

BDEA Aufgabe 1 - Analyse Amazon Video Games

Joel Staubach - 1820560

15.04.2023

Das Notebook analysiert Rezensionen der Plattform Amazon aus dem Zeitraum 1999 bis 2018. Es wurde für Apache Zeppelin 0.10.1 erstellt. Für die Verwendung muss lediglich die Datei *staubach_zeppelin.zpln* in eine lauffähige Apache Zeppelin [1] 0.10.1 Instanz importiert werden.

Für eine lokale Installation mit Anbindung an die Datenbank müssen Docker [2] und Docker-Compose [3] installiert werden. Die Entwicklung wurde mit der Docker-Version 23.0.1 und der Docker-Compose-Version 1.29.2 durchgeführt. Als Datenbank wird eine MongoDB [4] verwendet. Die Daten sind Rezensionen von Artikeln aus der Kategorie "Video Games" von Amazon [5]. Die Daten befinden sich fast alle im gültigen JSON-Format, was den Import in die MongoDB vereinfacht. Der Zeppelin-Container benötigt die Mongo-Shell und muss deshalb selbst gebaut werden. Außerdem wird der MongoDB-Container gebaut, in dem automatisch der Datensatz heruntergeladen wird, sowie über das Skript *format.py* in das passende Schema für die MongoDB überführt wird.

Zum Ausführen der lokalen Installation müssen folgende Schritte durchgeführt werden:

1. Das Repository <https://github.com/Lightningstream/bdea-a1> klonen oder die mitgelieferte ZIP-Datei *bdea-a1* extrahieren.
2. In einem Terminal im Ordner *bdea-a1* folgende Befehle nacheinander ausführen:
 - a. `docker-compose build` # Erstellt die Docker images
 - b. `docker-compose up -d` # Führt die images hoch
 - c. `docker exec -it bdea-mongo-amazon mongoimport --username root --password example --authenticationDatabase admin --db test --collection video_games --file video_games.json --jsonArray` # Importiert die Daten in die Datenbank
3. Die Webseite auf *localhost:8090* öffnen.
4. Oben rechts auf den Nutzernamen klicken -> *Interpreter* -> Nach *mongodb* suchen und folgende Werte eintragen sowie speichern:
 - a. *mongo.shell.path*: mongosh
 - b. *mongo.server.host*: mongo oder die Ip-Adresse des MongoDB Containers
 - c. *mongo.server.database*: test
 - d. *mongo.server.authenticationDatabase*: admin
 - e. *mongo.server.username*: root
 - f. *mongo.server.password*: example
5. Danach auf *Zeppelin* oben links klicken -> *import note* -> *Select JSON File/IPYNB File* -> *notebook_staubach.zpln* auswählen und öffnen.

Mit *Run all paragraphs* die einzelnen Paragraphen ausführen. Bei Fehlern bitte erneut die Interpreter-Konfiguration oder den Zustand der Docker-Container überprüfen.

Literaturverzeichnis

- [1] (2023). "Apache Zeppelin" [Online]. Verfügbar: <https://zeppelin.apache.org/> (besucht am 15/04/2023).
- [2] (2023). "Docker" [Online]. Verfügbar: <https://www.docker.com/> (besucht am 15/04/2023).
- [3] (2023). "Docker-Compose" [Online]. Verfügbar: <https://docs.docker.com/compose/> (besucht am 15/04/2023).
- [4] (2023). "MongoDB" [Online]. Verfügbar: <https://www.mongodb.com/de-de> (besucht am 15/04/2023).
- [5] Jianmo Ni und Jiacheng Li und Julian McAuley, Justifying recommendations using distantly-labeled reviews and fine-grained aspects. Empirical Methods in Natural Language Processing (EMNLP), 2019 . [Online]. Verfügbar: <https://cseweb.ucsd.edu/~jmcauley/pdfs/emnlp19a.pdf>
- [6] (2023). "Diablo 3 Launch" [Online]. Verfügbar: <https://www.gamespot.com/articles/how-diablo-3-went-from-disastrous-launch-to-a-hell-of-a-good-time/1100-6503426/> (besucht am 15/04/2023).
- [7] (2023) "Diablo 3 Amazon" [Online]. Verfügbar: <https://www.amazon.com/Diablo-III-pc/dp/B00178630A> (besucht am 15/04/2023)
- [8] (2023) "GameCube Controller" [Online]. Verfügbar: <https://www.amazon.de/GameCube-Controller-White-JP-Import/dp/B0017KIBAI> (besucht am 15/04/2023)