

HashtagHunter



Hauptseminar:
Data Science 2016
Ludwig-Maximilians-Universität München

Julian Kolarz, Florian Wirth, Michael Spitzer



- Hashtags:
 - Abgeschlossene Informationseinheit
 - Durch User vergeben

- Geodaten:
 - Ursprung
 - Lokale Verbreitung
 - Raum-zeitliche Dynamik

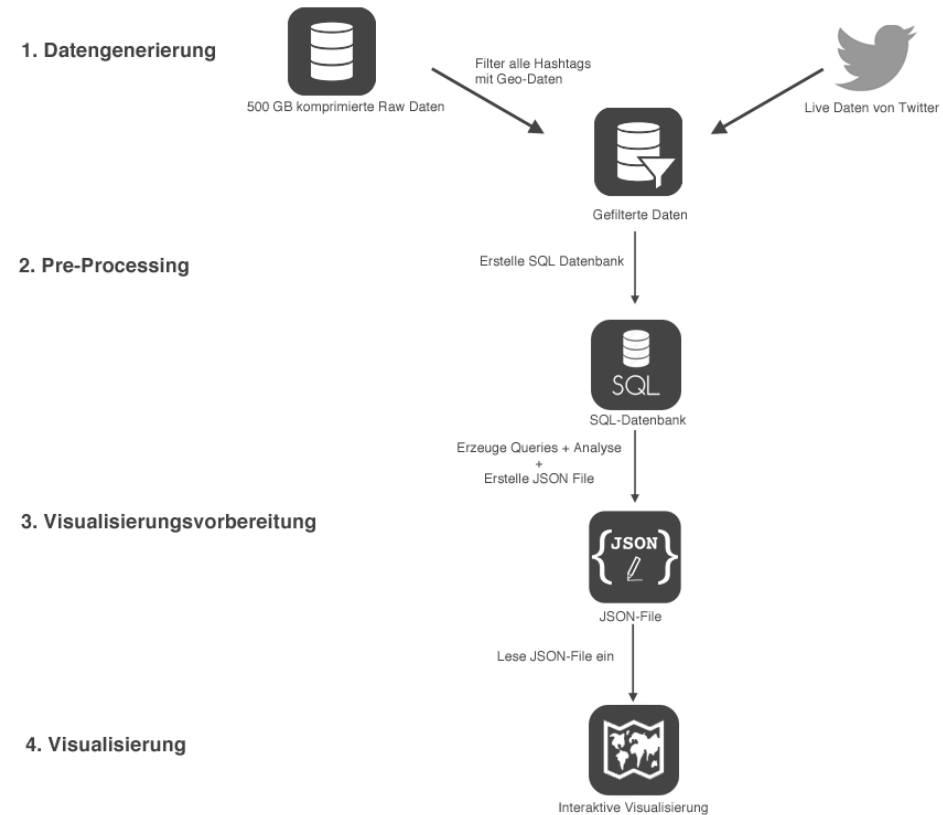
=> Visualisierung auf einer Karte



Pre-processing

- ## Auswahl betrachteter Hashtags über Heuristik:
- Unregelmäßige Ereignisse VS gleichmäßiges Rauschen
 - Ermittlung über Varianz in der Anzahl von Tweets pro Tag

