

Projektseminar Angewandte Informationswissenschaft 08.08.16 – 11.08.16

Thorsten Brückner

Datenanalyse mit Hilfe von Twitter API, Python und Networkx zum Thema: Release von World of Warcraft – Legion am 30.08.16

Bereits fertig (Zusammenarbeit mit Raphael Katschke):

- Twitter-Crawler in Python auf Basis der Twitter-API programmieren.
 1. Twitter API-Key erstellen (Access Token)
 2. Crawler in Python programmieren
 3. Python OAuth2/Tweepy library benutzen
 4. Accessing Data

To Do:

- Tweets zum Thema WoW-Legion erkennen und extrahieren über den Zeitraum von XX.XX.XX bis zum Release am 30.08.2016.
 - Mit Hilfe von GEO Tags, die Herkunftsländer von den Tweets analysieren und mit Networkx in Python grafisch darstellen.
 1. Key-attributes/fields wählen
 - a. Text
 - b. Geo-Tag
 - i. Coordinates
 - c. ID
 2. Tweets fetchen
 3. Daten in CSV speichern
 4. Daten in Python importieren
 5. Visualisierung mit Networkx & Gephi umsetzen
 - a. Verschiedene Länder anhand von GEO-Tags unterscheiden und anzeigen
 - i. Filtern von Geo-Tag/Coordinates
 - ii. Filtern von Tweets ohne Geo Tags
 1. Bekommt separate Notes/Edges
 - iii. Benutzen von „Complete Graph“ and „Bipartite Graph“
 - iv. Mögliche Graph Layouts (Entscheidung fällt mit Ergebnissen):
 1. Open Ord für „besser unterscheidbare Cluster“
 2. Yifan Hu für „dynamisches Map Clustern“
 3. Force Atlas für „Galaktisches Clustern“
 4. Fruchterman Rheingold für „Mass-Partikel Clustern“
 - v. Allgemeinen und Twitter Graphen generieren
 1. Hinzufügen von Notes und Edges anhand von Geo-Tags/Coordinates/ID
- Ebenfalls soll mit Hilfe von NLTK die Sprache der Tweets analysiert und über Networkx dargestellt werden.

1. Stoppwortliste
 - a. Englisch
 - b. Deutsch
 - c. Französisch
 - d. Spanisch
2. Langdetect Lib für Python
 - a. Wörterbücher erweitern mit nötigen „Game und Szenespezifischen Wörtern/Tags in erforderlichen Sprachen“
 - b. Text + ID (duplicates vermeiden)
 - i. Funktion schreiben, die „Text“ als String entgegen nimmt und mit Hilfe von der Langdetect Lib die Sprache ausgibt/identifiziert
 - ii. Anhand der Sprache wird danach zwischen verschiedenen „Dictionaries“ unterschieden und eingetragen
 - iii. Nicht erkannte Texte manuell überprüfen und Sprache markieren
3. Matplotlib
 - a. Als Visualisierungs-Lib benutzen um Sprachen in Diagrammen anzuzeigen
 - i. Importieren von matplotlib und numpy
 - ii. Generieren von „candidates column“
 - iii. „Bar plot“ erstellen
 1. Anpassen vom plot
 2. Labels hinzufügen
4. Über PHP die Ergebnisse auf Homepage anzeigen lassen