

Konzeption und Aufbau eines Crowdsourcing-Portals zur Visualisierung geocodierter Meldungen eines sozialen Nachrichtendienstes am Beispiel von Twitter

Concept and Construction of a Crowdsourcing Portal to Visualize Geocoded Messages of a Social Network Based on the Example of Twitter (English summary)

Stefanie MÄNNL, Mareike WENDTLAND und Martin BECKER

Summary

Maps are changing our world and the view on complex issues in ways we could hardly imagine just a few years ago. Nowadays, the combination of communication channels of Web 2.0 with cloud services and crowdsourcing enables completely new ways of building interactive maps and information visualization e.g. real-time generated information maps. In connection with the spreading number of GPS sensors in mobile devices, also untrained people get the chance to publish GPS supported images, tracks and messages on the internet. Challenges in the enormous potential of usage are shown by the way of structuring, harmonizing and quality checks of the huge amount of collected data. Beyond every crowdsourcing application the basic concept of filtering the amount of information consists of three general steps. The typical process is divided into the elevation of information by the community, as well as the filtering and weighting as the second and third step of collecting information.

The following paper offers the concept, construction, technical design and essential workflow of a prototype, which handles and visualizes geospatial messages of the news broadcast Twitter in a web application. Because of the various scenarios of usage, the development of geo mashups, which is how applications like this are called, experienced an upswing on the internet in the recent years. In comparison to existing examples in the web, the potential of the introduced application lies in the persistent storage of crowdsourcing information in a spatial database and the resulting opportunities of structuring and functional harmonization of the collected data. New and various evaluated potentials could be emerged by linking these information with other spatial data. Another advantage would be the possibility of temporal spatial analyses by using such stored data. The project has shown that the development of a crowdsourcing application could completely be built out of OpenSource software and that the potential of usage is given by various amounts of scenarios e.g. in sustainable land management or landscape architecture.

1 Einleitung

Karten verändern die Welt in einer Art und Weise, wie es vor wenigen Jahren nur schwer vorstellbar war. Informationskanäle des Web 2.0 in Kombination mit Cloud-Diensten und Crowdsourcing ermöglichen völlig neue Formen der Kartenerstellung bspw. aus Echtzeitdaten generierte Informationskarten. In Verbindung mit der sich ständig ausbreitenden Zahl von mobilen, GPS-gestützten Erfassungssensoren können auch wenig ausgebildete Nutzer der Technik raumgestützte Informationen wie z. B. GPS-gestützte Fotografien, GPS-Tracks oder auf GPS- basierende Textmeldungen erstellen und über das Internet verfügbar machen. Die Herausforderung ist dabei, das riesige Potenzial dieser Daten unterschiedlicher Art und Qualität durch Strukturierung und Verknüpfung in nutzbare Information zu wandeln. Positive Anwendungsbeispiele, u. a. im Crisis Mapping mit der Bündelung und Visualisierung von Informationen aus Krisengebieten in Karten (FRICKEL, 2010), verdeutlichen die Eignung und das Potenzial die Daten aus solchen Erfassungsmethoden auszuwerten. Der vorliegende Artikel stellt das Konzept und den Aufbau einer Anwendung dar, welche Meldungen des Nachrichtendienstes Twitter geocodiert und in einer prototypischen Webapplikation visualisiert. Das Projekt wurde im Rahmen einer Bachelorarbeit an der Hochschule Anhalt am Institut für Geoinformation und Vermessung im Jahr 2011 realisiert. Die Ziele waren:

- Beschreibung des wissenschaftlichen und technischen Standes in Bezug auf Geoapplikationen im Web 2.0.
- Erstellung von Konzepten für den Aufbau einer Internetplattform (Client-Server Architektur).
- Prototypische Umsetzung des Crowdsourcing-Portals mit der Visualisierung geocodierter Twitter-Meldungen.

