# D:\Studium\fhb_logo_neu.png

**Fachbereich Informatik und Medien**

**Begleitmaterial zur Projektarbeit „Twitalyse“**

Vorgelegt von: Andy Klay 2009…

Michael Koppen 2009…

Christoph Ott 2009…

Daniel Schmidt 20090063

Inhalt

Einleitung -Was ist Twitter?

-Motivation

-Was kann man mit den Daten machen?

Umsetzung -Techniken, Zusammenspiel derer, Probleme

Fazit -sind die gewonnenen Daten nützlich, wer profitiert davon

Einleitung

# Twitter kurz erklärt

In der heutigen Zeit ist so gut wie jeder Internetnutzer in einem sozialen Netzwerk aktiv oder hat zumindest schon mal von einem solchen gehört. Dieser Begriff umschreibt die Vernetzung mehrerer Menschen auf einer bereitgestellten Plattform im World Wide Web. Dort lassen sich Freundschaften schließen bzw. am Leben halten oder alte Bekannte wiederfinden. Mit diesen teilt man sich dann verschiedene Interessen, lädt Fotos von der letzten Party hoch oder verabredet sich zum Kaffee. Bekanntestes Beispiel für ein soziales Netzwerk mit bald 960 Millionen Nutzern ist wohl Facebook.

Auch nutzen viele User die Möglichkeit sogenannte Blogs zu schreiben. Diese verhalten sich quasi wie Online-Tagebücher und können von anderen Nutzern gelesen und kommentiert werden. Der Inhalt solcher Blogs variiert genau so stark, wie die individuellen Nutzer die diese schreiben. So können dies etwa Tutorials (Anleitungen) zu bestimmten Problemen sein, oder einfach nur ein Bericht aus dem letzten Urlaub.

Als „kleinere aber dennoch komplexere“ Variante der Blogs könnte man Twitter bezeichnen. Hier hat der Nutzer die Möglichkeit, in maximal 140 Zeichen eine für ihn passende Statusmeldung abzusetzen (Tweets). Andere User haben auch hier wieder die Möglichkeit, Kommentare abzusetzen. Zusätzlich können die Ersteller der Beiträge abonniert werden. Mit diesem Schritt wird man nun zum „Follower“ des gewünschten Nutzers und sieht in einer Übersicht alle Aktivitäten des Tweet-Erstellers. Weiterhin können User Freundschaften schließen. Twitter wird somit zum sozialen Netzwerk.

# Mehr als nur ein Status

Setzt nun jemand eine Statusmeldung ab, so ist auf den ersten Blick nichts weiter als diese zu erkennen. Jedoch „hängen“ an dieser noch viele sogenannte Meta-Daten. So kann man, mit der richtigen Technik, zum Beispiel sehen, aus welchem Sprachraum der Poster kommt. Auch kann man auslesen, wie viele Freunde/Follower User X besitzt. Neben diesen Daten gibt es noch weitere, auf die später noch eingegangen wird.

Was kann man nun mit diesen angehängten Informationen machen? Zum einen könnte man einfach nur Statistiken für ein Projekt sammeln und auswerten. Firmen allerdings könnten ein viel größeres Interesse daran haben (was sie auch tun). So können diese zum Beispiel speziell auf den User zugeschnittene Werbung präsentieren. Die Chance auf einen Klick und somit Einnahmen sind auf diese Art viel höher.

Auch lassen sich Trends ermitteln. Nehmen wir an, in Russland haben User von einem Produkt aus den USA gehört und Posten bei Twitter, dass eben dieses „total toll und angesagt“ ist aber in ihrem Heimatland nicht zur Verfügung steht. Häufen sich diese Meldungen zu einem gewissen Grad, so könnte der Hersteller anhand dieser Informationen sein Verkaufsgebiet auf das entsprechende Land ausweiten. Risiken würden so reduziert.

# Ziel dieser Arbeit

Das Ergebnis des Projekts Twitalyse soll eine konfigurierbare Menge an Daten liefern, die Aufschluss über das Verhalten der Twitter-User darstellen. Statistiken über die durchschnittliche Anzahl von Freunden oder Follower der Nutzer, veröffentlichte Tweets in den einzelnen Sprachräumen bzw. deren Antworten sind ebenfalls geplant.

So ist es möglich…

Umsetzung -Ressourcen

In diesem Teil des Dokuments werden die verwendeten bzw. ausgemusterten Techniken kurz erklärt. Die gesamte Programmierung erfolgt in der Sprache Java, da diese relativ leicht zugänglich ist.

…

# Hadoop

# Besser: Storm

Anders als Hadoop kann man mit Storm die gestreamten Daten „on the fly“ auswerten. Dieser Umstand passt um Längen besser in das geplante Projekt.

# Redis

Hierbei handelt es sich um eine sogenannte In-Memory-Datenbank. Diese benutzt den Arbeitsspeicher eines Computers als Datenspeicher. Vorteil dabei ist, dass somit eine viel höhere Schreib/Lesegeschwindigkeit erreicht werden kann (ca. 100.000 Schreib- und 80.000 Lesevorgänge pro Sekunde). Es lassen sich auch besser Vorhersagen treffen was die Zugriffszeiten angeht. Dadurch ist diese Datenbank auch sehr gut skalierbar. Die Daten sind allerdings nicht persistent, was jedoch bei diesem Projekt keine Rolle spielt, da die Daten sowieso im Stream verarbeitet werden und später nicht mehr gebraucht werden. Da es sich bei dem Stream von Twitter um sehr große Datenmengen handelt, ist Redis die perfekte Datenbank.

Fazit

Quellen:

Facebook - Nutzerzahlen: <http://allfacebook.de/userdata/>