



FernUniversität in Hagen

– Fakultät für Mathematik und Informatik –
Lehrgebiet Datenbanken und Informationssysteme

Analyse und Optimierung der Datenbankabfragen der Webseite des Wedekind Projektes

Abschlussarbeit zur Erlangung des akademischen Grades
Bachelor of Science (B.Sc.)

vorgelegt von

Marco Galster

Matrikelnummer: 8335710

Referentin : Prof. Dr. Uta Störl

Betreuer : —

ERKLÄRUNG

Ich erkläre, dass ich die Expose selbstständig und ohne unzulässige Inanspruchnahme Dritter verfasst habe.

Ich habe dabei nur die angegebenen Quellen und Hilfsmittel verwendet und die aus diesen wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht.

Die Versicherung selbstständiger Arbeit gilt auch für enthaltene Zeichnungen, Skizzen oder graphische Darstellungen.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form weder derselben noch einer anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Mit der Abgabe der elektronischen Fassung der endgültigen Version der Arbeit nehme ich zur Kenntnis, dass diese mit Hilfe eines Plagiatserkennungsdienstes auf enthaltene Plagiate geprüft werden kann und ausschließlich für Prüfungszwecke gespeichert wird.

Hagen, 01. Oktober 2023

Marco Galster

INHALTSVERZEICHNIS

1	Expose	1
1.1	Ausgangslage	1
1.2	Ziel	2
1.3	Aktueller Forschungsstand	2
1.4	Vorgehen bei der Umsetzung	2
1.5	Vorläufige Gliederung der Abschlussarbeit	4
	Literatur	5

EXPOSE

1.1 AUSGANGSLAGE

Die Editions- und Forschungsstelle Frank Wedekind (EFFW) wurde 1987 in der Hochschule Darmstadt gegründet. Ihr Intention ist es, den lange vernachlässigten Autor der europäischen Moderne in die öffentliche Aufmerksamkeit zu bringen. Die Publikation der »Kritischen Studienausgabe der Werke Frank Wedekinds. Darmstädter Ausgabe« im Verlag Jürgen Häuser wurde 1994 direkt nach der Erschließung der Wedekind-Nachlässe in Aurau, Lenzburg und München begonnen und im Jahre 2013 abgeschlossen (8 Bände in 15 Teilbänden, jetzt in Wallstein Verlag). Die EFFW wurde im Sommer 2015 an die Johannes Gutenberg-Universität Mainz umgezogen.

Da Frank Wedekind heute zu einem der bahnbrechendsten Autoren der literarischen Moderne zählt, aber bisher sehr wenig erforscht wurde, soll sich diese nun ändern. Die nationalen und internationalen Korrespondenzen von und an Wedekind zeigen eine starke Vernetzung in der europäischen Avantgarde. Dies zeigt, dass die Wissenschaft um die Korrespondenzen von Wedekind eine immer größere Rolle spielen. Aktuell sind lediglich 710 der 3200 bekannten Korrespondenzstücke veröffentlicht worden.

Um diese zu verändern entstand das Projekt »Edition der Korrespondenz Frank Wedekind als Online-Volltextdatenbank« [Mar23], welches bei der EFFW angesiedelt ist und als Kooperationsprojekt an der Johannes Gutenberg-Universität Mainz, der Hochschule Darmstadt und der Fernuni Hagen umgesetzt wird und durch die Deutsche Forschungsgemeinschaft (Bonn) gefördert wird.

Hierbei werden sämtliche bislang bekannte Korrespondenz in die Online-Edition überführt. Diese Korrespondenz beinhaltet substantiell das literarhistorische und kulturgeschichtliche Wissen über die Kultur zwischen 1880 und 1918, indem das überlieferte Material zum einen transkribiert, editiert und editionswissenschaftlich kommentiert wird. Und zusätzlich durch Kommentar die den historischen Kontexten inhaltlich erschließen.

Hierfür entstand das Pilotprojekt der Online-Volltextdatenbank für Briefe von und an Frank Wedekind, welches 2015 als Beta-Version freigeschaltet wurde. Diese Projekt kann aktuell unter <http://briefedition.wedekind.h-da.de> eingesehen werden.

Die benutzerfreundliche Erfassung und Annotation der Briefe, ist eines der Hauptziele der konzeptionierten technischen Architektur. Dies ist der Grund, warum die Präsentation-, Recherche- und Erstellungsebene vollständig webbasiert umgesetzt wurde. Die Briefe selbst, werden im etablierten TEI-Format gespeichert. Dies muss von den Editoren und Editorinnen nicht selbst eingegeben werden, sondern kann über einen entstandenen WYSIWYG-Editor direkt eingegeben werden, welcher es bei der Speicherung in das

TEI-Format wandelt. Ebenfalls wurde hierbei auf eine modulare und unabhängige Architektur geachtet, wodurch die Komponenten im Nachgang auch von anderen Briefeditionen genutzt werden können.