2 Stand der Wissenschaft

Der Begriff Crowdsourcing wurde erstmals im Jahr 2006 geprägt und reflektiert sich aus den Wörtern „Crowd“, einer unbestimmten Masse von Menschen und „Outsourcing“, der Auslagerung von Prozessen und Aufgaben ins Internet (HOWE, 2008). Crowdsourcing als „Baustein einer virtuellen Wertschöpfungskette“ (ROSKOS, 2010) ist ein Prozess, welcher durch das Nutzen des Internets, orts- und zeitunabhängige Inputs seitens der Nutzer zulässt. Die Einbindung der „Kraft, der Kreativität der Vielfalt der Menschen im Internet mit deren Ideen, kulturellen Backgrounds, beruflichen Qualifikationen und Erfahrungen“ führt zu neuen Ergebnissen und Mehrwerten, welche bisher nicht möglich waren (ROSKOS, 2010). Eine grundsätzliche Voraussetzung für das Funktionieren stellen die Informationskanäle des Web 2.0 dar (ROSKOS, 2009). Speziell für die freiwillige, kollektive Erfassung von Geodaten hat sich der Begriff „Volunteered Geographic Information“ (VGI) etabliert, welcher Menschen als „human sensors“ beschreibt (GOODCHILD, 2007). Aufgegriffen und erweitert wurde der Begriff durch (RAMM, 2009), der den Crowdsourcing-Begriff strukturiert und Beispielprojekte zuordnet (vgl. Abb. 1).

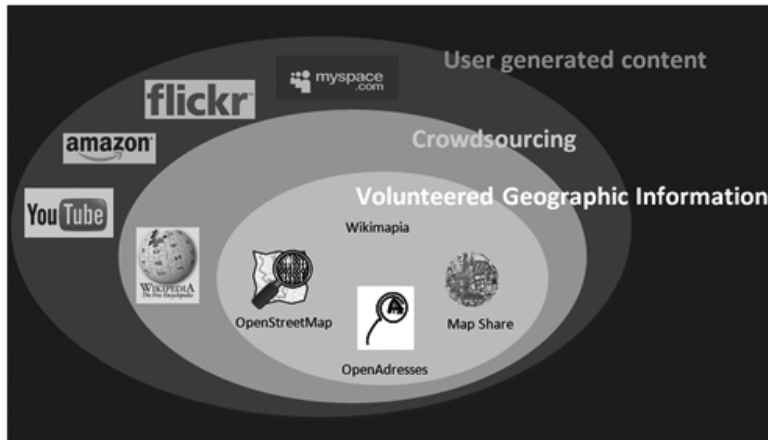


Abb. 1: Strukturierung des Crowdsourcing Begriffs (RAMM, 2009)

Als Standards für Übertragungskanäle von Geodaten haben sich GeoSMS¹, GeoRSS² und Twitter³ etabliert, welche sich u. a. in Projekten wie ushahidi⁴ und dem USGS Twitter Earthquake Detector⁵ bewährt haben. Für die prototypische Umsetzung wurde sich auf den Nachrichtendienst Twitter beschränkt. Beim 2006 gegründeten Nachrichtendienst Twitter steht bei den Anwendern die zentrale Frage „Was machst du gerade?“ im Mittelpunkt. Diese Echtzeitempfindungen oder -meldungen, welche sowohl von Privatpersonen, als auch Institutionen und Unternehmen zur Kommunikation von aktuellen Ereignissen genutzt werden, bilden mit der maximal 140 Zeichen umfassenden Nachricht – auch bezeichnet als Microblogging – eine prägnante und fest strukturierte Statusmeldung für Nutzer oder auch Anwendungsentwickler.

Der Umstand, dass die geposteten Feststellungen, Fakten, Erfahrungen und Ereignisse der Nutzer mit dem gesamten weltumspannenden, sozialen Netzwerk geteilt werden und anschließend durch die offengelegte Entwicklerschnittstelle analysiert und ausgewertet werden können, definiert das Netzwerk bereits als Crowdsourcing-Portal – einer Plattform zur Integration, Auswertung und benutzerspezifischen Anpassung von Inhalten, Diensten und Funktionen zu bereitgestellten Informationen im Web (vgl. KIRCHHOF, 2004). In Verbindung mit GPS-Informationen führte diese Technologie bereits zu einer Reihe von Applikationen im Internet – sogenannten GeoMashups – welche die Visualisierung von Ereignissen mit Raumbezug implizieren. Beispiele dafür bilden die Anwendungen My TWEET Map⁶, Middle East Protest – Tweets Mapped⁷ oder twitvision⁸.

