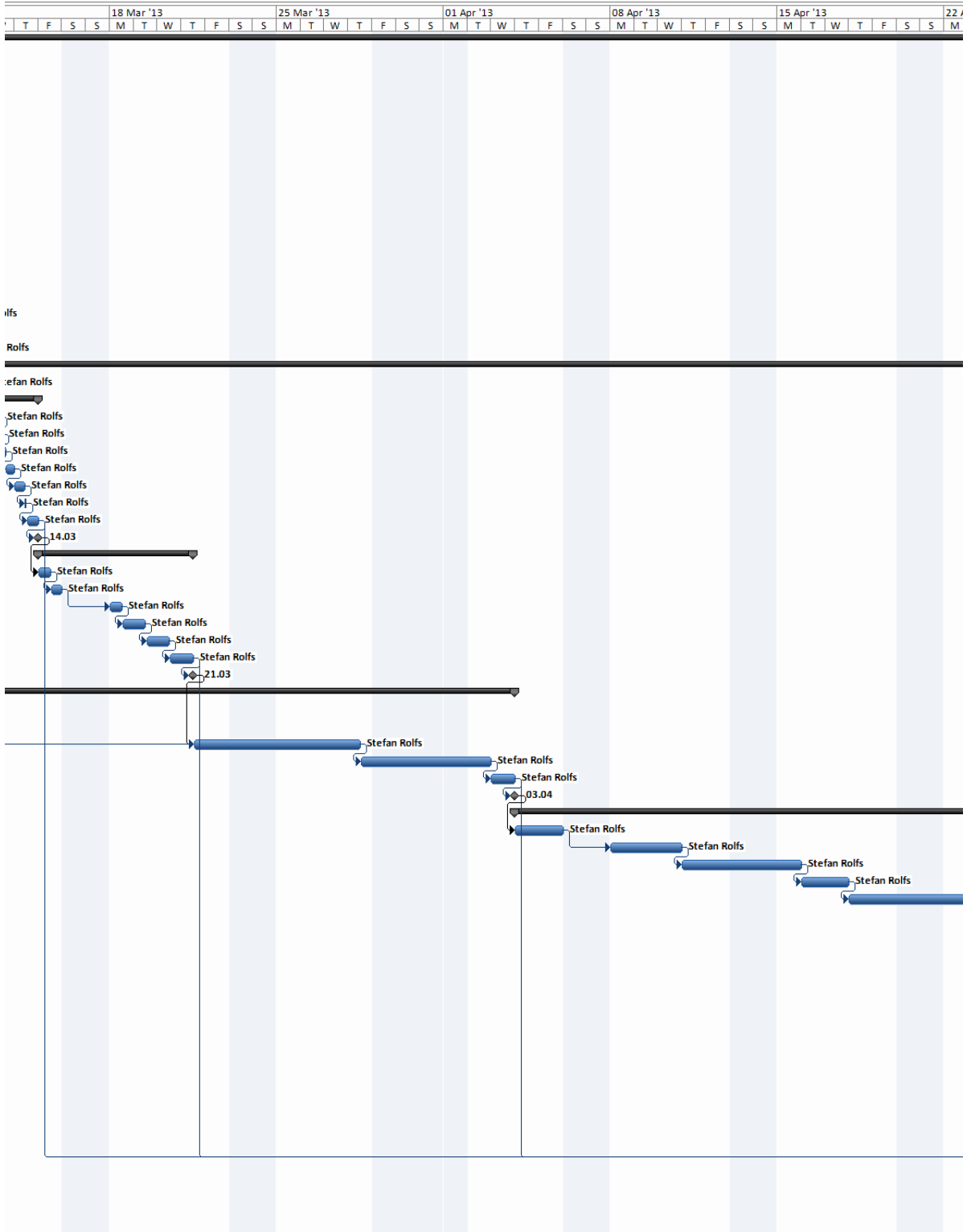


Abbildung 3: Gantt-Diagramm (Fortsetzung)

