

## Simpsons Suchmaschine: Einleitung:

Für den Bau einer Suchmaschine im Rahmen des Moduls 10-201-2316 „Information Retrieval“ entschied sich unsere Gruppe für das Erstellen einer Suchmaschine über einen Datensatz, der sich mit der berühmten Zeichentrickserie „The Simpsons“ befasst. Diese Entscheidung wurde vorrangig von unserem persönlichen Interesse motiviert, da unsere Generation mit der gelben Familie aufgewachsen ist, allerdings ist auch die kulturelle Relevanz und Präsenz der Serie bedeutend genug, dass die Suchmaschine auch ein breiteres Publikum ansprechen könnte. Der Datensatz selbst wurde über die Website [kaggle.com](https://www.kaggle.com) gefunden und basiert auf den Skripten der einzelnen Folgen, speziell besteht er aus CSV-Dateien, die unter das aus der Vorlesung bekannte Konzept „1 File – Many Units“ fallen. Die CSV-Dateien werden also in diesem Falle in Dokumentenunits aufgeteilt. Dabei werden von uns vor allem die Dialoge in den einzelnen Szenen als Grundlage des Suchprozesses verwendet, sodass sowohl eine interpretative Suche, aber dank der Datenstruktur auch datenbankähnliche Abfragen möglich sind. Diese nehmen in diesem Projekt natürlich nur einen deutlich kleineren Teil der Anfragen ein. Bei der Implementierung der Suchmaschine haben wir uns als Grundlage für die Open Source Software Apache Lucene für C# entschieden.