

Philosophische Fakultät III

Sprach- , Literatur- und Kulturwissenschaften

Institut für Information und Medien, Sprache und Kultur (I:IMSK)  
Lehrstuhl für Informationswissenschaft

Vertiefungsseminar Information Retrieval

Modul: INF -04.3

SS 2015

Leitung: Florian Meier M.A.

**Twitter4Beer**A Search-Engine that connects.

Dominik Altmann, Thomas Hardieck, Paul Wunderlich

1676209

4. Semester B.A. Informationswissenschaft

Tel.: [Telefonnummer (z.B. 0941/9999)]

Email: [Emailadresse (z.B.: max.mustermann@stud.uni-regensburg.de)]

Abgegeben am [Abgabetermin der Arbeit]

Inhalt

1. Einleitung 1,5
2. Theoretischer Teil
   1. IR Facetten 0,5
   2. Informationsbedürfnisse und Informationssuchverhalten der Nutzer 2
   3. Leitfragen für das eigene Projekt 1
3. Praktischer Teil
   1. Solr
      1. Datengrundlage und Indexierung 3
      2. Textanalyseprozess und Analyzer-Definition 2
      3. Anpassung des Request Handler 1
   2. SUI Design
      1. Search-Patterns und UI-Features 2
      2. Umsetzung mit AJAX-Solr 2
   3. Systemevaluation
      1. Vorgehensweise 2
      2. Ergebnisse 1
4. Quellenverzeichnis

Abbildungen

[Abbildung 1 – Wüste (Norman, 2010, S. 110) **Fehler! Textmarke nicht definiert.**](#_Toc284864361)

[Abbildung 2 – Blume (Norman, 2010, S. 31) 3](#_Toc284864362)

# Einleitung

Twitter gehört zu den erfolgreichsten Sozialen Netzwerken unserer Zeit. Nach eigenen Angaben kann die Plattform über 316 Millionen Besucher pro Tag , die über 500 Millionen Tweets absenden, verzeichnen.(“Twitter Usage,” 2015) Die Seite *Twitter.com* wurde 2006 von Gründer Jack Dorsey ins Leben gerufen. Am 19. April 2007 wurde „Twitter aus der Obvious Corp. ausgegliedert und [war] fortan ein eigenständiges Unternehmen“, das den Namen *Twitter, Inc.* trug. (“Twitter Meilensteine,” n.d.)

Die Suchmaschine *Twitter4Beer* hat sich zum Ziel gesetzt, eine eigene, verbesserte Suche rund um das Thema Bier und Braukultur zu auf Twitter abgesendeten Tweets zu liefern. Besonders zur Jahreszeit der Volksfeste, d.h. August, September, Oktober, werden in Deutschland überdurchschnittlich viele Tweets rund um das Thema Bier abgegeben, welche von *Twitter4Beer* somit unabhängig abgerufen werden können. Nachteile der Seite Twitter.com wie: Spam, Werbung, aggressive following und Datenschutz-Bruch, sollen in dieser Suchmaschine außen vor gelassen werden, da sie sich allein auf die Tweets und ihre Inhalte konzentriert.

Diese Arbeit soll als Dokumentation zu Erstellung, Funktion und Ziel der, im Rahmen des Vertiefungsseminars Information Retrieval, erstellten Suchmaschine *Twitter4Beer* dienen.

# Theoretischer Teil

## IR Facetten

Twitter bietet seinen Usern die Möglichkeit Nachrichten in Echtzeit-Übertragung zu senden und zu empfangen. Diese Nachrichten, so genannte Tweets, können maximal 140 Zeichen enthalten und werden auf dem persönlichen Profil angezeigt. Abgesendete Tweets sind standartmäßig öffentlich und können auch von nicht-registrierten Usern gelesen werden. Ein angemeldeter User jedoch kann die Tweets anderer User retweeten, also auf dem eigenen Profil teilen, favorisieren oder auf einen abgesendeten Tweet antworten. Um auf den Tweet eines bestimmen Nutzers zu antworten, wird dem eigenen Text ein *@* als Präfix vorangesetzt, um den Nutzer in dem Tweet zu markieren und damit den Text für ihn und seine Follower sichtbar zu machen. Durch den @-Präfix kann außerdem die Suche nach Tweets bestimmter User vereinfacht werden. Begriffe, denen ein @-Symbol vorausgesetzt werden, werden mit einem Hyperlink versehen und führen direkt auf das Profil des entsprechenden Nutzers. Des Weiteren kann ein User einem anderen User folgen und bekommt so dessen Tweets automatisch auf seiner Startseite angezeigt. Ein weiteres Feature sind private Nachrichten, die unter den Usern versendet werden können. Diese sind nicht öffentlich und können nur von den, an der Konversation beteiligten Usern gelesen werden. (“Twitter.com,” n.d.)

