

# Projektseminar I4 - Angewandte Informationswissenschaft

## Projekt | Projektplan

<b>Thematischer Kontext</b>	Entwicklung eines Datenbank gestützten Tools zur Social Benchmarking-Analyse für Unternehmen. Auf Basis des Social-Media-Auftritts und den daraus auswertbaren Parametern (wie z.B. bestimmte Klickzahlen) soll das Unternehmen messbar gemacht werden.
<b>Programmiersprache, Framework und Datenbanksprache zur Entwicklung der Funktionalitäten und der Visualisierung sowohl der Oberfläche als auch der Daten:</b>	Python, Jinja2, SQLite
<b>Von der Programmiersprache u.a. bereitgestellte Bibliotheken, die von Nutzen sein können:</b>	<p>matplotlib</p> <p>openpyxl (für Umsetzung der Idee, dass Daten aus der Datenbank in eine Excel importiert werden können).</p> <p>python-pptx (für Umsetzung der Idee, dass Daten aus der Datenbank in eine PowerPoint importiert werden können).</p> <p>nltk (für Auswertung der Tweets im Hinblick auf Inhalt lastige Wörter).</p>
<b>API und deren Bibliotheken zur Datengewinnung:</b>	Twitter API: tweepy, python-twitter
<b>Relevante Methoden der API:</b>	<p>Mit Hilfe des SocialMedia-Benchmark-Tool soll der Social-Media Auftritt des Unternehmens messbar gemacht werden. Hierzu werden folgende Parameter in Betracht gezogen, die wichtig sind, um die Relevanz und den Stellenwert zu bemessen, den das Unternehmen sich selbst im sozialen Netzwerk gibt und den es durch die User in Form von Resonanz erhält:</p> <ul style="list-style-type: none"> <li>○ Anzahl der Tweets &lt;-&gt; Anzahl der Retweets</li> <li>○ Anzahl der Follower</li> <li>○ Anzahl derer, denen das Unternehmen folgt</li> <li>○ Anzahl der Gefällt mir Angaben</li> <li>○ Anzahl der Kommentare</li> <li>○ Eventuell die Inhalte der Kommentare</li> </ul> <p>In diesem Sinne sind die „get...“-Methoden der API wichtig, um an die Daten zu kommen, worunter z.B. zählen:</p> <ul style="list-style-type: none"> <li>○ API.retweets(id[, count])</li> <li>○ API.direct_messages([since_id][, max_id][, count][, page][, full_text])</li> <li>○ API.direct_messages([since_id][, max_id][, count][, page][, full_text])</li> </ul>
<b>Konzeptioneller Rahmen der Idee:</b>	Der Nutzer (jemand aus dem Unternehmen, der sich mit Social Media und Benchmarking befasst) hat die Möglichkeit in der Datenbank einen Account anzulegen, zu dem nur er Zugang hat

	<p>und unter dem alle Daten gespeichert (archiviert) und aber auch wieder gelöscht werden können. Das fertige Projekt soll gewährleisten, dass Nutzernamen und Passwörter bei der Anmeldung übereinstimmen.</p> <p>Das Projekt soll so gestrickt sein, dass der Nutzer über einen der bereit gestellten Buttons eine Benchmark Analyse generieren kann, die auf Basis von speziellen und untereinander verglichenen Parametern (wie z.B. die Anzahl von Tweets pro Tag und die Retweets) graphisch aufbereitet werden.</p> <p>Als Nice-to-have soll der Nutzer, neben dem Abspeichern, Archivieren und Löschen der Daten in der Datenbank die Möglichkeit haben, die generierten Daten im Format einer PowerPoint und/oder einer Excel lokal auf seinen Rechner herunterzuladen.</p> <p>Es ist so vorgesehen, dass der Nutzer selber aktiv werden muss, damit Daten erhoben und analytisch aufbereitet werden (durch Klick auf Button). Er muss also in seinem Account in der Datenbank angemeldet sein. Es ist nicht vorgesehen, dass die Datenerhebung(en) automatisiert, ohne Zutun des Nutzers, abläuft/ablaufen.</p> <p>Die in der Datenbank pro Tag erhobenen Informationen können sowohl einzeln, also als pro Tag gewonnener Informationspool als auch in Form einer Zeitreihe ausgewertet und visuell abgebildet werden. Die erste Option ist zwangsläufig relevant, um die Informationen explizit an den relevanten Tagen in die Datenbank zu holen. Die zweite Option eignet sich, wenn man den in der Datenbank vorab gespeicherten Stand der Dinge z.B. zum 2. eines Monats mit dem ebenfalls vorab gespeicherten Stand der Dinge z.B. zum 25. eines Monats vergleichen möchte oder z.B. den Stand der Dinge zum 28.08.2016 mit dem Stand der Dinge am 23.09.2016 vergleichen möchte. Die erste Option eignet sich also zur Datengewinnung, Datenauswertung sowie zur Datenvisualisierung, um den Stand der Dinge an einem bestimmten Tag graphisch aufbereitet vorliegen zu haben und die zweite Option eignet sich zur erweiterten Datenauswertung und Datenvisualisierung, um die Veränderungen graphisch vor Augen zu haben, die sich innerhalb eines Zeitintervalls ergeben haben können.</p>
--	---