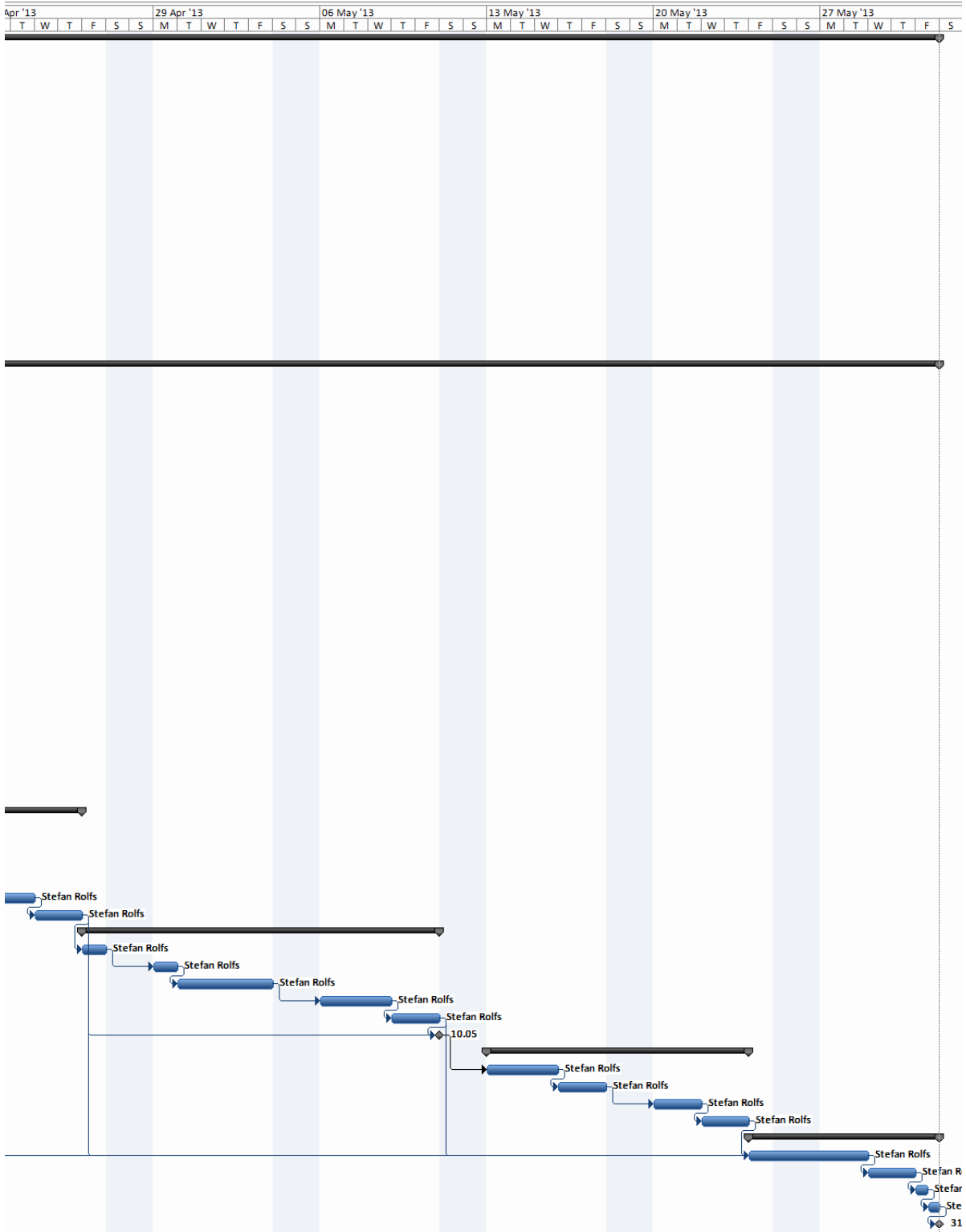


Abbildung 3: Gantt-Diagramm (Fortsetzung)

Abbildungsverzeichnis

1	Datenbankstruktur	7
2	Komponenten-Diagramm	9
3	Gantt-Diagramm	10

Literatur

- [1] Keith Andrews. Evaluating information visualisations. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, BELIV '06, pages 1–5, New York, NY, USA, 2006. ACM.
- [2] Karl Anders Ericsson and Herbert Alexander Simon. *Protocol analysis: Verbal reports as data (rev. ed.)*. MIT Press, 1993.
- [3] John Gantz and David Reinsel. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC*, 2012(December):1–16, 2012.
- [4] A. Kemper and A. Eickler. *Datenbanksysteme*. Oldenbourg, 2006.
- [5] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012.
- [6] Chris North. Toward measuring visualization insight. *IEEE Comput. Graph. Appl.*, 26(3):6–9, May 2006.
- [7] Ken Peffers, Tuure Tuunanen, Marcus Rothenberger, and Samir Chatterjee. A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3):45–77, 2008.
- [8] Margit Pohl, Sylvia Wiltner, and Silvia Miksch. Exploring information visualization: describing different interaction patterns. In *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaluation methods for Information Visualization*, BELIV '10, pages 16–23, New York, NY, USA, 2010. ACM.
- [9] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages, VL '96*, pages 336–, Washington, DC, USA, 1996. IEEE Computer Society.