¹ Siehe <http://www.opengeospatial.org/standards/opengeosms>

² Siehe http://georss.org/Main_Page

³ Siehe <https://dev.twitter.com>

⁴ Siehe <http://www.ushahidi.com>

⁵ Siehe <http://recovery.doi.gov/press/us-geological-survey-twitter-earthquake-detector-ted>

⁶ Siehe <http://www.mytweetmap.com>

⁷ Siehe <http://www.mibazaar.com/meprotests.html>

⁸ Siehe <http://twitvision.com>

Eine umfangreiche Reflexion über den Nutzen von Crowdsourcing-Technologien im Umfeld der Landschaftsarchitektur zeigt (HEWITT, 2012). Auf Grundlage von weltweiten Analysen von Webseiten wurden Schwerpunktthemen und Technologien aus den Bereichen Architektur, Landschaftsarchitektur und Städtebau identifiziert um daraus soziale und geographische Hotspots und Betrachtungen zu den Themengebieten definieren zu können.

3 Konzeption der Webanwendung

Für die Konzeption der Webapplikation namens „geo.tweet“ wurden unterschiedliche Bearbeitungsbereiche identifiziert, welche in einen Arbeitsworkflow (siehe Abb. 2) gemündet

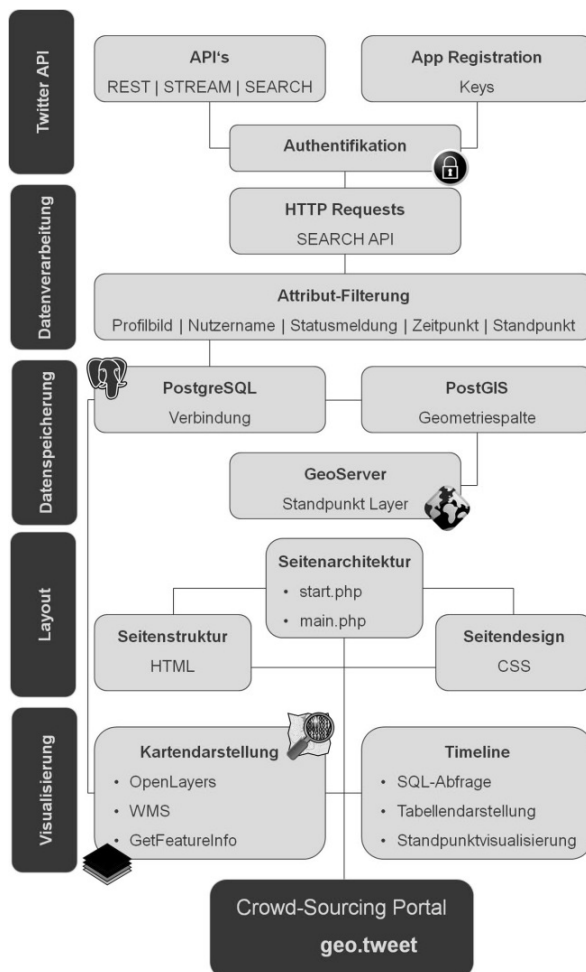


Abb. 2: Workflow zur prototypischen Umsetzung der Internetplattform (eigene Darstellung)

sind. Nachfolgend werden die Herangehensweisen in der Entwicklung für die durchgeführten Arbeitsbereiche erläutert. Hauptansatzpunkte waren dabei die Twitter API sowie Aufbau und Design, welches sich wiederum in die Bereiche Datenverarbeitung, Datenspeicherung, Layout und Visualisierung untergliedert.

3.1 Twitter API

Das „Twitter Application Programming Interface“ ist eine von Twitter offengelegte Programmierschnittstelle mit der Möglichkeit, auf Twitter-Daten zugreifen zu können.

Der Zugriff auf die Twitter API ist mit allen gängigen Programmiersprachen und den dafür existierenden Bibliotheken möglich. Für die Umsetzung des Projekts wurde die Programmiersprache PHP und die dazugehörige Bibliothek von Abraham Williams⁹ genutzt. Der Twitter Web Service ermöglicht die Nutzung von drei unterschiedlichen API's. Die REST (Representational-State-Transfer) API stellt alle Basisfunktionen bereit, die auch unter <http://twitter.com> zur Verfügung stehen. Hierzu zählen das Posten von Tweets, das Ansehen einer Timeline (Liste von Statusmeldungen) sowie die Möglichkeit, anderen Nutzern zu folgen oder sie der eigenen Freundesliste hinzuzufügen. Um die Twitter-Suche zu ermöglichen, wird die SEARCH API verwendet. Nach WEIGERT (2010) „erscheinen neue Tweets nicht unmittelbar (...), sondern werden in Intervallen von einigen Sekunden oder manchmal auch Minuten in 'Paketen' aktualisiert.“ Diesen Effekt soll die im Sommer 2010 erschienene STREAM API verhindern und Tweets in Echtzeit liefern.

