

## Übung 3 - Elasticsearch mit Movie-Datensatz

### 1. Kaggle-Datensatz herunterladen

1. Besuchen Sie die [Kaggle-Seite des Movie-Datensatzes](#).
2. Laden Sie den Datensatz herunter und entpacken Sie ihn.
3. Prüfen Sie die Struktur der CSV-Dateien, um sich mit den verfügbaren Daten vertraut zu machen.

---

### 2. Ziel-Datenstruktur und Mapping

#### 1. Ziel-Datenstruktur festlegen:

- Überlegen Sie, welche Felder aus dem Datensatz für Ihre Anwendung relevant sind (z. B. `title`, `genres`, `release_date`, `overview`, `vote_average`, `popularity`).
- Erstellen Sie ein Mapping für diese Felder:
  - `title` : `text` mit `keyword` Sub-Field.
  - `genres` : `nested` Typ mit `keyword`.
  - `release_date` : `date`.
  - `overview` : `text`.
  - `vote_average` : `float`.
  - `popularity` : `float`.

#### 2. Mapping implementieren:

- Erstellen Sie ein Elasticsearch-Mapping basierend auf Ihrer Ziel-Datenstruktur.
- Nutzen Sie die [Mapping API](#).

---

### 3. Analyzer definieren

#### 1. Analyzer für den Anwendungsfall:

- Definieren Sie einen benutzerdefinierten Analyzer für das Feld `title`, um die Suchfunktion zu optimieren.
- Fügen Sie einen weiteren Analyzer für das Feld `overview` hinzu, der einen `synonym` Filter enthält.

## 2. Analyzer implementieren:

- Verwenden Sie die [Analysis API](#), um Ihre Analyzer zu konfigurieren und zu testen.

////////////////////////////////////

## 4. Datensatz indexieren

### 1. Python-Script vorbereiten:

- Nutzen Sie eine Bibliothek wie `elasticsearch` oder `elasticsearch-py`.
- Laden Sie die Daten aus der CSV-Datei, verarbeiten Sie sie und indexieren Sie die Dokumente in den Elasticsearch-Cluster.
- Achten Sie auf folgende Schritte:
  - Überprüfen Sie, ob alle Felder dem Mapping entsprechen.
  - Konvertieren Sie Datumsangaben und Listen korrekt.
  - Indexieren Sie in Batches, um Performance-Probleme zu vermeiden.

### 2. Indexierung durchführen:

- Starten Sie das Script und stellen Sie sicher, dass alle Daten erfolgreich in Elasticsearch gespeichert wurden.

////////////////////////////////////

## 5. Abfragen und Analyse

### 1. Experimentieren mit Anfragen:

- Stellen Sie folgende Beispielanfragen:
  - Alle Filme eines bestimmten Genres.
  - Freitextsuche in den Feldern `title` und `overview`.

### 2. Probleme analysieren:

- Dokumentieren Sie Herausforderungen und Einschränkungen, die bei der Indexierung oder Abfrage auftreten (z. B. Suchgenauigkeit, Performance).
- Schlagen Sie Lösungen oder Verbesserungen vor (z. B. Anpassung der Analyzer, Änderungen im Mapping).