

Inhaltsverzeichnis

1	LastFM	2
1.1	Hintergrund	2
1.2	Geschichtlicher Überblick	2
2	Systemarchitektur	3
2.1	Überblick der LastFM API-Features	3
2.2	API Features	3
2.3	Tools und Frameworks	3
3	Auswertung der Daten	4
3.1	Datenquellen	4
3.2	Feature Extraction	4
3.3	Ähnlichkeitsmessungen	4
3.3.1	Cosine-Similarity	4
3.4	Visualisierung	4
4	Ergebnisse	5
5	Referenzen	6

1 LastFM

1.1 Hintergrund

LastFM ist heute ein personalisiertes Online-Radio - abspielbar mit einer eigenen, proprietären Software. Dieses ermöglicht es den Usern einen auf ihren Geschmack zugeschnittenen Musikstream abzuspielen. Angegeben kann der momentan gewünschte Musikstil via Tags - entweder auf einen Künstler bezogen oder auf einen Musikstil bzw. Genre. Ebenso ist es möglich über Plug-Ins das Abspielverhalten von Usern in anderen Audio-Playern zu beobachten um damit die Erstellung des Nutzerprofils zu unterstützen. Weiters ist *LastFM* auch ein großes soziales Netzwerk, dass vor allem auf das zusammenbringen von Menschen mit ähnlichem Musikgeschmack ausgelegt ist. Die User der Community tragen zur Klassifikation des Musikbestandes durch tagging, Wiki-Beträge und einfach ihr Hörverhalten bei. Mehr dazu in Abschnitt 2.2.

1.2 Geschichtlicher Überblick

LastFM wurde in den späten 90er Jahren als Online-Musiklabel gegründet und bot als Feature die Möglichkeit, sich durch die Art der konsumierten Musikstücke ein Profil über seinen Musikgeschmack zu erstellen. Die Firma *Audioscrobbler*, hervorgegangen aus einem Informatikprojekt, hatte sehr ähnliche Ideen, worauf hin beide Unternehmen sehr eng zusammenarbeiteten, bis sie schließlich im Jahr 2005 fusionierten und unter dem Namen *LastFM* die Funktionen von beiden Technologien zur Verfügung stellen. 2007 wurde *LastFM* um den Preis von 280 Millionen Dollar an das US-amerikanische Medienunternehmen CBS verkauft. Diese Übernahme gehört damit zu den größten dotcom Aquisitionen bisher. Seit April 2009 ist die Benutzung des Radiodienstes nur mehr in den USA, Großbritannien und Deutschland kostenlos möglich. In anderen Ländern muss ein Entgelt von 3 Euro für die monatliche Nutzung erbracht werden.

2 Systemarchitektur

2.1 Überblick der LastFM API-Features

2.2 API Features

Hauptquelle für Informationen über Musikstücke und Künstler sind die User der Community. Mittels tagging oder einfach dem übermitteln der Information, welche Musikstücke man konsumiert (dem sogenannten *scrobbling*), wird ein riesiger Metadatenpool ermittelt. Auf dieser Basis werden die Künstler und Bands Genres zugeordnet, sodass Ähnlichkeiten zwischen diesen ermittelt werden können. Weiters kann man über das Online-Portal auch eine Menge Hintergrundinformationen und aktuelle Auftritte (künstler- oder regionsbezogen) erfahren. Die meisten dieser Infos lassen sich direkt über die Webseite, oder eine Webserice-API abrufen. Die API ist zu denselben Bedingungen verfügbar wie die Nutzung des Radiodienstes, d.h. außerhalb US, UK und DE muss man einen geringes monatliches Entgelt entrichten. Nur wenn man die API in großem Standard nutzen will sind eigene Nutzungsbedingungen mit den Betreibern auszuverhandeln. Technisch basiert die Web-API auf dem REST-Standard (XML), Abfragen sind somit unkompliziert via HTTP möglich. Unter anderem lassen sich Künstlerdaten (Bio, Alben, Events, Top-Songs, Tags, ...), Geodaten (Events in der Umgebung, beliebteste Künstler, ...), Userdaten (Top-Artists, Friends, Neighbours, ...) und viele mehr. Für eine Dokumentation mit Anwendungsbeispielen sei auf <http://www.lastfm.de/api/intro> verwiesen.

Es existieren eine Reihe von Wrapper-APIs von Dritten für unterschiedliche Technologien, damit man die Web-API auch unkompliziert in anderen Anwendungen nutzen kann. Die Qualität dieser Implementierungen ist unterschiedlich und es kann durchaus sein dass diese nicht am aktuellen Stand der Web-API sind.

2.3 Tools und Frameworks

3 Auswertung der Daten

3.1 Datenquellen

3.2 Feature Extraction

3.3 Ähnlichkeitsmessungen

3.3.1 Cosine-Similarity

Ein Maß zur Bestimmung der Ähnlichkeit der Künstler in der LastFM Datenbank ist die Auswertung der Tags und ein Vergleich auf deren Übereinstimmung. Dazu wurden die Tags der Künstler mit der API ausgelesen und anschließend die Ähnlichkeit mit der aus der LV bekannten Formel der Cosine Similarity bestimmt:

$$sim(a, b) = \frac{a * b}{|a| * |b|} \quad (1)$$

Die API gibt auf Anfrage die Tags eines Künstlers/Band mit einer Gewichtung zwischen 100 und 1 zurück. Zwei kleine Probleme traten hier aber auf: Zum einen werden auf jeden Fall eine Menge Tags zurückgegeben. Nach ein paar Versuchen wurde aber klar, dass die Tags mit Gewicht 1 nur zufälliger Natur sind und in die Berechnung nicht mit einzubeziehen sind (Mozart hat die Attribute *psychedelic* bzw. *Heavy-Metal* eher nicht verdient). Weiters lieferte unsere Wrapper-API für Java leider nur die Tags, aber ohne Gewichtung. Die API musste daher geringfügig modifiziert werden.

3.4 Visualisierung

4 Ergebnisse

5 Referenzen