Aus Gründen der Sicherheit und Überwachung ist es für Entwickler notwendig die eigenen Applikationen bei Twitter zu registrieren, um nach dem Erhalt eines Authentifizierungsschlüssels den Zugriff auf die Daten zu gewährleisten. 2010 gab es dazu eine Umstellung von der einfachen Basic Authentication auf die komplexere Authentication. Vormalig reichten Benutzername und Passwort zur Identifikation aus, welche jedoch mit dem neuen, sicheren Einloggen über den erwähnten Schlüssel angepasst wurde. Demnach werden die Zugangsdaten des Nutzers der Applikation nicht mitgeteilt, sodass keine Speicherung dieser Daten erfolgen kann.

3.2 Aufbau und Layout

Datenverarbeitung

Mittels HTTP-Abfragen in der SEARCH API können über spezifische Abfrageparameter z. B. Gebietsabhängigkeit gefiltert werden. Hierfür wurde die SEARCH API verwendet mit der unter anderem gebietsabhängige Filterungen möglich sind. Der Befehl zur Abfrage lautet „geocode“ und beinhaltet die Angabe von Längen- und Breitengrad sowie dem Radius. Des Weiteren besteht die Möglichkeit, nach themenbezogenen Hashtags zu filtern. Eine Beispielabfrage sieht folgendermaßen aus:

```
http://search.twitter.com/search.json?q=&tag=DLA&geocode=52%2C12%2C100km
```

Das Resultat eines Requests sind 15 Nutzer-Profile, dessen Informationen separat abgefragt werden können. Für das Projekt relevante Angaben sind das Profilbild, der Name des Nutzers, der Zeitpunkt, der jeweilige Tweet und der Standort. Letzteres bildet das Hauptkriterium zur Filterung, falls der Standort in Koordinatenform vorliegt.

⁹ Siehe <https://github.com/abraham/twitteroauth>

Datenspeicherung

Nach der Filterung der Daten ist eine Speicherung in einer Datenbank vorgesehen. Dazu wurde die Open-Source Software PostgreSQL in Verbindung mit PostGIS genutzt. Um Standpunkte in einer Karte visualisieren zu können, wird ein MapServer benötigt. Allein durch die standardmäßige PostGIS-Anbindung erwies sich der GeoServer als vorteilhaft und ermöglichte eine unkomplizierte Verbindung zwischen Datenbank und MapServer. Des Weiteren ist das Erstellen eigener Styles möglich, das für die Erzeugung eines Symbols für die Standpunkte benötigt wurde.

Layout

Die grundlegende Struktur des Portals beschränkt sich auf zwei Seiten. Nachdem sich der Nutzer auf der Startseite mit seinem Twitter-Log-in angemeldet hat, wird er auf die Hauptseite zur Datenvisualisierung geleitet. Neben der Funktionalität wurde auf eine ansprechende Gestaltung des Layouts geachtet. Hierfür ist die Trennung von Struktur und Design ausschlaggebend. Während die logische Struktur der Seite über HTML realisiert wird, bedient man sich beim Design der Cascading Style Sheets.

Visualisierung

Die Informationsvisualisierung bezieht sich einerseits auf die Darstellung einer Karte mit den geocodierten Meldungen und andererseits auf die einer Liste von Statusmeldungen, in einer sogenannten Timeline. Für die Kartendarstellung wird die JavaScript-Bibliothek OpenLayers und der freie Kartendienst OpenStreetMap genutzt. Die im GeoServer vorgehaltenen Standorte der Statusmeldungen werden mittels eines Web Map Service (WMS) in die Karte integriert. Für die Anzeige von Informationen über einen bestimmten Punkt in der Karte ist die getFeatureInfo-Abfrage des WMS zuständig. Hierbei wird eine Abfrage-URL generiert und an den GeoServer gesendet. Über einen Proxy-Server lassen sich die Rückgabewerte empfangen, die in einem Popup-Fenster visualisiert werden. Für die Timeline-Darstellung werden die Daten (Profilbild, Koordinaten, Nutzernamen, Status-Meldung und Zeitstempel) aus der Datenbank abgefragt und neben der Karte in Tabellenform, zeitlich absteigend dargestellt.

