

# Modellierung eines Epidemie- Frühwarnsystems mit Social Media Mining

Nicolas With  
Master Seminar  
WS 2012/13

# Agenda

- Einstieg
  - Motivation
  - Abgrenzung
  - Ziel
- Status
  - Projekt 1
  - Projekt 2
- Ausblick
  - Chancen
  - Risiken
- Zusammenfassung

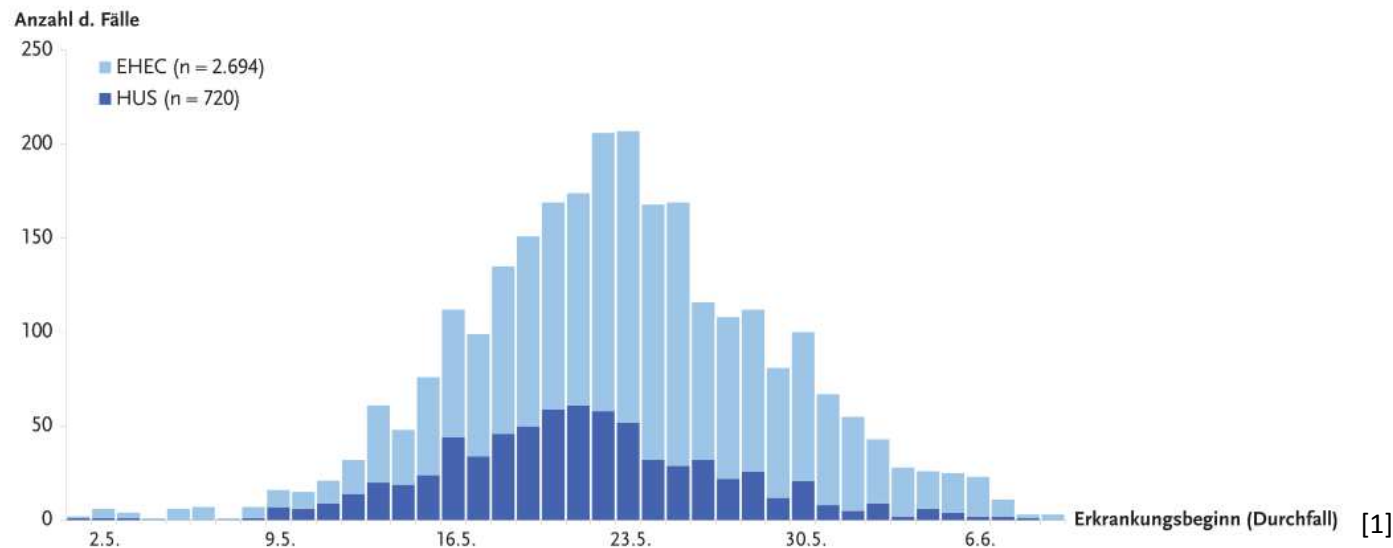
## Aufgabe

*Ein generisches System zu erstellen, welches über Data Mining und Textanalyse frühzeitig Informationen über eine anstehende Epidemie ausgibt und präsentiert.*

- Stichwort: „Epidemic Intelligence“

## Motivation

- Mai 2011: größter EHEC-Ausbruch in Deutschland
  - Am 1. Mai treten erste EHEC-Fälle auf
  - Am 19. Mai informiert das RKI das Gesundheitsamt
  - Am 23. Mai werden erste Maßnahmen ergriffen



## Motivation

- These: *Wenn ein Mensch krank wird, dann teilt er das dem Internet mit, noch bevor er zum Arzt geht.*<sup>[10]</sup>

Das führt zu:

- Frühere Erkennung von Erkrankungsfällen
- Schnellere Warnung einer potentiellen Epidemie
- Weniger Erkrankte und weniger Todesfälle

# Einstieg

## Social Media in Deutschland <sup>[2] [3] [4]</sup>



März 2011: 480.000 aktive Nutzer  
November 2012: 825.000 Nutzer



Oktober 2012: 24.600.000 Nutzer  
7% der Nutzer posten täglich (1.722.000 Posts)