1.2 ZIEL

Die aktuelle Umsetzung beinhaltet die bisher definierte Anforderungen komplett. Darunter fallen die Recherchemöglichkeiten, sowie auch die Eingabe und die Verarbeitung der Briefe. Ein größeres Problem hierbei ist die Performance der Oberfläche. Durch die langen Abfragedauern des Datenbestandes leidet die Akzeptanz der Anwendung.

Das Ziel der Arbeit ist es, die Abfragedauern zu verringern, wodurch die Performance der Oberfläche signifikant verbessert wird. Anhand der Quellen müssen hierfür die Abfragen ermittelt und bewertet werden. Die Abfragen mit den höchsten Zeitbedarf werden dann analysiert und optimiert.

Hierbei ist auch ein Vergleich mit anderen Technologien angedacht.

1.3 AKTUELLER FORSCHUNGSSTAND

1.4 VORGEHEN BEI DER UMSETZUNG

Zuerst werden alle Abfragen ermittelt und deren Performance in Abhängigkeit der Häufigkeit der Aufrufe und Datenmengen bestimmt. In der Auflistung werden dann die 5-7 Anfragen bestimmt, bei denen die größten Performance-Optimierungen möglich ist. Diese Abfragen werden dann genauer untersucht. Hierbei werden verschiedene Blickwinkel betrachtet um die Performance zu optimieren.

Darunter fallen zum einen die Cache-Algorithmen der JDBC-Engine, sowie auch die Einstellungen am Datenbanksystem. Hierbei ist noch ein besonderes Augenmerk auf die vorhandene Serverkonstellation mit zu beachten, da diese enormen Einfluss auf die Einstellungen bewirkt. Ebenso werden die Aufrufe im ganzen überprüft und untersucht um zu prüfen ob die Anfragen sich gegenseitig durch Transaktionen oder Locks sperren. Hierfür wird ebenfalls die interne Protokollierung der Aufruf aktiviert und dessen Ausgabe untersucht und analysiert.

Danach werden die Abfrage selbst untersucht und auf Optimierungen überprüft. Hierbei wird als erstes der Aufruf der Abfrage betrachtet. Dann wird die Abfrage selbst genauer untersucht. Dabei wird beachtet ob die Anfragen selbst viel Zeit für die Bearbeitung benötigen oder auf Ressourcen warten. Zum anderen wird geprüft ob durch gezielte Umstellung oder Einfügen von Zwischenergebnissen schnellere Abfragen möglich sind, wie es in der Abhandlung *Optimizing Iceberg Queries with Complex Joins* [WRY17] gezeigt wird. Zum Schluss werden noch die Abfragekriterien und die vorhanden, beziehungsweise genutzten, Indizierungen überprüft.

Als letztes wird noch die Art des Aufrufes betrachtet. Hierbei wird die Art und Weise der Aufrufe genauer betrachtet. Ob zum Beispiel eine vorhandene

Anfrage mehrfach verwendet wird und diese besser auf 2 ähnliche Abfrage aufgeteilt werden kann. Oder ob an den Stellen ein Paging eingebaut werden kann, um die Datenmengen zu reduzieren.

1.5 VORLÄUFIGE GLIEDERUNG DER ABSCHLUSSARBEIT

1. Einleitung
2. Grundlagen
3. Konzept
4. Performance-Untersuchung
 - 4.1 Einbau von Performancecounter
 - 4.2 Aktivieren von Performancecounter am Postgresql
 - 4.3 Laufzeitanalyse starten
5. Optimierung
 - 5.1 Anpassung der Konfiguration
 - 5.2 Veränderung der Abfragen
 - 5.3 Veränderung der Webseite
6. Evaluierung
 - 6.1 Erneute Laufzeitanalyse starten
 - 6.2 Vergleich der Ergebnisse vor und nach der Optimierung
7. Zusammenfassung
8. Ausblick

LITERATUR

- [Mar23] AK Prof. Dr. Ariane Martin. *Frank Wedekind | FB 05 - AK Prof. Dr. Ariane Martin*. 2023. URL: <https://www.martin.germanistik.uni-mainz.de/forschung/frank-wedekind/> (besucht am 24.09.2023).
- [WRY17] Brett Walenz, Sudeepa Roy und Jun Yang. "Optimizing Iceberg Queries with Complex Joins". In: *Proceedings of the 2017 ACM International Conference on Management of Data*. SIGMOD '17. Chicago, Illinois, USA: Association for Computing Machinery, 2017, 1243–1258. ISBN: 9781450341974. DOI: 10.1145/3035918.3064053. URL: <https://doi.org/10.1145/3035918.3064053>.