4 Visualisierungsergebnis

Das Endprodukt des erstellten Crowdsourcing-Portals geo.tweet ist in Abbildung 3 ersicht-lich. Im Header-Bereich befindet sich ein Info-Symbol, welches bei Klick Auskunft über das Projekt gibt. Im Footer-Bereich wird über Icons auf verwendete Technologien verwiesen, die zusätzlich mit einem Hyperlink zur entsprechenden Homepage versehen sind. Der linke Content-Bereich beinhaltet die Kartendarstellung mit der Standpunktvisualisierung der geocodierten Twitter-Meldungen. Ergänzend dazu ist im rechten Contentbereich die Timeline als interaktive Tabelle dargestellt. Um Tabelleneinträge lokalisieren zu können, wird bei Klick auf eine Zelle der dazugehörige Standpunkt in der Karte rot markiert. In Analogie dazu besteht ebenso die Möglichkeit, die Standpunkte in der Karte zu identifizieren. Wählt man einen Standpunkt in der Karte, so erscheint ein Popup-Fenster mit Nutzernamen und Statusmeldung.



Abb. 3: Darstellung des Crowdsourcing-Portal „geo.tweet“ im Browser (eigene Darstellung)

5 Zusammenfassung und Ausblick

Crowdsourcing ermöglicht durch die Vielzahl an Nutzern, in kurzer Zeit mit vergleichbar geringen Kosten eine enorme Fülle an Informationen zu sammeln. Der Arbeitsaufwand für den einzelnen innerhalb der Community ist gering, da ohne große Hilfsmittel mitgewirkt werden kann. Nachteilig ist der Umstand, dass nicht jede Information als sinnvoll und brauchbar angesehen wird oder Fehler entstehen – ob mutwillig oder ungewollt. Hierfür sind bei einer projektbezogenen Umsetzung Vorgehensweisen zur Qualitätssicherung zu schaffen. Ansätze hierbei sind das Erkennen von Dubletten, das Filtern über Hashtags sowie das Ignorieren von Tweets ohne Angabe exakter Koordinaten.

Eine schematische Darstellung der grundlegenden Funktionsweise der Anwendung ist in Abbildung 4 ersichtlich. Die Community gibt über Twitter Statusmeldungen ab. Die Funktionen des Portals ermöglichen es diese Meldungen zu filtern und in einer Datenbank zu speichern. Ruft der Nutzer geo.tweet auf, erfolgt eine Anmeldung über die Log-in-Daten von Twitter. Anschließend werden die Datensätze aus der Datenbank mithilfe einer Web-Mapping-Anwendung im Portal dargestellt.

Verbesserungen des Portals können beispielsweise durch eine Optimierung der realisierten WMS-getFeatureInfo Abfrage geschaffen werden. Anstatt eines Popup-Fensters würde bei Klick auf einen Standpunkt in der Karte, der betreffende Timeline-Eintrag markiert und darauf gescrollt werden. Eine weitere denkbare Zusatzfunktion wäre außerdem eine Zeitskala – so könnten Tweets manuell in Abhängigkeit von der Zeit ein- und ausgeblendet werden. Vorteilhaft wäre außerdem eine Realisierung weitergehenden Auswertemöglichkeiten in Form von Grafiken oder Diagrammen.

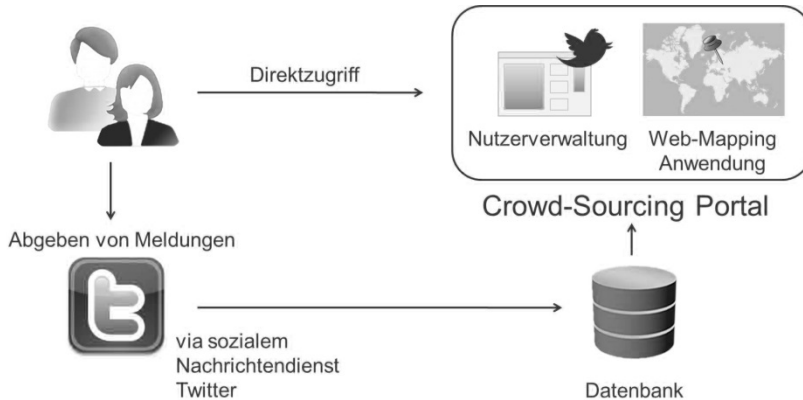


Abb. 4: Darstellung der grundlegenden Funktionsweise des Crowdsourcing-Portals (eigene Darstellung)