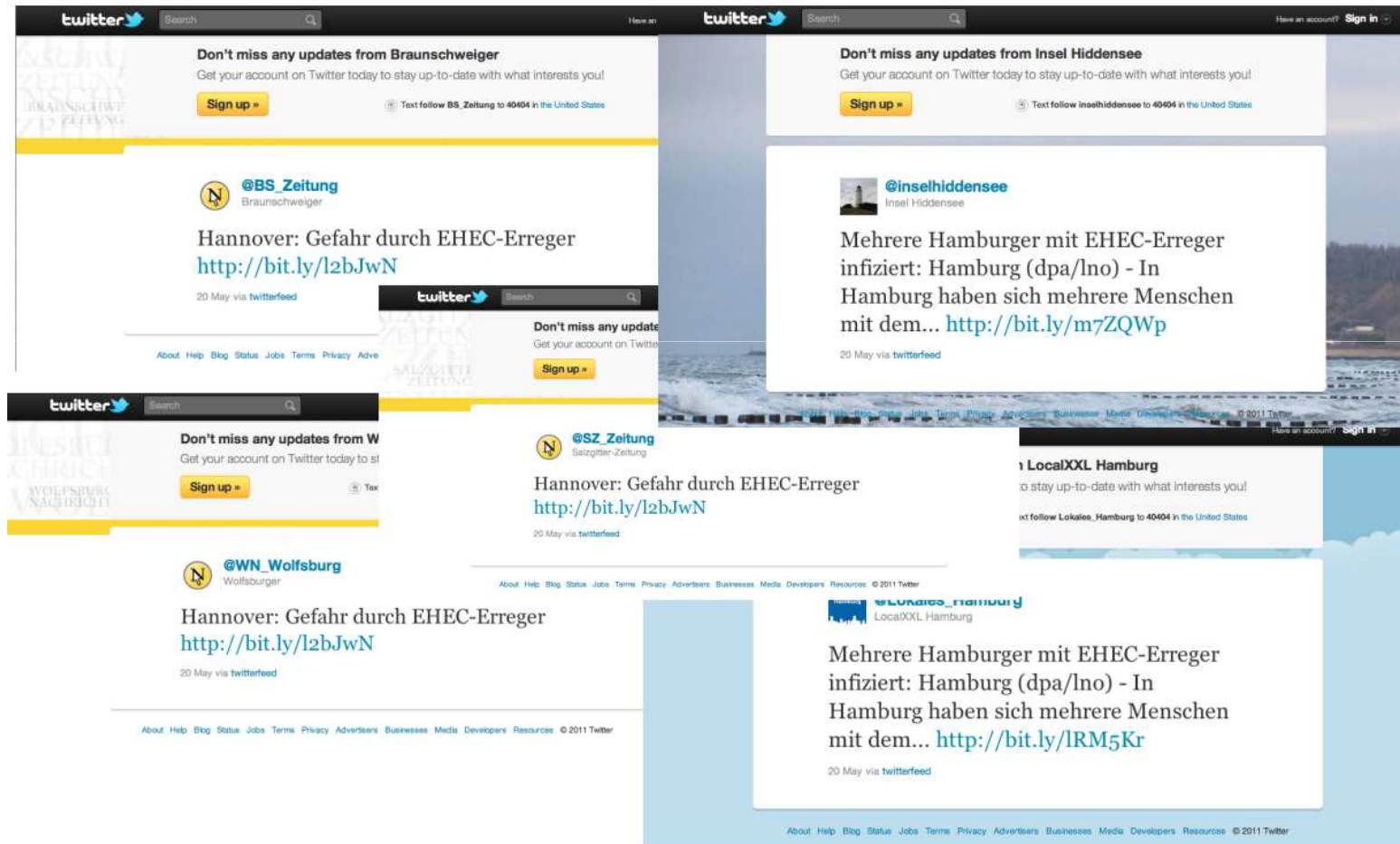


Oktober 2012: 2.100.000 Nutzer



Ca. zwischen 500.000 – 1.000.000 Blogs

# Einstieg



[5]

