

Exposé

der Bachelorarbeit zum Thema

Analyse und Optimierung der Datenbankschicht des Content- Management-Systems TYPO3 unter besonderer Berücksichtigung von rekursiven Tabellenausdrücken

vorgelegt von

Adrian Schönfeld

dem **Fachbereich VI – Informatik und Medien**
der Berliner Hochschule für Technik Berlin
Studiengang: **Medieninformatik Online (B.Sc.)**

Tag der Abgabe 10. September 2024

Betreuer: Sebastian Kreideweiß, M.Sc. Technische Hochschule Brandenburg



Inhaltsverzeichnis

Abkürzungsverzeichnis	III
1 Problemstellung	1
2 Erkenntnisinteresse	2
3 Fragestellung	3
4 Ziele und Hypothesen.....	4
5 Methodik	5
6 Gliederungsentwurf	6
7 Vorläufiges Literaturverzeichnis.....	7
8 Zeitplan.....	8

Abkürzungsverzeichnis

CMS	Content-Management-Systeme
GSB	Government Site Builder
BHT	Berliner Hochschule für Technik
CTE	common table expression
DB	Datenbank

1 Problemstellung

Content-Management-Systeme (CMS) bilden heutzutage mit fast 70% die Basis aller Webanwendungen [W³Techs, 2024]. Mit ihnen ist es Betreibern von Webseiten möglich, ohne Programmierkenntnisse, Inhalte zu erstellen und zu verwalten.

Im Bereich der öffentlichen Verwaltung hat sich, in Deutschland, das CMS TYPO3 seit vielen Jahren etabliert [Kölbel, 2022]. Typo3 ist ein open source CMS, das seit 2001 entwickelt wird.

Zudem wird TYPO3 künftig die Basis des Government Site Builder (GSB) bilden. Der GSB ist die zentrale Content-Management-Lösung für die Webangebote der deutschen Bundesverwaltung [ITZ Bund, 2024]. Er basierte zunächst auf einem kommerziellen CMS, das im Jahr 2018 mit der GSB Version 10 durch open source Komponenten abgelöst wurde. Der GSB 11, mit TYPO3 als Basis, befindet sich aktuell in der Entwicklung.

Daten, die durch TYPO3 verwaltet werden, werden in einer Datenbank gespeichert. Die Datensätze weisen an mehreren Stellen hierarchische Strukturen auf. Diese Strukturen können dazu führen, dass die jeweilige CMS Anwendung schlecht skaliert, da diese Strukturen rekursiv aufgelöst werden müssen. Aktuell werden hierarchische Datenbankstrukturen durch TYPO3 selbst aufgelöst, d.h. das Ergebnis eines Datenbankzugriffs wird von TYPO3 verarbeitet und es erfolgt ein weiterer Datenbankzugriff auf die nächste Ebene. Je mehr Hierarchien existieren, desto mehr Datenbankzugriffe werden durchgeführt. Dadurch kann es zu merklichen Performanceeinschränkungen kommen.

Es ist davon auszugehen, dass TYPO3 Instanzen in Zukunft immer größere Datensätze verwalten werden. Ebenso ist davon auszugehen, dass die deutsche Bundesverwaltung große Datenmengen mit dem GSB 11, mit Typo3 als Basis, verwalten wird. Dies führt bei Anwendungen mit vielen hierarchischen Datenstrukturen bzw. Zugriffen auf diese Strukturen zu Performanceeinschränkungen und einer sinkenden Akzeptanz von TYPO3.

Es werden bereits an verschiedenen Stellen Gegenmaßnahmen zu den Performanceeinschränkungen ergriffen, wie z.B. Caching, aber es lassen sich nicht immer Wartezeiten aufgrund von Datenbankzugriffen vermeiden.

2 Erkenntnisinteresse

Ziel ist es, die Verwendung von hierarchischen Datenstrukturen im CMS TYPO3 zu identifizieren und deren Zugriff mittels recursive common table expressions (CTE) umzusetzen. Durch die Verwendung von recursive CTE soll der Aufwand zum Auflösen dieser Strukturen in die Datenbank verlagert werden. Dadurch erhält TYPO3 direkt das rekursiv aufgelöste Ergebnis von der Datenbank und muss dieses nicht selbst mit vielen einzelnen Datenbankabfragen realisieren.

Die Anzahl der Datenbankzugriffe soll dadurch deutlich verringert werden und es soll ein Performancegewinn erreicht werden.

Die Arbeit ist ein Baustein, um die Akzeptanz von TYPO3 als open source CMS zu steigern. Dadurch profitieren alle Projekte, die auf dem CMS aufbauen, wie z.B. der GSB 11.

3 Fragestellung

Hauptfragen

a) Datenbankzugriffe in TYPO3

- I. Wie werden Datenbankzugriffe durch TYPO3 realisiert?
- II. Welchen Methoden greifen auf hierarchische Datenstrukturen zu?
- III. Wie können die DB-Zugriffe in den Methoden durch recursive CTE realisiert werden?

b) Messung der Umsetzung

- I. Wie viele DB-Abfragen werden pro Seitenaufruf ausgelöst?
- II. Wie lässt sich das Ergebnis messen?

Nebenfragen

c) Datenbankoptimierung außerhalb des TYPO3 Quellcodes

- I. Welche Systemeinstellungen gibt es um den Datenbankserver zu optimieren?
- II. Welche Einstellungen gibt es um die Datenbankanwendung zu optimieren?
- III. Welches Datenbankschema liegt in Typo3 vor?
- IV. Wie lässt sich das Datenbankschema verbessern?

