

Datenverlust - Songlyrics

Patrick Zivkovic

May 19, 2020

Nicht alle Alben berücksichtigt

Durch das Limit = 50 Parameter der Funktion des get-artist-albums haben wir nicht alle Alben extrahiert. Mit dem Offset Parameter haben wir das korrigieren können und etwa 2500 weitere Matches hervorbringen können, damit erhöht sich die Größe unseres Datensatzes von 34116 auf 36932. (ursprünglich 57650 Lieder - also arbeiten wir mit etwa 64% des vollständigen Datensatzes)

Matching: Stringdist = 2

Hat bei 5994 Titeln eine Korrektur machen müssen. Meistens waren es bloß Fehler in der Zeichensetzung wie das folgende Beispiel zeigt

artist	original_songname	songname
ABBA	Ahe's My Kind Of Girl	She's My Kind Of Girl
ABBA	Gonna Sing You My Lovesong	Gonna Sing You My Love Song
ABBA	Hey Hey Helen	Hey, Hey Helen
ABBA	Knowing Me Knowing You	Knowing Me, Knowing You
ABBA	Money Money Money	Money, Money, Money

Probleme mit Anhängsel

Die meisten Lieder gingen uns verloren weil das Lied in Spotify nur mit einem Anhängsel wie etwa "Live", "Outtake", "Single" oder derartigem verfügbar war. Etwa kommt das Lied "Love Isn't Easy" von Abba in dieser Form in dem Original-Datensatz vor in Spotify jedoch nur folgendermaßen

```
## [1] "Love Isn't Easy (But It Sure Is Hard Enough)"
```

Ein weiteres Beispiel zeigt das Lied von den Beatles "All I've Got To Do"

```
## [1] "All I've Got To Do - Remastered 2009"
```

Verfügbarkeit

Die restlichen Lieder wurden nicht gematcht, weil sie tatsächlich nicht in der Spotify Datenbank verfügbar sind. Zum Beispiel einige Lieder von ABBA ("Burning My Bridges" oder "Dream World")