Da es sich hierbei um einen Prototyp handelt, ist das Portal nicht online erreichbar. Das Potenzial zur Nutzung der Plattform zeigt sich einerseits in der rasanten Entwicklung der Bereitstellung von Informationskarten mit Echtzeitdaten im Internet, und andererseits durch die Vielzahl an GPS-Sensoren in heutigen, mobilen Endgeräten als Grundvoraussetzung der Datenerfassung. Daraus ergeben sich vielfältige Anwendungsszenarien, sodass die Plattform auch in weiteren Hochschulprojekten ausgebaut und genutzt werden wird. Der innovative Faktor und große Vorteil gegenüber vergleichbaren Anwendungen besteht in der persistenten Speicherung der Crowdsourcing-Informationen in einer räumlichen Datenbank mit den sich daraus resultierenden Möglichkeiten der Strukturierung und fachlichen Harmonisierung der Daten sowie der Ansatz der raumzeitlichen Abbildung der Meldungen über einen individuell auswählbaren Zeitraum. Ähnliche Applikationen verfolgen nur das Konzept der Realtime Darstellung von Informationen. Jedoch können sich mit der Verknüpfung zu weiteren Geodaten neue, vielfältige und noch nicht absehbare Auswertungspotenziale für verschiedene Nutzergruppen herauskristallisieren.

Insbesondere für die Landschaftsarchitektur ergibt sich mit dem Portal eine Vielzahl von denkbaren Anwendungsfällen. So könnten landschaftsbezogene Twitter-Meldungen über Hashtags gefiltert und anschließend räumlich visualisiert werden, sodass möglicherweise regional unterschiedliche Diskussionsthemen und Trends rund um die Landschaftsarchitektur erkennbar wären. Auch in der Phase des Nachgangs der Fertigstellung eines Projektes, bspw. in der Umgestaltung einer Parkanlage, ließen sich neben der ökologischen Neugestaltung auch die sozialen Aspekte sammeln und untersuchen. Nachrichten mit einem bestimmten Stichwort oder der räumlichen Eingrenzung rund um die Parkanlage können extrahiert werden und in die direkte Bewertung des Projektes eingehen. Umgekehrt können im Vorfeld einer geplanten Gestaltungsmaßnahme die Tweets der Leute räumlich eingegrenzt gefiltert und die Informationen in einem Monitoringverfahren erworben werden. Diese Art von Nutzungs- und Stimmungsanalyse kann wiederum in landschaftsarchitektonische Planungen einfließen und die Zustimmung zu einem Projekt steigern.

Literatur

- Frickel, C. (2010), *Krisen-Plattform Ushahidi – Datensammler für den guten Zweck*.
http://www.focus.de/digital/internet/tid-18595/krisen-plattform-ushahidi-datensammler-fuer-den-guten-zweck_aid_518248.html (15.02.2013).
- Goodchild, M. F. (2007), *Citizens as sensors: The world of volunteered geography*.
http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Goodchild_VGI2007.pdf
(14.02.2013).
- Hewitt, R. (2012), *Finding Multi-Centers: Using Crowdsourcing Technologies to Define Communities of Landscape Architecture*. Council of Educators in Landscape Architecture.
- Howe, J. (2008), *Crowdsourcing: Why the power of the crowd is driving the future of business*. Crown Publishing Group.
- Kirchhof, A. et al. (2004), *Was ist ein Portal?* http://pub-379.bi.fraunhofer.de/Images/Was_ist_ein_Portal_tcm379-129040.pdf (15.02.2013).
- Ramm, F. (2009), *Crowdsourcing Geodata*. <http://www.geofabrik.de/media/2009-09-08-crowdsourcing-geodata.pdf> (17.06.2011).
- Roskos, M. (2009), *Was ist Crowdsourcing? 12 Eckpunkte, die den Rahmen abstecken*.
<http://www.socialnetworkstrategien.de/2009/08/was-ist-crowdsourcing-12-eckpunkte-die-den-rahmen-abstecken> (17.06.2011).
- Roskos, M. (2010), *Crowdsourcing – Mehrwerte, Chancen, Definition*.
<http://www.socialnetworkstrategien.de/2010/07/crowdsourcing-mehrwerte-chancen-definition> (15.02.2013).
- Weigert, M. (2010), *Anschnallen für den Echtzeitstream*. <http://netzwertig.com/2010/07/30/twitters-neue-user-streams-api-anschnallen-fuer-den-echtzeitstream/> (15.08.2011).