

Identifikation typischen Benutzerverhaltens in digitalen Studienformaten

Bachelorarbeit zur Erlangung des akademischen Grades Bachelor of Science
Berliner Hochschule für Technik · Fachbereich VI · Informatik und Medien

AUTOR

Werner Breitenstein
Matrikelnr.: 866059

BETREUER

Prof. Dr. Petra Sauer

GUTACHTER

Prof. Dr. Heike Ripphausen-Lipa

ABGABE

dd.mm.2022

Inhaltsverzeichnis

1	Einleitung	7
2	Grundlagen	8
2.1	Theorie	8
2.2	Technik	8
2.3	Datenbasis	9
2.3.1	Vorbereitung der Daten	9
2.3.2	Beschreibung der Daten	10
2.3.3	Visualisierung der Daten	13
3	Umsetzung	14
4	Ergebnisse	15
5	Fazit	16
6	Ausblick	17
	Literaturverzeichnis	18
	Erklärung zur Urheberschaft	19
	Inhalt des beigefügten Datenträgers	20

Abbildungsverzeichnis

Tabellenverzeichnis

1	Schema des Datenbestandes mit Erläuterungen	11
---	-------------------------------------------------------	----

Quellcodeverzeichnis

Zusammenfassung

...

Abstract

...

1 Einleitung

...

2 Grundlagen

Hier steht ein kurzer einleitender Text, der die nachfolgenden Kapitel inhaltlich skizziert.

...

2.1 Theorie

...

2.2 Technik

...

2.3 Datenbasis

Gegenstand der Untersuchungen zu dieser Arbeit ist ein durch die *Virtuelle Fachhochschule* (VFH) zur Verfügung gestellter anonymisierter Datenbestand aus dem Wintersemester 2021/2022¹. In diesem enthalten sind die Moodle-Nutzungsdaten von Studenten der *Berliner Hochschule für Technik* (BHT) sowie der *Alice Salomon Hochschule Berlin* (ASH) aus den folgenden vier Online-Studiengängen:

- Master-Studiengang Medieninformatik Online (MMIO)
- Bachelor-Studiengang Wirtschaftsingenieurwesen Online (BWIO)
- Bachelor-Studiengang Wirtschaftsinformatik Online (BWINF)
- Bachelor-Studiengang Soziale Arbeit Online (BSAO)

2.3.1 Vorbereitung der Daten

Um die Übersichtlichkeit der Daten und deren Untersuchung im Rahmen dieser Arbeit zu erleichtern, wurden diese in einem ersten Schritt aus der umfangreichen Datenbank des VFH-Moodle (Green, 2022) extrahiert und in einer einzigen Relation zusammengefasst.

Hierbei wurden Merkmale, die für diese Arbeit erwartungsgemäß keinen Mehrwert besitzen bereits eliminiert, während z. B. das Attribut *Studiengang* als neue Spalte in die Tabelle aufgenommen wurde, um die Zuordnung der verschiedenen Datensätze zu den jeweiligen Studiengängen unmittelbar erkennen zu können. Des weiteren wurden vorab die beiden Merkmale *course_module_type* und *instanceid* eingefügt, um auch deren Informationsgehalt sinnvoll nutzen zu können.

¹ Das gesamte Semester musste nach der SARS-CoV-2-Infektionsschutzmaßnahmenverordnung des Berliner Senates unter erhöhten Sicherheitsbedingungen stattfinden. Die Regelungen für das Lehr- und Prüfungsgeschehen wurden an der BHT infolgedessen wie folgt angepasst:

- keine Lehrveranstaltungen und Prüfungen in Präsenz
- keine Zählung des Semesters als Fachsemester
- keine Zählung von Prüfungsfehlversuchen

2.3.2 Beschreibung der Daten

Damit die Daten in einem beliebigen IT-Umfeld einfach weiterverarbeitet werden können, wurden sie im Anschluss an ihre Vorbereitung in einem für diesen Zweck typischen CSV-Format exportiert. Übergeben wurden die CSV-Daten schließlich als offene und komprimierte Textdateien in ASCII-Kodierung (Cerf, 1969), in der die Daten entgegen der üblichen Praxis jedoch nicht durch Kommata, sondern durch Semikola strukturiert waren.

Die freie Wahl eines Trennzeichens ist beim CSV-Format möglich, weil dieses nur allgemein beschreibt, wie die Tupel einer Relation und darin enthaltene Werte in der Regel interpretiert werden. Das Format definiert aber keinen verbindlichen Standard (Shafranovich, 2005), so dass die Daten entgegen ihrer Definition als Comma-Separated Values nicht zwingend nur durch Kommata zu strukturieren sind.

Der zur Verfügung gestellte Datenbestand umfasst insgesamt 288.152 Datensätze. Dabei handelt es sich genau betrachtet um eine spezifische Teilmenge von Loggings auf dem Moodle-Server der VFH, mit denen client- und serverseitige Aktionen fortlaufend protokolliert werden. Typische Aktionen, die so u. a. aufgezeichnet werden sind das Aufrufen eines Kursmoduls, das Starten eines Uploads, das Senden einer Nachricht oder auch die Bewertung einer Aufgabe.

