API’s für den Zugriff auf Daten

aus sozialen Netzwerken

zur Sentimentanalyse bzw.

zum Option Mining.

Julian Nischler, B.Sc.

(S1310457017)

SEMINARARBEIT

Eingereicht am

Fachhochschul-Masterstudiengang

Information Engineering Management

in Hagenberg

im Juni 2014

Diese Arbeit entstand im Rahmen des Gegenstands

Seminar aus Information Management

(SIM)

im

Sommersemester 2014

Betreuer:

DI. Dr. Gabriel KRONBERGER

Inhalt

[1 Einführung 1](#_Toc390108171)

[1.1 Motivation 1](#_Toc390108172)

[1.2 Fragestellungen 2](#_Toc390108173)

[1.3 Zielsetzung und Methodik 2](#_Toc390108174)

[2 „soziale“ API’s 3](#_Toc390108175)

[2.1 Funktionsbeschreibung und Entwicklung 3](#_Toc390108176)

[2.2 Übersicht populärer API’s 7](#_Toc390108177)

[2.3 Facebook Open Graph 7](#_Toc390108178)

[2.4 Twitter Stream 7](#_Toc390108179)

[2.5 Historische Daten 7](#_Toc390108180)

[3 Sentimentanalyse 8](#_Toc390108181)

[3.1 klassische Verfahren 8](#_Toc390108182)

[3.2 Frameworks 8](#_Toc390108183)

[4 Praxisbeispiele 9](#_Toc390108184)

[4.1 Facebook Open Graph 9](#_Toc390108185)

[4.2 Twitter Stream 9](#_Toc390108186)

[4.3 Sentimentanalyse von Stream Daten mittels DatumBox 9](#_Toc390108187)

[5 Zusammenfassung 10](#_Toc390108188)

[5.1 Ergebnisse 10](#_Toc390108189)

[5.2 Allgemeines Resümee 10](#_Toc390108190)

[5.3 Persönliches Resümee 10](#_Toc390108191)

[6 Literatur 11](#_Toc390108192)

# Einführung

## Motivation

Soziale Netzwerke sind von der derzeitigen „Thumb“-Generation nicht mehr weg zu denken. Netzwerke wie Facebook erfahren einen enormen Aufschwung an Nutzern und Verwendung. Kaum jemanden sind diese Netzwerke kein Begriff. Diese Arbeit beschäftigt sich nicht mit den Vor- bzw. Nachteilen solche Netzwerke, vielmehr mit den nicht direkt ersichtlichen Möglichkeiten, den API’s. Viele Unternehmen fürchten einen Auftritt in solchen Netzwerken da sich negative Informationen bzw. Meinungen sehr schnell verbreiten und nur sehr schwer eindämmen lassen. In vielen Fällen lässt sich eine Präsenz jedoch nicht vermeiden, so auch in meinem Arbeitsumfeld. Tätig im Eventbereich sind wir auf solche Netzwerke zur Kundengewinnung stark angewiesen. Da diese Netzwerke völlig neue Marketingtechniken eröffnen. Es ist verhältnismäßig sehr günstig neue Kunden zu bewerben. Leider sind auch wir immer wieder mit der Problematik der Bildung einer negativen Meinung konfrontiert. Mittels der API’s der Netzwerke soll versucht werden negative Meinungen schnell zu erkennen um eine Ausbreitung zu verhindern.

Aus der einleitenden Motivation ergeben sich folgende Kernthemen für die Seminararbeit:

* Einführung in die Thematik „soziale“ API‘s
* Evaluierung der einzelnen API’s
* Kombination mit der Thematik „Sentiment Analyse“

## Fragestellungen

Folgende zentrale Fragestellungen sollen in der Seminararbeit beantwortet werden.

1. Sozial API spezifische Fragen
   1. Was ist eine „soziale“ API?
   2. Welche API’s existieren? (Auszug)
   3. Welche API für welchen Zweck geeignet?
   4. Wie werden die API’s verwendet?
2. Fragen in Verbindung mit Sentiment Analyse
   1. Klassische Ansätze des Sentiment Analysis?
   2. Wie gut eignen sich die API’s dafür?
   3. Problematiken der Sentiment Analyse?
   4. Welche Frameworks existieren?

Welche API’s existieren, wie werden diese verwendet, sowie eignen sich diese zur Sentimentanalyse sprich zur Erkennung von Meinungsbildungen?