4 Ziele und Hypothesen

Im Rahmen der Bachelorarbeit soll die Datenbankschicht vom Content-management System TYPO3 optimiert werden. Im Quellcode sollen Methoden identifiziert werden, die auf hierarchische Strukturen in der TYPO3 Datenbank zugreifen. Zugriffe die über mehrere einzelne CTE erfolgen, sollen durch recursive CTE ersetzt werden.

Die Verwendung von recursive CTE soll beim Zugriff auf hierarchischen Datenbankstrukturen deutlich performanter sein als mehrere einzelne CTE. Es wird vermutet, dass sich durch Umstellung der betroffenen Methoden ein merklicher Performancegewinn einer TYPO3 Anwendung einstellt.

Ziel ist es, eine allgemeingültige Performancesteigerung zu erreichen, die sich auf die meisten TYPO3 Instanzen anwenden lässt in denen auf hierarchische Datensätze zugegriffen wird.

Es ist **nicht** das Ziel Hierarchien im Datenbestand einer TYPO3 Instanz aufzulösen um die Anzahl der CTE zu reduzieren. Ebenso sollen keine Einträge gelöscht werden um den Datenbestand zu verringern.

5 Methodik

Es wird eine Recherche durchgeführt um einen Wissensstand, bezugnehmend auf das Thema aufzubauen.

Es wird eine aktuelle TYPO3 Instanz lokal installiert, an der die Optimierungen vorgenommen werden. Die Anpassungen am TYPO3 Quellcode erfolgt in Abstimmung mit Entwicklern des TYPO3 Core. Je nach Umfang der Anpassungen erfolgt ggf. nur eine Teilumsetzung um den Rahmen der Bachelor-Arbeit nicht zu überschreiten.

Um ein Aussagekräftiges Ergebnis über den Erfolg der Anpassungen zu erhalten, ist es wünschenswert auf einen Testdatensatz einer großen TYPO3 Instanz zugreifen zu können. Die Performancemessung der Umsetzungen erfolgt ebenfalls mit lokal installierten Tools.

6 Gliederungsentwurf

- 1 Einleitung
 - 1.1 Motivation
 - 1.2 Zielsetzung
 - 1.3 Forschungsfragen
- 2 Grundlagen
 - 2.1 Aufbau der Typo3 Datenbankschicht
- 3 Methodik
 - 3.1 Systemumgebung
 - 3.2 Typo3 Instanz
 - 3.3 Entwicklungsumgebung
 - 3.4 Analysetools
 - 3.5 Testdatensätze
- 4 Erkenntnisse
 - 4.1 Allgemeine Optimierungen
 - 4.1.1 Serverebene
 - 4.1.2 Datenbankanwendung
 - 4.1.3 Datenbankschema
 - 4.2 Anpassung der Datenbankzugriffe
- 5. Analyse
- 6. Diskussion
- 7 Zusammenfassung der Ergebnisse
 - 7.1 Ausblick

7 Vorläufiges Literaturverzeichnis

- W³Techs. (2024). Usage Statistics and Market Share of Content Management Systems, September 2024, [online]
<https://www.theoceancleanup.com/updates/whales-likely-impacted-by-great-pacific-garbage-patch/> [abgerufen am 09.09.2024]
- Kölbel, J. (2022). Statistische Verteilung von Content-Management-Systemen im öffentlichen Bereich in Deutschland (Bachelorarbeit, Informatik und Medien). Berliner Hochschule für Technik, Berlin
- ITZ Bund. (2024). GSB 11 - Informationstechnikzentrum Bund, Informations Technik Zentrum Bund, [online] <https://produkt.gsb.bund.de/gsb11> [abgerufen am 09.09.2024]

8 Zeitplan

Folgender Zeitplan besteht. Wichtige Meilensteine sind fett hervorgehoben.

09.09.2024	Erstellung und Bereitstellung Exposé an Betreuer
10.09.2024	Überarbeitung und Abnahme Exposé
11.09.2024	Anfrage Zweitbetreuer
13.09.2024	Abgabe Anmeldung Bachelor-Arbeit im Prüfungsamt
01.10.2024	Beginn der Bearbeitungszeit (12 Wochen)
01.10.2024	Analyse der Datenbankschicht von Typo3
08.10.2024	Zugriffe auf hierarchische Datenbankstrukturen identifizieren
14.10.2024	Umsetzung von recursive CTE in den identifizierten Methoden
31.10.2024	Prüfung und Umsetzung - Optimierungen Server
04.11.2024	Prüfung und Umsetzung - Optimierungen Datenbankanwendung
08.11.2024	Prüfung und Umsetzung - Optimierungen Datenbankschema
12.11.2024	Umsetzung aller Optimierungen 100%
13.11.2024	Beginn Messung und Auswertung der Ergebnisse
18.11.2024	Messung und Auswertung der Ergebnisse 100%
08.12.2024	Alle Kapitel auf 100%
09.12.2024	Korrekturlesen durch Dritte
16.12.2024	Einarbeitung Korrekturen
23.12.2024	Einreichung Abschlussarbeit auf Dokumentenserver (OPUS) des Prüfungsamtes
24.12.2024	Ende der Bearbeitungszeit
31.12.2024	Anfertigung von Präsentationsunterlagen
07.01.2024	Phase Gutachtenerstellung der Betreuer
14.01.2024	Kolloquium BSc-Thesis