Visualisierungs-Pipeline





Visualisierung



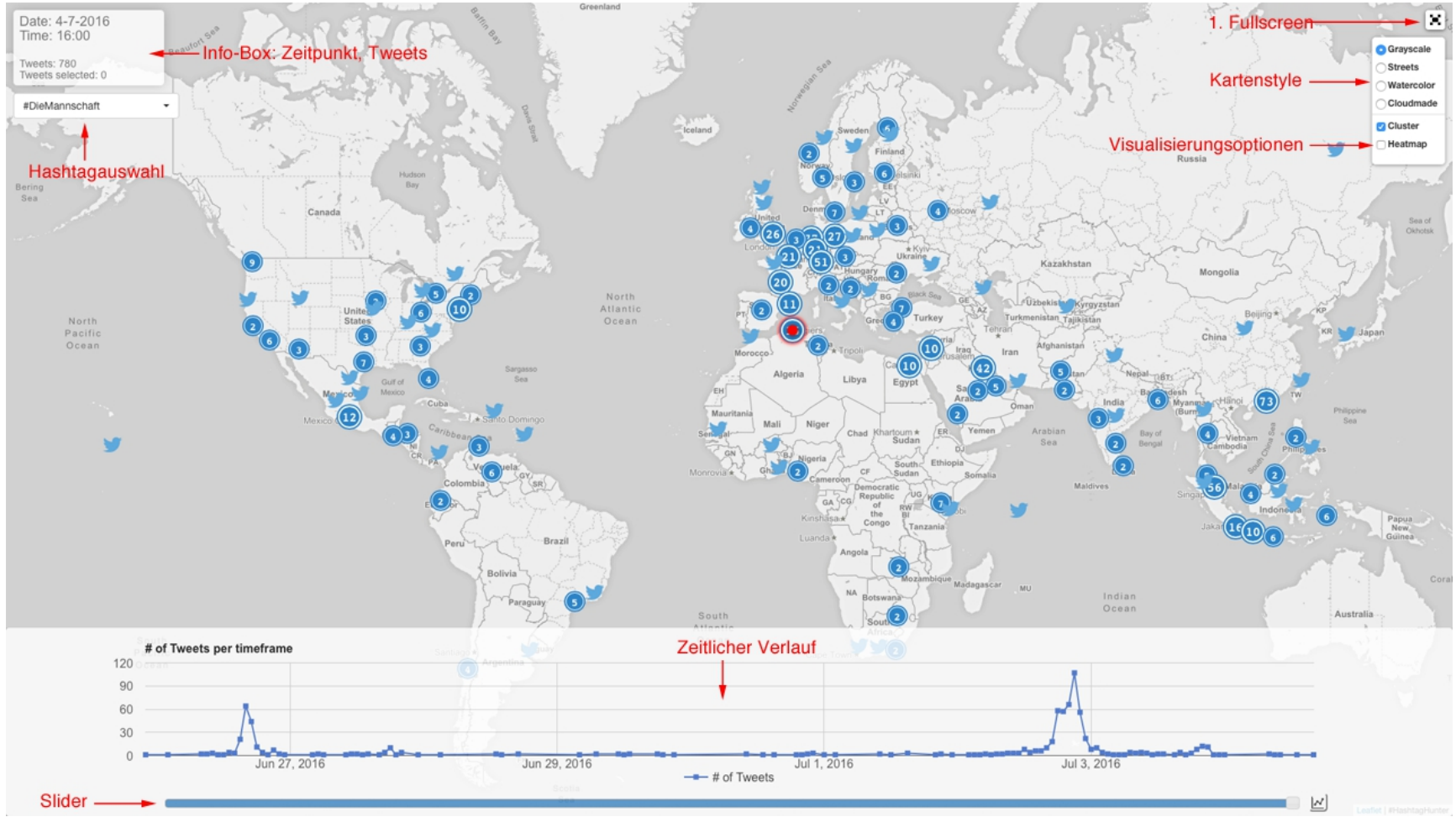
Features:

- ClusterMap
- HeatMap
- verschiedene Kartenstyles
- Time-Slider
- Zeit-Graphen
- Fullscreen-Modus
- ...



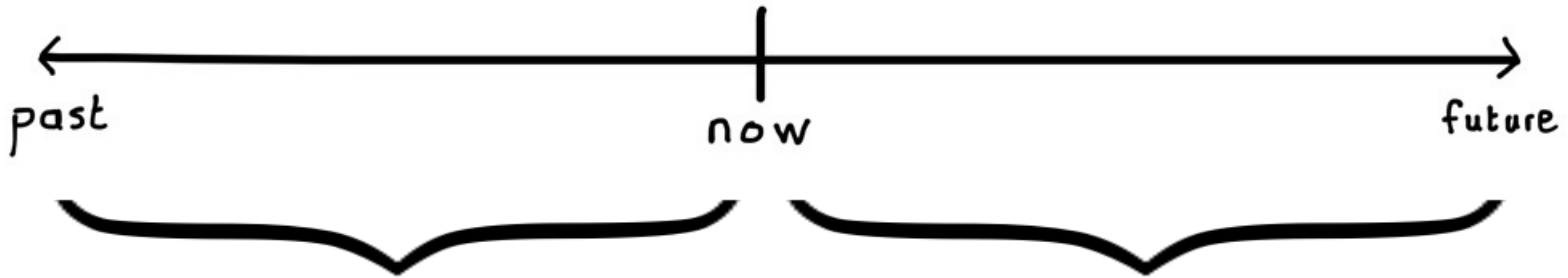
Problem:

- sehr große Menge an Geo-Daten (200k +) auf einer interaktiven Karte zu visualisieren





Live-Betrieb



Use-Case #1:



Use-Case #2:





Use-Case #1:

Nachträgliche Analyse vergangener Daten

- Verwendung der REST API oder eigenem Datensatz
- Filtern nach gewünschtem Hashtag
- Speichern und Weiterverarbeiten der Daten

Probleme:

- Geringe Speicherdauer der Tweets
- Gespeicherte Tweets sind vorselektiert
- Limitierte Anzahl an Queries

REST API

Filtere nach Hashtag
+ Speichern

erzeuge Json



lese Daten ein



Use-Case #2:

Aufzeichnung kommender Tweets

- Verwendung der Streaming API
- Filtern nach gewünschtem Hashtag
- Speichern und Weiterverarbeiten der Daten

Problem:

- Hardware / Skript muss kontinuierlich aktiv sein





Future Work



- Implementierung Live-Betrieb
- Hinzufügen neuer Features (Sentiment-Analysis, ...)
- ...

Live-Demo

<http://www.cip.ifi.lmu.de/~kolarz/demo.html>

Happy Hunting

