



# Datenbanksysteme

## Projekt: 1. Iteration - Modellierung

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

**Abgabe** bis spätestens **Freitag**,  
den **12.05.2017, 14:00 Uhr**.

Aufgabe	1	2	3	4	5	$\Sigma$
Punkte	20	10	30	30	10	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: Projektdokumentation

1. Erstellen Sie eine geeignete Projektdokumentation in LaTeX für Ihr Projekt (bspw. in "Overleaf" zur kollaborativen Nutzung).
2. Dokumentieren Sie alle wichtigen Schritte in Ihrem Projekt. Beginnen Sie mit dem Projektziel und einer kurzen Vorstellung des Teams.

## 2. Aufgabe: Explorative Datenanalyse

1. Laden Sie sich den Datensatz "*american-election-tweets.csv*" oder "*american-election-tweets.xlsx*" aus dem KVV herunter.
2. Schauen Sie sich den Datensatz an und versuchen Sie die Zusammenhänge im Datensatz zu verstehen. Der Datensatz enthält mehr Daten als für das Projektziel relevant sind.
3. Beschreiben Sie den Datensatz.

## 3. Aufgabe: ER-Modellierung

1. Die finale Projekt-Web-Anwendung soll ein Netzwerk aller "*Hashtags*" aus dem Datensatz "*american-election-tweets*" grafisch darstellen und benutzerdefinierte Anfragen auswerten können. Typische Abfragen an die finale Web-Anwendung könnten sein:
  - Welche "Tweets" waren am Wichtigsten?
  - Welche "Hashtags" wurden am meisten verwendet?
  - Wann traten insgesamt am meisten "Hashtags" auf?
  - Welche Hashtags treten besonders häufig gemeinsam auf?
  - Wie hat sich die Häufigkeit der Verwendung eines speziellen "Hashtags" über die Zeit entwickelt?

Es sollen die absoluten Häufigkeiten aller "*Hashtags*" erfasst werden. Zusätzlich sollen die Häufigkeiten des paarweisen Auftretens von zwei "*Hashtags*" erfasst werden. Werden beispielsweise in einem "*Tweet*" drei "*Hashtags*" #A, #B und #C genannt, sollen die Häufigkeiten der

Paare (#A,#B), (#A,#C), (#B,#C) erfasst werden. Des weiteren soll erfasst werden, wann ein "*Hashtag*" in einem "*Tweet*" aufgetreten ist und wie oft.

Erstellen Sie ein geeignetes "Entity Relationship Model".

2. Fügen Sie das erstellt ER-Modell Ihrer Projektdokumentation hinzu.
3. Schreiben Sie eine kurze Erklärung zu ihrem ER-Modell, in dem Sie Ihre wichtigsten Entscheidungen begründen.

## 4. Aufgabe: Relationales Modell

1. Erstellen Sie zu Ihrem ER-Modell, das dazu passende relationale Modell.
2. Fügen Sie das erstellt relationale Modell Ihrer Projektdokumentation hinzu.
3. Schreiben Sie eine kurze Erklärung zu Ihrem relationalen Modell, in dem Sie Ihre wichtigsten Entscheidungen begründen.

## 5. Aufgabe: Datenbank erstellen

1. Erstellen Sie eine Datenbank "Election".