Ein Charakteristikum von Twitter ist die Verwendung des Hashtags in Tweets. Das erste Hashtag wurde im August 2007 von dem Nutzer *Chris Messina* verwendet. (“Twitter Meilensteine,” n.d.) Heute wird das Hashtag vor allem verwendet um in einem Tweet Schlagwörter und Themen hervorzuheben. Diese können dann später von anderen Nutzern zur Kategorisierung und Suche verwendet werden. Durch den Hashtag ist eine sehr schnelle und gezielte Informationssuche möglich, wodurch sich vor allem Trends und aktuelle Themen leicht verfolgen lassen.(“Twitter Support - Hashtag,” 2015)

Wird eine Standart-Suche ausgeführt, lassen sich Ergebnisse durch einen Klick auf Top, Live, Accounts, Fotos oder Videos filtern. Twitter äußert sich auf seiner Support-Seite dazu wiefolgt: " Wenn Du auf Top klickst, werden beliebte Tweets angezeigt, die viele andere Nutzer auf Twitter als interessant und nützlich empfunden haben. [...]Wir haben einen Algorithmus entwickelt, der die Tweets findet, die die Aufmerksamkeit anderer Nutzer geweckt haben. Wir wählen die Top-Tweets nicht von Hand aus."(“Twitter FAQs zu Top-Suchergebnissen,” n.d.) Des weiteren werden dem Nutzer weitere Optionen angeboten, unter denen man die Suche geodaten-basiert einschränken, oder sich nur Tweets von Nutzern denen man folgt anzeigen lassen kann. Außerdem kann man die getätigte Suche speichern. Auf der linken Seite werden dem Nutzer unter den Punkten "Related searches" und "Trends" weitere Facetten in Form von Hashtags angeboten. Darüber hinaus werden dem Nutzer andere User vorgeschlagen, die er eventuell kennen könnte und denen er eventuell folgen möchte.

Twitter bietet seinen Nutzern somit eine Vielzahl Von technisch sehr ausgereiften Suchmöglichkeiten.

Eine der wichtigsten Facetten

## Informationsbedürfnisse und Informationssuchverhalten der Nutzer

Der klassische Tweet ist durch seine 140 Zeichen Begrenzung im Vergleich zu Webseiten sehr kurz gehalten. Webseiten hingegen entwickeln sich auch nach ihrer Erstellung immer weiter, ein Tweet jedoch verändert sich meist nicht mehr nachdem er gepostet wurde.

So unterscheiden sich auch die Informationsbedürfnisse und das Informationssuchverhalten der Twitter-Nutzer essentiell von denen der webbasierten Suche, wie Teevan et al. (2011) in ihrer Arbeit erforschten.

Im Gegensatz zur Websuche konzentriert sich das Informationsbedürfnis der Twitter-User vor allem auf sehr zeitnahe Information, wie zB. die aktuellen Nachrichten aus der Region. Durch die Twitter-Suche soll der Output vor allem dazu dienen auf dem neuesten Stand der aktuellen Ereignisse zu sein. Die klassische Websuche hingegen dient laut Teevan, et al. (2011) vor allem der Weiterbildung und des Anreicherns von Wissen im Bereich eines bestimmten Interessenfeldes.

Ein weiterer Unterschied zwischen den beiden Suchverhalten haben Teevan, et al. (2011) in der Formulierung der Twitter-Suchanfragen ausgemacht. Diese seien verbreiteter, werden häufiger wiederholt und ändern sich seltener als bei Suchanfragen der Websuche.

Besonders der gesellschaftliche Aspekt steht bei der Twitter-Suche im Vordergrund. Während die Websuche mehr auf rationale Fakten und Navigationsanfragen, wie zB. "Google", abzielt, beeinhalten die Suchergebnisse der Twitter-Suche meist vor allem soziale Inhalt und Informationen über bevorstehende Veranstaltungen, bzw. aktuelle Ereignisse.