# Einstieg



**krümelinchen** @cookikruemel  
Eine Runde Mitleid bitte, denn das Krümelchen hat Fieber und bekommt grad eine Grippe :(

**Naturheilpraxis** @HaugHenseler  
Gehäufte EHEC-Fälle in Belgien:  
[bit.ly/KVyUPS](http://bit.ly/KVyUPS)

**Ana Rosa López** @mivozmipalabra  
Schweinegrippe fordert für #Bolivien - latina press Na  
[is.gd/gk76rP](http://is.gd/gk76rP) via @latinap

**Orrrr, Dandy** @dandyliving  
Ich will nicht sagen, dass ich schlimmen Durchfall habe, aber ich glaube ich hab auf der Toilette gerade den Tschernobylunfall nachgespielt.

1 RETWEET

Antworten Retweeten Folgen Favorisieren

[6]



## Abgrenzung

- Google Flu Trend (nur jährliche Influenza)<sup>[7]</sup>
- M-Eco<sup>[8]</sup>
  - Universität Hannover / Robert-Koch-Institut
  - „Personalized Event-based surveillance“
- HealthMap<sup>[9]</sup>
  - Zeigt Infektionsfälle auf einer Karte an
  - Daten kommen von medizinischen Medienquellen (ProMED, Eurosurveillance etc.)

# Einstieg



The screenshot displays the HealthMap website interface. At the top, the logo "HealthMap" is accompanied by the tagline "Global Health, Local Knowledge". Navigation tabs for "Global", "Local", and "News" are visible, along with a search bar and "Advanced Search" options. The main area features a world map with numerous orange and red pins indicating disease activity across various countries. A legend at the bottom left shows the "Activity Index" with color-coded pins for "Low" (yellow) and "High" (red) activity, and distinguishes between "Country level" and "Province or local level". Below the map, there are sections for "Outbreak Missing?", "flu near you", and "Quick Views" which includes "Alerts from past week", "Avian Influenza (past month)", and "Drought in Africa 2011".

[9]

# Ziel

- Generisches System aufbauend auf Arbeiten von M-Eco
  - Analyse und Erkennung von Krankheitsmustern
  - Extraktion von relevanten Informationen
  - Rückprojektion auf die Ursache der Epidemie
- Intermediale Vernetzung der Personen
- Nutzung von Services
  - Google Maps
- Frühere Warnungen als RKI / Gesundheitsamt

# Agenda

- Einstieg
  - Grundlage
  - Abgrenzung
  - Ziel
- Status
  - Projekt 1
  - Projekt 2
- Ausblick
  - Chancen
  - Risiken
- Zusammenfassung

## Projekt 1

- Entwicklung eines branchenspezifischen Frameworks für eHealth-Plattformen
  - Einheitliche Schnittstelle zu eHealth-Plattformen
  - Einheitliche Bearbeitung von eHealth-Records
- Generisches Modul zur Zustandsspeicherung
  - Speicherung von großen Zustandsräumen

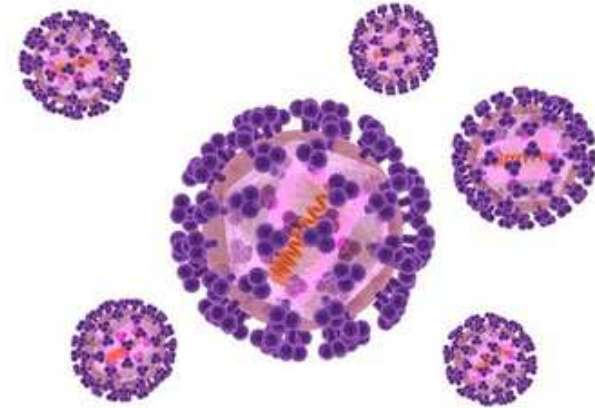
# Projekt 2

## Eingrenzung

- Krankheiten (EHEC / Hanta-Virus)
- Soziale Medium (Twitter)
- Architektur des Systems
  - Komponenten
  - Anforderungen
  - Services