## Zielsetzung und Methodik

Ziel dieser Arbeit ist, dem interessierten Leser einen Einstieg in die Verwendung von „sozialen“ API’s zu vermitteln, sowie in die Thematik der Sentimentanalyse. Aufbauend auf diesen Erkenntnissen soll gezeigt werden wie durch Sentimentanalyse der verfügbaren Daten profitiert werden kann und welcher Mehrwert aber auch welche Risiken dadurch entstehen. Anhand von Praxisbezogenen Beispielen wird versucht die Thematik zu veranschaulichen.

# „soziale“ API’s

## Funktionsbeschreibung und Entwicklung

Dieses Kapitel soll einen Überblick der am Stärksten verbreiteten API’s schaffen. Das Internet ist heutzutage eines der am Meisten verwendeten Medien zum Verbreiten von Information. Vor allen durch soziale Netzwerke wie Facebook, Twitter, YouTube und Co. Um dem Benutzer eine noch bessere Erfahrung zu ermöglichen gibt es unzählige Anwendungen welche die einzelnen Plattformen verknüpfen. Eine sehr bekannte solche Anwendung ist „Instagram“. Mittels dieser ist es möglich sehr einfach und schnell Fotos mit einem Smartphone zu erstellen, bearbeiten und vor allem zu verteilen. Dabei wird das neue Foto mit nur einem Klick in nahezu allen sozialen Netzwerken verteilt oder geteilt. Solche Anwendungen werden von den Entwicklern meist „mashups“ genannt, sprich die nahtlose Kombination von Plattformen und Innhalten.

So gut wie alle sozialen Netzwerke stellen leistungsstarke API’s zur Verfügung, wodurch solche „mashups“ überhaupt erst ermöglicht werden. Fast ausschließlich fungieren diese in Form von Web Services[[1]](#footnote-1). Die Anzahl der verfügbaren API’s ist in den letzten Jahren förmlich explodiert, so stieg die Anzahl der API’s von etwa 200 Anfang 2006 auf über 10000 im Jahr 2013 und dabei handelt es sich lediglich um auf „ProgrammableWeb“ registrierte API‘s. Abbildung 1 veranschaulicht das Wachstum der letzten Jahre. (Programmable Web, 2013)

Viele frühen API’s verwendeten SOAP[[2]](#footnote-2) ein Netzwerkprotokoll mit dessen Hilfe Daten zwischen Systeme ausgetauscht werden können, zur Repräsentation der Daten wird XML[[3]](#footnote-3) verwendet.

Durch das enorme Wachstum von Webanwendungen in den letzten Jahren wurden einfachere Protokolle gefordert, welche mittels einfacher Skriptsprachen wie JavaScript verwendet werden können. SOAP war nie für Webanwendungen gedacht, dessen Anwendungsbereich liegt im Enterprise Bereich.

Weshalb sich in den letzten Jahren das REST[[4]](#footnote-4) Programmierparadigma durchsetzte und mittlerweile von mehr als 70% aller API’s angeboten wird. Abbildung 2 zeigt die Entwicklung von REST in Vergleich zu SOAP in den letzten Jahren. (Programmable Web, 2013)

REST beschreibt strenggenommen kein Protokoll auch keine Norm es beschreibt eigentlich lediglich die Methodik einer Web URL genau einer Serverseitigen Aktion zuzuordnen. Als Protokoll wird HTTP oder HTTPS verwendet. Dabei werden die folgenden HTTP Befehle *(HTTP-Verb)* verwendet.

* GET
  + Anfordern der angegeben Ressource (URL), es wird nur gelesen.
* POST
  + Zusätzlich zum Anfordern der Ressource werden zusätzlich Nutzdaten vom Client an den Server übertragen. Es wird gelesen und geschrieben.
* DELETE
  + Löscht die angeforderte Ressource.

Sowie die zusätzlichen Befehle *PATCH, PUT, HEAD, OPTIONS, CONNECT* und *TRACE* auf diese wird im Zuge dieser Arbeit nicht weiter eingegangen da Diese im Kontext nicht von Bedeutung sind.

REST Spezifiziert auch nicht die Form der Datenrepräsentation. Ein sehr praktisches und einfaches Datenformat stellt JSON dar, da es direkt von JavaScript interpretiert werden kann. Weshalb sich JSON nahezu gleich mit REST verbreitet so bieten mittlerweile weit über 50% der API’s das JSON Format an. Viele Moderne API’s wie Facebook Open Graph bieten nur noch JSON an.