Die folgende tabellarische Übersicht zeigt, welche Informationen in den Feldern der verschiedenen Merkmale des Datenbestandes enthalten sind und in welchen Beziehungen diese innerhalb der relationalen Datenbank des VFH-Moodle stehen; siehe hierzu auch die Moodle Entity Relationship Documentation (Moodle-ERD) (Green, 2022):

Merkmal	Information / Beziehung innerhalb des VFH-Moodle
courseid	Studienmodul, das im WS 2021/2022 belegt wurde. <i>Fremdschlüssel zur Identifikation eines bestimmten Studienmoduls in der Relation course.</i>
Studiengang	Studiengang, in dem aktuell studiert wird. <i>Frei gewählte Kennziffer zur eindeutigen Unterscheidung der Studiengänge; bedeutet keine Referenz auf eine andere Entität.</i>
userid	Kennzahl zur Identifikation des Benutzers. <i>Hash-Code zur Anonymisierung der Benutzerkennung, über die sonst ein Benutzer in der Relation user konkret referenziert werden kann.</i>
action	Interaktion, die im Moodle-System ausgeführt wurde. <i>Eindeutiger Wert (Fremdschlüssel?), der aus einer Relation außerhalb des eigentlichen VFH-Moodle bezogen wird.</i>
eventname	Mehrteiliger Bezeichner für das ausgelöste Event. <i>Ausgelöst durch eine Interaktion wird ein Bezeichner durch die drei Werte modulename, instance und eventtype der Relation event generiert und eingetragen.</i>
objecttable	Relation zur Verwaltung von Objekttabellen. <i>Abhängig von der Art des Kursmoduls und der Interaktion werden die durch Verwendung bestimmter Objekte tangierten Tabellen dokumentiert, z. B. assign_grades, course_modules oder forum_discussions</i>
objectid	Kennzahl zur Identifikation des verwendeten Objekts. <i>Fremdschlüssel zur Identifikation des durch die Interaktion tangierten Objekts in der zugehörigen Relation objecttable.</i>
timecreated	Zeitpunkt der ausgeführten Interaktion. <i>10-stelliger Unix Epoch Timestamp, der vergangene Sekunden seit Donnerstag, dem 01.01.1970, 00:00 Uhr UTC zählt.</i>
course_module_type	Typ des verwendeten Kursmoduls. <i>Zur Anreicherung des Informationsgehalts aus der Relation course_modules entnommener Bezeichner des Modultyps, z. B. assign, forum, label oder resource</i>
instanceid	Kennzahl zur Identifikation des Kursmodultyps. <i>Fremdschlüssel zur Identifikation des Kursmodultyps in der zugehörigen Relation course_modules.</i>

Tabelle 1: Schema des Datenbestandes mit Erläuterungen

Weitergehende Einblicke in die Art, den Umfang und die Struktur der zu untersuchenden CSV-Daten ergeben sich nach deren Import in eine MySQL-Datenbank durch erste einfache statistische SQL-Abfragen:

1. Abfrage: Art und Umfang der implementierten MySQL-Datenbank

```
mysql> SELECT table_schema, table_name, engine,
      (SELECT COUNT(*) FROM moodle_data) AS table_rows, table_collation
      FROM information_schema.tables WHERE table_name = "moodle_data";
```

table_schema	table_name	engine	table_rows	table_collation
vfh_moodle_ws21	moodle_data	InnoDB	288152	ascii_general_ci

2. Abfrage: Datentypen und Constraints der importierten CSV-Daten

```
mysql> DESCRIBE moodle_data;
```

Field	Type	Null	Key	Default	Extra
courseid	int(11)	YES		NULL	
Studiengang	varchar(11)	YES		NULL	
userid	varchar(32)	YES		NULL	
action	varchar(10)	YES		NULL	
eventname	varchar(57)	YES		NULL	
objecttable	varchar(27)	YES		NULL	
objectid	int(11)	YES		NULL	
timecreated	int(11)	YES		NULL	
course_module_type	varchar(18)	YES		NULL	
instanceid	int(11)	YES		NULL	

3. Abfrage: Summe aller Datensätze pro Studiengang

```
mysql> SELECT Studiengang, COUNT(Studiengang) AS "total number records"
      FROM moodle_data
      GROUP BY Studiengang;
```

Studiengang	total number records
	12051
1	14115
2	53706
3	111163
4	97117

4. Abfrage: Summe aller Kurse in den Studiengängen 1 bis 4

```
mysql> SELECT COUNT(DISTINCT courseid) AS "total number courses"
        FROM moodle_data
        WHERE Studiengang != "";
```

total number courses
108

5. Abfrage: Summe aller Benutzer in den Studiengängen 1 bis 4

```
mysql> SELECT COUNT(DISTINCT userid) AS "total number users"
        FROM moodle_data
        WHERE Studiengang != "";
```

total number users
107

2.3.3 Visualisierung der Daten

...

3 Umsetzung

...

4 Ergebnisse

...

5 **Fazit**

...

6 **Ausblick**

...

Literaturverzeichnis

Cerf, V. (1969, Oktober). *ASCII format for network interchange* (Nr. 20). RFC 20. RFC Editor. Zugriff auf <https://www.rfc-editor.org/info/rfc20> doi: 10.17487/RFC0020

Green, M. (2022). *The Moodle Database. Table and relationship documentation generated from moodle source code*. Zugriff am 2022-04-08 auf <https://www.examulator.com/er/>

Shafranovich, Y. (2005, Oktober). *Common Format and MIME Type for Comma-Separated Values (CSV) Files* (Nr. 4180). RFC 4180. RFC Editor. Zugriff auf <https://www.rfc-editor.org/info/rfc4180> doi: 10.17487/RFC4180

Erklärung zur Urheberschaft

Ich habe die Arbeit selbständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie alle Zitate und Übernahmen von fremden Aussagen kenntlich gemacht.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt.

Die vorgelegten Druckexemplare und die vorgelegte digitale Version dieser Arbeit sind vollkommen identisch.

Heidelberg, dd.mm.2022

Unterschrift

Inhalt des beigefügten Datenträgers

Verzeichnis / Beschreibung

/1_ ...

/2_ ...

/3_ ...