## EHEC

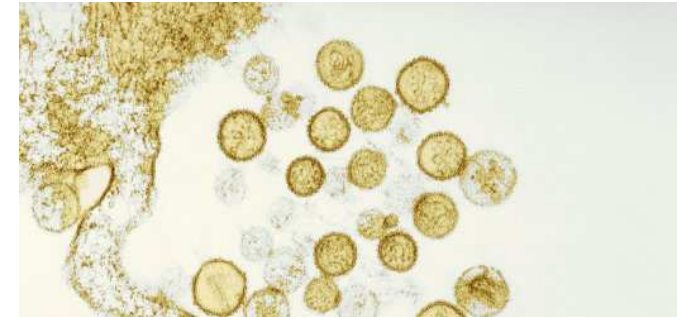
- Darminfektion
- Symptome
  - Übelkeit, Erbrechen, Durchfall
- Inkubationszeit
  - 3-4 Tage
- Übertragung
  - Mensch zu Mensch, Tier zu Mensch
- Auftreten
  - Ältere Personen und Kinder mit geschwächtem Immunsystem



# Projekt 2

## Hanta-Virus

- Lungenerkrankung, Nierenversagen
- Symptome
  - Grippeähnlich, Kopf-, Bauch-, Rückenschmerzen
- Inkubationszeit
  - 5-60 Tage
- Übertragung
  - Tier zu Mensch, Rötelmäuse
- Auftreten
  - Mittleres Alter, Forstwirtschaft, Mai bis September