In Abbildung 3 wird ein einfaches Beispiel gezeigt. Mittels eines REST GET Request wird das Wetter von Salzburg abgefragt. Der Web Service antwortet wahlweise mit XML oder JSON. Es ist ersichtlich das JSON Daten leichter lesbar sind und wirken übersichtlicher. Auf den ersten Blick würde man nicht glauben das die JSON Daten kleiner als die XML Repräsentation. JSON brauch für die Darstellung Vertikal zwar viel mehr Platz ist jedoch deutlich kleiner. 441 Bytes JSON gegen 659 Bytes XML.

Abbildung 1 Wachstum Anzahl angebotene API’s auf ProgrammableWeb

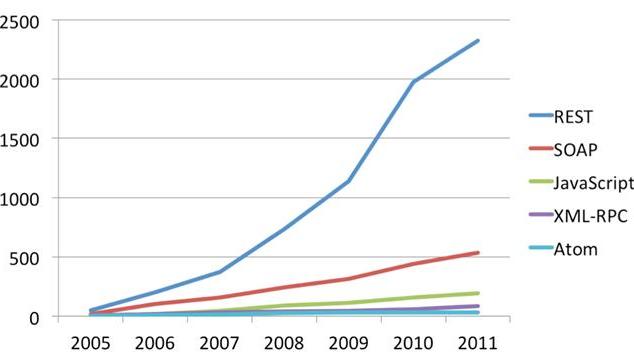


Abbildung 2 Vergleich REST zu SOAP und Alternativen

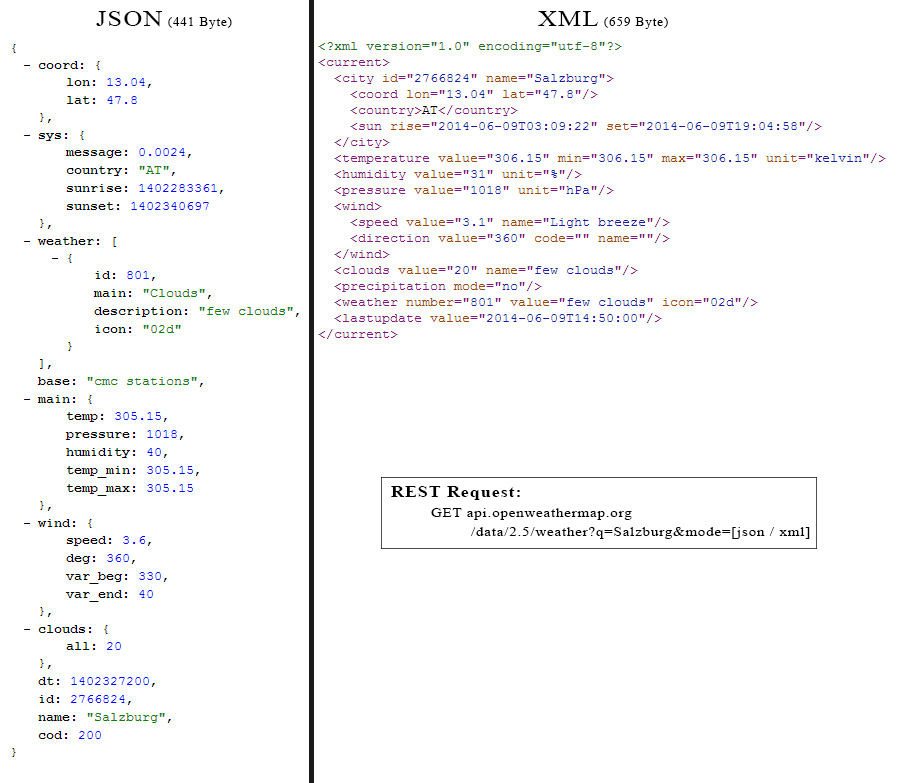


Abbildung 3 Beispiel für REST Request mit XML bzw. JSON Antwort

## Übersicht populärer API’s

### Facebook

Siehe 2.3 Facebook.

### Twitter

Siehe 2.4 Twitter.