Ein besonderer Fokus sollte auf Berühmtheiten bei Twitter gelegt werden. Nach Teevan, et al.(2011) sind die Namen berühmter Personen eine der führenden Suchkategorien auf Twitter. Hierbei nutzen jedoch die wenigsten User das Suchverfahren um sich in Bezug auf einen bestimmten Aspekt einer Person zu informieren, als viel mehr um sich über aktuelle Informationen über die betreffende Person auf dem Laufenden zu halten.

Etwas näher an der herkömmlichen Websuche befindet sich die Twitter-Suche laut Teevan, et al. (2011) in Bezug auf das Informationsbedürfnis nach spezifischen Gebieten, wie zB. Astronomie, auch wenn diese Gebiete wie die meisten Aspekte der Twitter-Suche erneut sehr zeit-gebunden und vor sozialem Hintergrund betrachtet müssen.

Im Hinblick auf das Suchverhalten der Twitter-User muss vor allem die spezielle Syntax betrachtet werden, die zur Formulierung einer Suchanfrage verwendet wird. Wie unter Gliederungspunkt 2.1 bereits erklärt, wird der @-Präfix von den Nutzern dazu verwendet auf den Nicknamen und damit das Profil anderer Nutzer hinzuweisen. Auch das Hashtag-Zeichen ist in den Suchanfragen der Twitter-Nutzer sehr verbreitet. Teevan, et al. (2011) setzen in ihrer Arbeit die Durchschnittslänge der Wörter in den Twitter-Suchanfragen in Zusammenhang mit der Verwendung von Hashtags, da diese meist besonders lange, zusammengebundene Wörter sind. Im Vergleich zur Websuche, ist hier die durchschnittliche Wortlänge der Twitter-Suche um mehr als einen Buchstaben länger.

Das @-Präfix und das Hashtag-Zeichen können wie Suchanfragen-Operatoren benutzt werden und so den Sucherfolg verbessern, indem die Mehrdeutigkeit reduziert wird.(Teevan, Ramage, & Morriss, 2011)

# Praktischer Teil

## Solr

### Überschrift 3

### Überschrift 3

#### Beispielabschnitt mit Abbildung

Die Breite der Abbildungen kann variabel sein: Wichtig ist die Lesbarkeit.

Auch den Abbildungen wird eine Formatvorlage zugewiesen. Dadurch werden die Abbildungen zentriert und automatisch der Abstand zum Text nach oben eingestellt: Es müssen keine Leerzeichen eingefügt werden!



Abbildung 1 – Blume (Norman, 2010, S. 31)

Dies ist ein Beispieltext nach einer Abbildung (nicht eingerückt). Nutzt man die Funktion zur Erstellung von Querverweisen, muss man sich keine Gedanken mehr um die Anpassung von Nummerierungen bei Austausch von Abbildungen (z.B. **Fehler! Verweisquelle konnte nicht gefunden werden.** und Abbildung 1) oder ganzen Abschnitten mehr machen, da alle Verweise automatisch aktualisiert werden können. Markieren Sie dazu den gesamten Text im Dokument mit Strg+A starten Sie die Aktualisierung mit F9.

#### Zweiter Abschnitt auf vierter Ebene

Sobald Sie eine Gliederungsebene beginnen, muss es auch einen zweiten Abschnitt geben. Beispielsweise darf nie ein Abschnitt 2.3.1 stehen, wenn es keinen Abschnitt 2.3.2 gibt.

## SUI-Design

# Weiterer Abschnitt

Literaturverzeichnis

American Psychological Association. (2010). Concise Rules of APA Style (6th ed.). Washington: United Book Press.

Norman, A. D. (1988). Infuriating by design. Psychology Today, 22(3), 52-56. Retrieved from http://www.psychologytoday.com/. [24.12.2010]

Wimmer, E., Hornung A., & Lukesch H. (2008). Hinweise zur inhaltlichen und formalen Gestaltung von Seminar- und Diplomarbeiten im Fach Psychologie. Retrieved from: http://www-cgi.uni-regensburg.de/Fakultaeten/Psychologie/  
Lukesch/downloads/Lehre/gestaltung\_seminararbeiten.pdf. [07.02.2011]