



# Datenbanksysteme

## Projekt: 3. Iteration - Data Mining und Visualisierung

Freie Universität Berlin  
Fachbereich Mathematik und Informatik, Institut für Informatik  
Arbeitsgruppe Datenbanken und Informationssysteme  
Prof. Dr. Agnès Voisard, Nicolas Lehmann  
Sommersemester 2017

**Abgabe** (Online im KVV) bis spätestens  
**Freitag, den 07.07.2017, 23:59 Uhr.**

Aufgabe	1	2	$\Sigma$
Punkte	50	50	100

*Hinweise:*

*Die Projekt-Iteration soll in Dreiergruppen bearbeitet werden und muss zu mindestens 70% erfüllt werden.*

*Stellen Sie Ihre Ergebnisse in Form einer 5 bis 10 minütigen Präsentation im Tutorium vor.*

## 1. Aufgabe: Data Mining: Clustering

1. Entwerfen Sie eine geeignete Metrik für die Ähnlichkeit von "Hashtags".
2. Verwenden Sie Ihre selbst definierte Metrik um ähnliche "Hashtags" zu identifizieren: Finden Sie alle Cluster von ähnlichen "Hashtags". Verwenden Sie hierzu den "k-means Algorithmus". Schreiben Sie ein Programm, das die Cluster berechnet.
3. Visualisieren Sie anschließend Ihre Lösung.

## 2. Aufgabe: Datenvisualisierung

1. Laden Sie sich die Javascript-Bibliothek "SigmaJS" herunter (erhältlich auf <http://sigmajavascript.org/>).
2. Verwenden Sie die Javascript-Bibliothek "SigmaJS" um ein "Hashtag"-Netzwerk zu visualisieren.

Die Knoten des Graphen sollen "Hashtags" sein. Zwischen zwei Knoten ("Hashtags") soll genau dann eine Kante existieren, wenn diese gemeinsam aufgetreten sind.

3. Finden Sie eine geeignete Visualisierung für Ihr "Hashtag"-Netzwerk, um die Ähnlichkeit von Hashtags ausdrücken zu können.
4. Visualisieren Sie die Häufigkeit des Auftretens aller "Hashtags" im Laufe der Zeit mit Balkendiagrammen entlang einer Zeitachse. Die kleinste Einheit entlang Ihrer Zeitachse soll ein Tag sein.
5. Visualisieren Sie die Häufigkeit des Auftretens eines auswählbaren "Hashtags" im Laufe der Zeit mit Balkendiagrammen entlang einer Zeitachse.