

Social Media as Sensors

Nikolai Bock, Klaus Böhm

Einleitung

Dieser Beitrag beschäftigt sich damit, die Analyse von Social Media Daten als eine Art „Human Sensor“ in einem Sensor Observation Service (SOS) bereitzustellen.

Hintergrund sind die Möglichkeit und das Potential, welche die Analyse von Social Media Daten aus den unterschiedlichen Netzwerken, wie Twitter, Facebook, Flickr oder Youtube bieten. Vor allem im Bereich der Marketinganalyse und des Katastrophenmanagements ist dies ein immer häufiger genutztes Werkzeug. Die Relevanz des Raumbezugs bei Social Media zeigt der Prototyp einer Arbeit, welcher die Ergebnisse einer gefilterten Twitter-Suche über einfache kartographische Darstellungsformen räumlich visualisiert (<http://tweet-map.fh-mainz.de>). Durch den Einsatz von Natural Language Processing (NLP), z.B. bei der Sentimental Analyse, können Social Media Daten hinsichtlich des Inhaltes untersucht werden. Aus diesen Untersuchungen lassen sich detailliertere Informationen, ähnlich zu physikalisch gemessenen Phänomenen, gewinnen. Dadurch lassen sich Social Media Daten zu einem Human Sensor wandeln. Unter Human Sensors versteht man ein Messmodell, in dem Menschen neben physikalischen Messungen, wie z.B durch Fitnessarmbänder, auch subjektive „Messungen“ wie Sinneseindrücke, Empfindungen oder persönliche Beobachtungen beitragen (vgl. Resch et al. [1]).

Die Nutzung von Social Media basierten Human Sensors gewinnt zudem an Bedeutung, wenn diese Daten in den Zusammenhang mit anderen Datenquellen, wie z.B. Umweltinformationen, gesetzt werden können. Eine Herausforderung dieser Zusammenführung ist die Interoperabilität zwischen den Daten. Diese kann durch den Einsatz von Standards erreicht werden.

Der hier vorgestellte Ansatz basiert auf dem Sensor Observation Service (SOS) aus dem Bereich des Sensor Web Enablements (SWE). Die Kombination der oben genannten Bereiche (Raum/Zeit, NLP, Human Sensor) ergibt hierfür eine Sensordatenquelle in einem SOS-Netzwerk. In diesem Netzwerk können die so verwalteten Informationen mit anderen räumlichen Sensordaten kombiniert werden.

Der Beitrag fokussiert dabei die Konzeption und Entwicklung des Systems FlexSensor, welches die Möglichkeit bietet, Daten aus Social Media Netzwerken (zunächst beispielhaft mit Twitter) und deren Analyse (NLP), als Beobachtungen in einem Sensornetzwerk als SOS zur Verfügung zu stellen. Zudem werden Ansätze gezeigt, die gesammelten Daten räumlich und zeitlich zu visualisieren.

Social Media Analyse as Sensor

Um in den Gesamtkontext des Beitrags einzuführen, werden nachfolgend die Aspekte Social Media, Social Media Analyse und Sensoren bzw. Sensornetzwerke einleitend erläutert.

Social Media

Social Media umschreibt grob Inhalte [2], bzw. Anwendungen zur Erstellung, Gestaltung und Austausch dieser Inhalte [3], welche „Benutzer-generiert“ sind. Social Media bedient sich hierbei der Techniken des Web 2.0. Betrachtet man diese Definition, gibt es eine sehr große Anzahl an Daten und Diensten, welche unter dem Begriff Social Media fallen. Einige der Bekanntesten sind Twitter, Facebook, Flickr, Instagram, Youtube, Wikipedia oder Google+. Aber auch OpenStreetMap, welches als klassisches Beispiel für „Volunteered Geographic Information“ (VGI) [4] gesehen wird, kann somit als Social Media verstanden werden.

Analyse (Social Media)

Die Analyse von Social Media lässt sich in zwei Bereiche aufteilen. Zum Einen lassen sich die Strukturen des Netzes untersuchen. Typischerweise spricht man dann von „Social Network Analysis“ (SNA). Zum Anderen lässt sich der Inhalt genauer analysieren. Dies geschieht derzeit hauptsächlich mit textlichem Inhalt durch natürlichsprachliche Methoden, bekannt als „Natural Language Processing (NLP)“.

