

## Projektseminar IM :: Wintersemester 2013/14

### Projekt 2: Codename „Larry Bird“

Das erste Projekt in diesem Seminar trägt den Codenamen „Larry Bird“. Auch dieses Projekt besteht sowohl aus Python-Programmierung und Dokumentation mit Hilfe von Sphinx. Ab diesem Projekt findet keine scharfe Aufgabentrennung zwischen „Programmieraufgabe“ und „Dokumentation“ statt, sondern es wird vorausgesetzt, dass eine Dokumentation des Projektes selbstverständlich dazugehört. Dazu gehört eine Anleitung für das Programm, die Beschreibung des Quellcodes sowie weitere Dokumentationsanteile (sofern gefordert).

### Projektaufgabe

Mit Hilfe von *tweepy* soll über die Twitter-API auf Twitter-Daten zugegriffen werden. Dazu werden zunächst zwei Twitter-Accounts ausgewählt. Jeder dieser Twitter-Accounts soll insgesamt mehr als 3200 Tweets haben. Die Personen/Institutionen/etc., welche hinter den beiden Twitter-Accounts stehen, sollen aus dem gleichen „Bereich“ kommen, also z.B. zwei Musiker/innen, zwei Schauspieler/innen, etc. Die Wahl des Bereichs/Genres ist euch völlig frei überlassen.

Die mit Hilfe von *tweepy* über die Timeline der Twitter-Accounts geholten Tweets werden in einer zur weiteren Verarbeitung geeigneten Struktur gespeichert. Mit Hilfe dieser Daten sollen Auswertungen vorgenommen, die sich auf Wörter und Hashtags in den Tweets sowie die generelle Menge abgesetzter Tweets beziehen.

Folgende Punkte müssen dabei berücksichtigt werden:

- Herunterladen der Tweets mit Hilfe von *tweepy* und Speichern in einer geeigneten Struktur
- Exportieren der Daten in eine externe Datei (in einem geeigneten Datenformat), welche zu einem späteren Zeitpunkt und bei allen Personen im Team wieder eingelesen werden kann
- Analyse der Tweets auf:
  - Anzahl je Jahr, Monat, Tag, etc. (Analyse mit Hilfe des *datetime*-Moduls)
  - Menge und Häufigkeit der verwendeten Wörter
  - Menge und Häufigkeit der verwendeten Hashtags
  - Menge und Häufigkeit der verwendeten Twitter-Clients
- Darstellung dieser Werte jeweils als Diagramme, die mit Hilfe von *matplotlib* erstellt wurden (vorab in Python oder direkt in Sphinx)
- Verwendung geeigneter Diagramme für die jeweilige Darstellung
  - z.B. Balkendiagramm für die Anzahl von Tweets pro Monat, in dem die Werte für beide Twitter-Accounts angezeigt werden
- Festlegen eines sinnvollen Grenzwertes, bis zu dem die Daten angezeigt werden
  - Diagramme sollen übersichtlich bleiben (z.B. Top 50, etc.)

## Projektseminar IM :: Wintersemester 2013/14

### Dokumentation des Tools

Die Dokumentation erfolgt erneut mit „Sphinx“. Die Grundlagen der Arbeit mit Sphinx sollten bereits bekannt sein. Neue Elemente in diesem Projekt sind das direkte Einbinden von *matplotlib*-Diagramm sowie die Erstellung von Tabellen.

Folgende Ziele sind zu erreichen:

- Dokumentation des Tools mittels Sphinx
  - Einleitung und Text entsprechend einem Handbuch
  - Beschreibung der Twitter-Accounts sowie Anzeige weiterer Daten, die zu den Twitter-Accounts über *tweepy* abgerufen werden können
    - die Auswahl der zu dokumentierenden Daten ist euch überlassen
    - Anzeige der Daten zu den Twitter-Accounts in Form einer Tabelle(!)
  - Verfassen eines Textes, in dem die Auswertungsergebnisse kurz dargestellt und miteinander verglichen werden
    - treten Ausreißer in den Daten auf, können Vermutungen für die Gründe angegeben werden
  - Verfassen eines kurzen Textes (~ 300 Wörter): Was ist ein API und wozu ist das gut?
    - Vorteile von Diensten mit API gegenüber Diensten ohne API
- Export der Sphinx-Dokumentation in HTML- und PDF-Format
  - PDF: Generierung von LaTeX, was mittels LaTeX-Umgebung (muss ebenfalls installiert werden) zu PDF verarbeitet wird; keine selbst erstellten PDF-Dateien

### Abgabe des Projekts

Die genauen Abgabemodalitäten werden im weiteren Verlauf der Veranstaltung rechtzeitig vor dem Abgabetermin (19.11.2013 23:59, CET) angegeben und ebenfalls online veröffentlicht.

### Fragen oder Probleme?

Sofern Fragen oder Probleme auftreten und diese nicht innerhalb des Teams zu klären sind, meldet Euch einfach im Forum, in den Übungsgruppen oder per E-Mail / persönlich (in dieser Reihenfolge) bei Eurem jeweiligen Dozierenden (Jens / Tobi) frühzeitig.

Frühzeitig bedeutet dabei, dass (grundsätzliche) Fragen zum Projekt möglichst nicht erst am Abgabetag gestellt werden.