### Google+

Google bietet mit Google+ ein soziales Netzwerk welches im Funktionsumfang Facebook ähnelt. Mittels der verfügbaren REST API ist es möglich Personen, Aktivitäten, Kommentare sowie Momente anzufragen und erstellen. Zusätzlich können Abfragen auch mittels Volltextsuche formuliert werden. Dadurch kann man sehr schnell Informationen einer Person oder Zielgruppe erhalten. Google verwendet vorzugsweiße JSON als Datenformat. (Google Inc, 2014)

### bitly

Bitly zählt zu den am häufigsten verwendeten „URL-shorten“ Diensten. Dieser Dienst wird vor allem gerne für Twitter-Nachrichten verwendet um die maximale Nachrichtenlänge nicht zu überschreiten. So werden Täglich rund 1 Million bitly links erstellt und über 4 Billionen pro Monat angeklickt. Bitly bietet API’s mit denen es Möglich ist die Populärsten Links ausfindig zu machen. Und gerade im Business Intelligence Bereich sehr interessant die Verbreitung von Phrasen bzw. Innhalten der Zielseiten. Als interessantes Beispiel zeigt bitly eine Echtzeitkarte, welche die Verteilung von Artikel der einzelnen Nachrichtenportale in den USA zeigt. (bitly inc, 2014)

### del.icio.us (delicious)

Delicious ermöglicht es dem Benutzer Lesezeichen abzulegen und diese von all seinen „smart - Devices“ abzufragen. Lesezeichen werden nicht in Ordner gespeichert sondern mit Tags versehen, dies geschieht Teilweise automatisch. Mit den von delicious zur Verfügung gestellten API ist es leider nur Möglich Daten des aktuellen Benutzers abzufragen, diese Einschränkung ist sehr bedauerlich. Da Benutzer die Links bereits mit Tags klassifizieren hätten diese Daten einen sehr hohen Mehrwert für unter anderem das Suchen und Finden von Referenzen oder Querverweisen.

### Foursquare

Foursquare der wohl bekannteste Geotagging Dienst. Grundsätzlich eine Weltkarte auf der Benutzer Plätze eintragen können und diese auch bewerten. In den Letzten Jahren ist Foursquare stark gewachsen und weißt mittlerweile eine der größten Geodatenbanken auf. Mittels der angebotenen API’s ist es möglich anhand von Koordinaten oder Suchbegriffe Plätze ausfindig zu machen. So ist es unter anderem Möglich dem Benutzer Informationen aus seiner Umgebung anzubieten.

## Facebook

<https://www.facebook.com/about/graphsearch>

FQL (alt)

GRAPH API

## Twitter

TWITTER API

TWITTER STREAM

## Historische Daten

qnip

# Sentimentanalyse

## klassische Verfahren

## Frameworks

# Praxisbeispiele

## Facebook Open Graph

## Twitter Stream

## Sentimentanalyse von Stream Daten mittels DatumBox

# Zusammenfassung

## Ergebnisse

## Allgemeines Resümee

## Persönliches Resümee

# Literatur

*Die Literatursuche hat sich für mich als sehr schwierig erwiesen. Ich könnte hier die Dokumentationen der einzelnen API’s aufführen aber ich denke diese haben hier nichts zu suchen. Somit führe ich bezüglich den API`s relativ wenig Literatur an.*

* Techniques and applications for sentiment analysis  
  R Feldman - Communications of the ACM, 2013 - dl.acm.org
* Social network analysis: Methods and applications  
  S Wasserman – 1994
* Sentiment analysis and subjectivity  
  B Liu - Handbook of natural language processing, 2010 - f3.tiera.ru
* Automatic Sentiment Analysis in On-line Text.  
  E Boiy, P Hens, K Deschacht, MF Moens - ELPUB, 2007 - law.kuleuven.be
* Twitter as a Corpus for Sentiment Analysis and Opinion Mining  
  A Pak, P Paroubek - LREC, 2010
* Twitter sentiment analysis: The good the bad and the omg!  
  E Kouloumpis, T Wilson, J Moore - ICWSM, 2011 - aaai.org
* Sentiment analysis and opinion mining  
  B Liu - Synthesis Lectures on Human Language Technologies, 2012 - morganclaypool.com
* A survey of opinion mining and sentiment analysis  
  B Liu, L Zhang - Mining Text Data, 2012 – Springer
* Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More  
  MA Russell - 2013
* A comparison of on and offline networks through the Facebook API  
  B Hogan - Available at SSRN 1331029, 2008 - papers.ssrn.com

1. Web Services stellen Maschine to Maschine Systeme dar welche Daten über Netzwerke übertragen. [↑](#footnote-ref-1)
2. Simple Object Access Protocol [↑](#footnote-ref-2)
3. Extensible Markup Language [↑](#footnote-ref-3)
4. Representational State Transfer [↑](#footnote-ref-4)