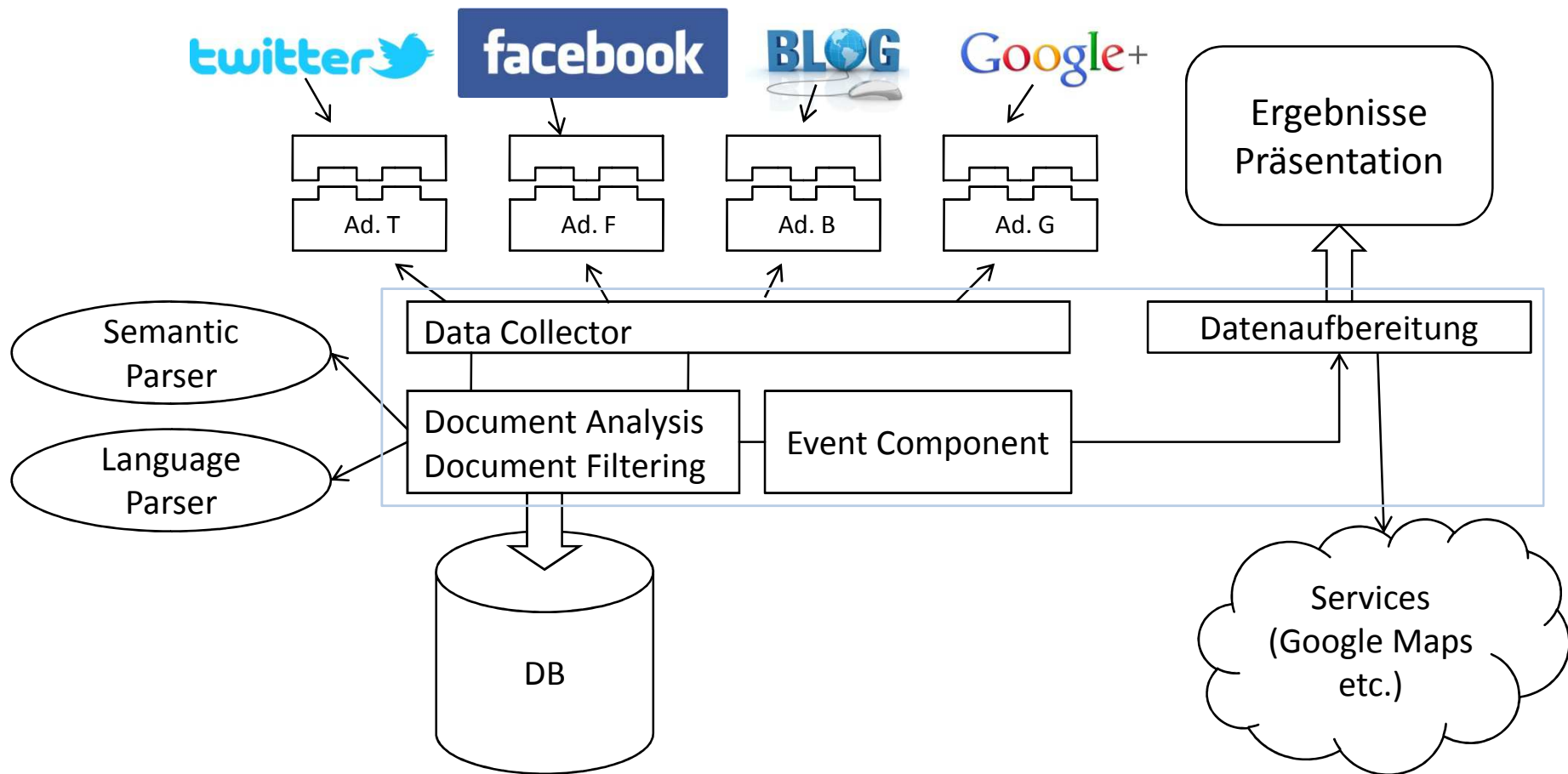


# Projekt 2

## Twitter

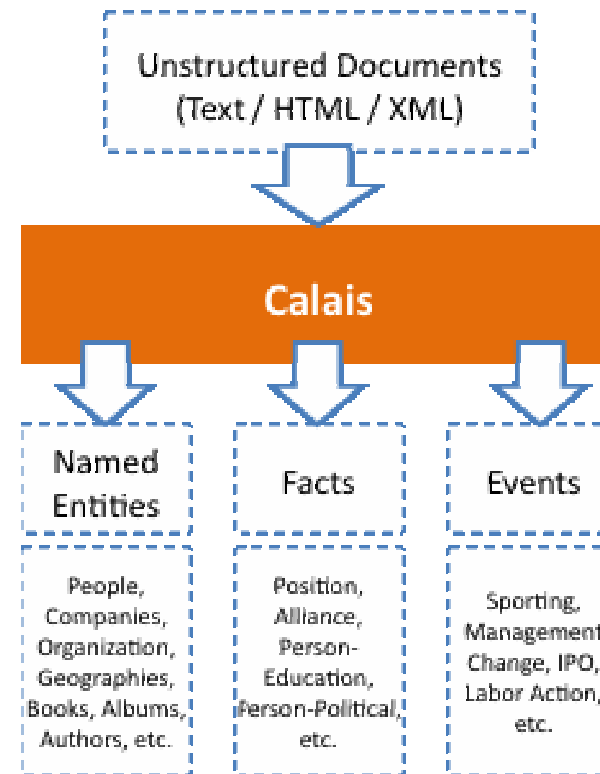
- Twitter API
  - Standard Twitter-Bibliothek
- Twitter Search API
  - Index von aktuellen Tweets (6 – 9 Tage alt)
  - Queries limitiert auf 1000 Zeichen
  - Beispiel: Tweets die @twitterapi and @anywhere beinhalten  
<http://search.twitter.com/search.json?q=%40twitterapi%20%40anywhere>

# Aufbau



## Analyse

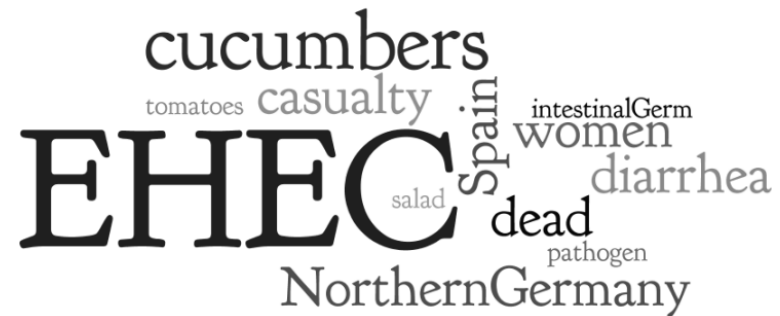
- OpenCalais<sup>[11]</sup>
  - Analysiert natürlichsprachige Dokumente
  - Gibt geordnetes Dokument als RDF zurück
  - Nur Englisch, Französisch, Spanisch
- Negra Corpus<sup>[12]</sup>
  - Linguistische Verarbeitung von deutschen Texten



