

# Social Data Mining

Albert Weichselbraun

Vienna University of Economics and Business Administration  
Department of Information Systems and Operations  
Augasse 2-6, 1090 Vienna

`albert.weichselbraun@wu.ac.at`

May 2009

# Agenda

## Aufgabenstellung

Facebook

Web 2.0 Datenquellen

Administration

## Technologie

Komponenten

Anonymisierung der Benutzerdaten

Facebook Application

Web 2.0 Datenquellen

Geographische Visualisierungen

# Aufgabenstellung

Ziel dieses Projekt ist es *anonymisierte* Profile von

- ▶ Benutzern und (facebook, studivz, ...),
- ▶ deren Interessen (flickr, digg, technorati, del.icio.us, ...)
- ▶ und weiteren Attributen (location, age, ...)

zu erstellen. Es soll möglich sein Verbindungen zwischen Benutzern sowie deren geographische Verteilung zu visualisieren und Abfragen nach den gespeicherten Attributen vorzunehmen.

# Arbeitspakete

- ▶ Eine Facebook Applikation zum Mining von sozialen Netzen
- ▶ Integration von Web 2.0 Datenquellen (flickr, del.icio.us, ...)
- ▶ ein Applikation zur Visualisierung der Ergebnisse

# Arbeitsaufteilung

## Kernaspekte

- ▶ Technische Aspekte: Facebook-Applikation, Web-Interface, technorati-Spider, Visualisierung, ...
- ▶ Organisatorische Aspekte: Zeitplan, Organisation von Treffen, Projektdokumentation, ...
- ▶ Strategische Aspekte: Planung, Modellierung, Marketing, Aufteilung der Workpackages, rechtliche Aspekte ...

Wöchentliches Reporting durch mindestens ein Gruppenmitglied

*Achtung:* Es wird erwartet, dass *jedes* Gruppenmitglied zumindest einen Programmiertask übernimmt.

# Facebook - Anforderungen

- ▶ Design, Erstellen und Verteilen einer Facebook Applikation
- ▶ Einhaltung der rechtlichen Bestimmungen (Datenschutz, ...)
- ▶ Marketing der Applikation (entsprechend grosser Benutzerkreis)
- ▶ Auswahl der zu verwendeten Felder
- ▶ Anonymisieren von allen personenbezogenen Daten

# Integration von Web 2.0 Datenquellen

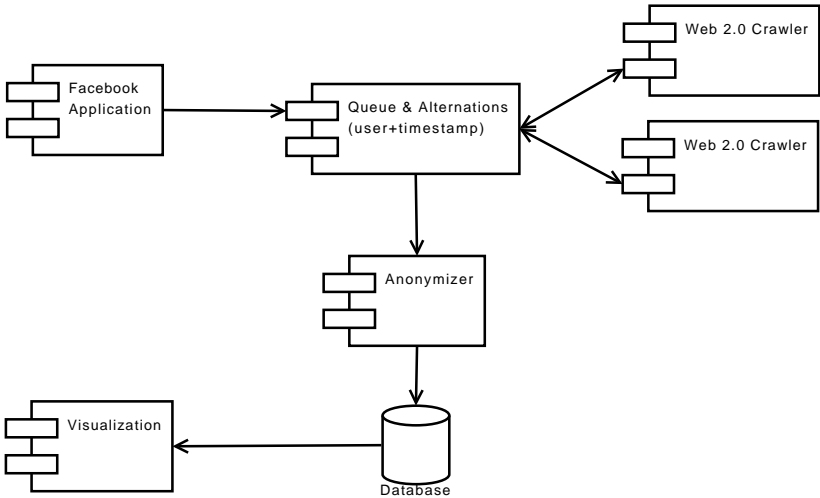
- ▶ Integration in den Facebook Data-Stream, Annotation er erhaltenen Daten
- ▶ Anonymisieren aller personenbezogenen Daten - Iteration über mögliche Benutzernamen; Schutz der Privatsphäre
- ▶ Recherche: mögliche nützliche Datenquellen (technorati, del.icio.us, flickr, digg, Amazon Reviews, IMDB Reviews, ...)
- ▶ Implementierung von zwei dieser Quellen und Integration der Quellen ins Datenschema

# Administration

- ▶ Web Applikation zum Zugriff auf das Data Repository
- ▶ Visualisierungen:
  - ▶ Geographische Verteilung der Benutzer
  - ▶ Beziehungen zwischen den Nutzern (Tabelle oder grafisch, ...)
- ▶ Abfragen anhand der gespeicherten Attribute



# Komponenten



# Anonymisierung der Benutzerdaten

Durch Bildung eines Hash-Wertes aus der Kombination von Benutzername und aktueller Zeit:

```
1    <?php
2        $s = $user . "." . time();
3        $userId = sha1($s);
4    ?>
```

# Facebook Application

Vorgangsweise:

- ▶ Facebook:

- ▶ Ausgangspunkt: <http://www.facebook.com/developers/>
- ▶ Neue Applikation erstellen
- ▶ Canvas
  - URL: [http://www.facebook.com/sql\\_example/](http://www.facebook.com/sql_example/)
  - Callback URL:  
<http://xmdimrill.ai.wu.ac.at/aweichse/projects/2009s/fbA/>

- ▶ xmdimrill:

- ▶ Example Code + Library auf Dimrill kopieren.
- ▶ Applikation mittels Canvas URL oder Canvas Callback URL testen.

## Web 2.0 Datenquellen

- ▶ Technologie: anwendungsspezifisch
- ▶ Beginn: Recherche und entsprechende Dokumentation
- ▶ Integration von mindestens einer externen Quelle

# Geographische Visualisierungen



# Geographische Visualisierungen

- ▶ Verwendung der Google MAPS API
- ▶ Vorgangsweise:
  - ▶ API Key Anfordern
  - ▶ KML File erstellen (dynamisch/statisch)
  - ▶ JavaScript Code in die Webseite integrieren und KML File angeben.

# Geographische Visualisierungen

```
1  <?xml version="1.0" encoding="UTF-8"?>
3  <kml xmlns="http://earth.google.com/kml/2.1">
4  <Document>
5  <name>Example</name>
6  <description>An Example KML File</description>
8  <Placemark>
9    <name>Perth</name>
10   <description>An Example Placemark</description>
11   <Point><coordinates>
12     115.833297729,-31.9333000183
13   </coordinates></Point>
14 </Placemark>
15 </Document>
16 </kml>
```

# Geographische Visualisierungen

```
1 <html xmlns="http://www.w3.org/1999/xhtml"><head>
3 <script
4 src="http://maps.google.com/maps?file=api&v=2&key=XXXX"
5 type="text/javascript"></script>
6 <script type="text/javascript">
7   function load() {
8     if (GBrowserIsCompatible()) {
9       var map = new GMap2(document.getElementById("map"));
10      var geoXml = new GGeoXml("http://t.at/myGeoFile.kml");
12      map.setCenter(new GLatLng(48.2,16.36), 3);
13      map.addControl(new GLargeMapControl());
14      map.addOverlay(geoXml);
15    }
16  }
17 </script>
19 </head>
20 <body onload="load()" onunload="GUnload()">
21   <div id="map" style="width: 680px; height: 400px"></div>
22 </body>
23 </html>
```