# Dokumentation Genius Crawler

(genaue Dokumentation der Funktionen im Code)

* Aufgrund unserer Repositorystruktur befinden sich das Skript, sowie alle zugehörigen Dateien, im python Ordner unter Punkt 2.5
* Ziel dieses Skriptes ist es die Songtexte der vom API Skript gefundenen Songs von [www.genius.com](http://www.genius.com) herunterzuladen und in einer CSV zu speichern
* Da für unsere Analyse nur Songs relevant sind, zu denen wir auch die passenden Audio Features von Spotify haben, nutzt dieses Skript die „hitlist\_updated.csv“ als Inputdatei
* Songs zu denen keine Spotify ID vorhanden ist werden daher ignoriert
* Aufgrund der hohen Fehlerquote des Skriptes (ca.50%), bedingt zum einen durch die Inputdatei (Schreibfehler, Umgang mit Sonderzeichen, Songs mit unterschiedlichen Interpreten) und zum anderen durch Genius (inkonsistente URL-Struktur) wird neben der eigentlichen Outputdatei zusätzlich noch eine error\_lyrics.csv erzeugt, welche den Songtitel, Interpreten und die Abfrage-URL an Genius speichert um eventuell Muster in den Fehlern erkennen und anschließend für alle betroffenen Songs beheben zu können (die Behebung von einzelnen Fehlern, wie z.B. Tippfehlern, ist aufgrund des zu großen Aufwandes kein relevanter Verwendungszweck dieser Datei).