# Projekt 2

## Präsentation

- Tag Cloud
- Graphen/Diagramme
  - Entwicklung
  - Häufigkeit passender Schlagwörter (per Region)
- Map
  - Service: Google Maps
  - Einfärben der betroffenen Gebiete -> Ausbreitung ersichtlich



# Agenda

- Einstieg
  - Grundlage
  - Abgrenzung
  - Ziel
- Status
  - Projekt 1
  - Projekt 2
- **Ausblick**
  - Chancen
  - Risiken
- Zusammenfassung

## Masterarbeit – Was muss getan werden?

- Generik
  - Verschiedene soziale Medien anbinden (mind. 2)
  - Einfache Verwaltung der Krankheiten
- Textanalyse ausbauen
- Datenaufbereitung
  - Präsentation über Google Maps
  - Rückschlüsse auf Ursprung der Epidemie
  - Rückschlüsse auf Ursache der Epidemie

## Chancen

- Frühes Warnsystem für die Gesundheitsämter
- Schnellere Reaktion / schnellerer Impfstoff
- Ausbreitungsmuster
  
- Fragen:
  - Wie früh lässt sich eine Epidemie erkennen?
  - Lassen sich Rückschlüsse auf den Werdegang und Ursprung der Epidemie geben?

## Risiken

- Deutschsprachige Posts
- Umfang
- Zugänglichkeit der Social Medias
  - Beschränkung der Social Media APIs – wie im Fall Twitter mit Version 1.1



# Zusammenfassung

- Epidemien-Frühwarnsystem
  - Aufbauend auf M-Eco
  - Begrenzt auf Deutschland / deutschsprachig
  - Generisches System
  
- Ziel
  - Frühere Epidemienwarnung als herkömmliche Wege
  - Werdegang der Epidemie
  - Ursprung / Ursache der Epidemie

Vielen Dank für Eure  
Aufmerksamkeit!

Und frohe Weihnachten!

# Quellen

1. [http://de.wikipedia.org/wiki/HUS-Epidemie\\_2011](http://de.wikipedia.org/wiki/HUS-Epidemie_2011)
2. <http://www.socialmediastatistik.de/825000-aktive-twitter-accounts-interview-mit-thomas-pfeiffer/>
3. [http://allfacebook.de/zahlen\\_fakten/deutschland-oktober-2012/](http://allfacebook.de/zahlen_fakten/deutschland-oktober-2012/)
4. <http://www.socialmediastatistik.de/google-100-millionen-aktive-nutzer-4-der-internetnutzer-in-deutschland-sind-aktiv/>
5. Diaz-Aviles E, Stewart A, Velasco E, Denecke K, Nejd W: Epidemic Intelligence for the Crowd, by the Crowd. The Sixth International AAAI Conference on Weblogs and Social Media ICWSM2012. Dublin, Ireland, June 4 - 7, 2012
6. <http://www.welt.de/gesundheit/article106735414/Epidemien-frueher-erkennen-mit-Twitter-amp-Co.html>
7. <http://www.google.org/flutrends/de/#DE>
8. <http://www.meco-project.eu/about>
9. <http://healthmap.org/en/>
10. [http://www.focus.de/gesundheit/ratgeber/zukunftsmedizin/tid-26294/social-media-spuert-epidemien-auf-wie-facebook-und-co-leben-retten-sollen\\_aid\\_772372.html](http://www.focus.de/gesundheit/ratgeber/zukunftsmedizin/tid-26294/social-media-spuert-epidemien-auf-wie-facebook-und-co-leben-retten-sollen_aid_772372.html)
11. <http://www.opencalais.com/about>
12. <http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/negra-corpus.html>