Eine dieser NLP-Methoden ist die Sentimental-Analyse, welche auch in der Arbeit genutzt werden soll. Dieser Algorithmus analysiert, ob ein Text positive, neutrale oder negative Emotionen ausdrückt. Es ist möglich, diese Information auf einen gesamten Text, oder auf bestimmte Subjekte oder Objekte im Text, zu beziehen.

Sensoren / Sensornetzwerke

Der Grundgedanke, die Analysedaten als Sensorinformation bereitzustellen, kommt aus dem Forschungsbereich „People as Sensors“. Hierbei handelt es sich um ein Messmodell (Resch et al. [1]), in dem Menschen neben physikalischen Messungen, wie z.B durch Fitnessarmbänder, auch subjektive „Messungen“, wie Sinneseindrücke, Empfindungen oder persönliche Beobachtungen, beitragen. In der Literatur findet man im selben Kontext auch die Begriffe „Human As Sensors“ (Forrest [5]) oder „Citizen As Sensors“ (Goodchild [4]).

Die Sammlung der subjektiven Messungen über dezidierte Anwendungen hat den Vorteil, dass die Informationen sehr gezielt und somit einfacher auswertbar sind. Jedoch ist in der Regel die Verbreitung, und somit eine hohe Messdichte, limitiert. Verwendet man Social Media als Information, sind diese Daten in einer höheren Dichte verfügbar. Allerdings sind die Anforderungen deutlich höher bei der Analyse, da der Kontext zu einer Beobachtung nicht immer eindeutig vorhanden ist. Allerdings sind durch immer neue Methoden bessere Qualitäten in der Analyse möglich. Somit lassen sich aus den unterschiedlichen Sensordatenquellen später Ansätze aus dem Collective Sensing (Resch [6]) oder Collective Human Behavior Patterns (Sagl et al. [7]) umsetzen.

Die Verwendung von OGC-Standards aus dem Bereich des SWE, wie dem SOS, bringt den Vorteil, dass diese Daten mit anderen heterogenen Sensorinformationen verknüpft werden können.

FlexSensor Ansatz

In den folgenden Kapiteln soll der gewählte Ansatz näher beschrieben werden. Hierzu werden zunächst Anforderungen erläutert. Anschließend wird das Konzept kurz vorgestellt. Abschließend werden zwei Komponenten des Konzepts detaillierter beschrieben.

Anforderungen für FlexSensor

Aus der Problemstellung, Social Media Analysedaten interoperabel mit anderen (Geo-) Daten nutzen zu können, ergaben sich folgende Anforderungen:

- Social Media Analyse Information über eine standardisierte Schnittstelle zur Verfügung stellen
- Flexible Möglichkeit, unterschiedliche Social Media Quellen, (Geo-) Dienste und Analysen kombinieren zu können, um sie der Schnittstelle zuführen zu können
- Berücksichtigung späterer Skalierbarkeit im Kontext von Big Data.
- Benutzeroberfläche, welche unterschiedliche Datenquellen visualisieren, verwalten und eine Interaktion zwischen diesen herstellen kann
- Benutzeroberfläche mit unterschiedlichen Möglichkeiten, Analyseprozesse zu initiieren

Systemkonzept

Auf Basis der Anforderungen wurde ein Systemkonzept entwickelt, welches, wie in Abbildung 1 zu sehen ist, aus 3 Hauptkomponenten besteht.

Das Konzept sieht vor, dass die Social Media Analyse Daten in einem SOS bereitgestellt werden. Somit ist ein standardisierter Zugriff gewährleistet. Zu den 3 Hauptkomponenten gehören zum Einen, als ein wichtiger Bestandteil, die (Geo-) Dienste, welche als Datenquelle, zur Datenverarbeitung und zur Datenspeicherung verwendet werden. Hierzu zählen die bekannten OGC-Standards, wie Web Map Service (WMS), oder der bereits erwähnte SOS für die Sensordaten. Als Datenquelle für Social Media wurde zudem die Twitter API verwendet. Weiterhin sieht das Konzept für die Verarbeitung der Quelldaten zu Sensorinformationen in einem Sensornetz ein System vor, welches es ermöglicht, die Analysen als eine Art Workflow zu steuern. Dank der Umsetzung mit Hilfe von Spring XD [8] ist dies möglich und eine hohe Flexibilität und Skalierbarkeit gewährleistet. Auch Big Data Ansätze sind bereits in Spring XD vorgesehen. Als dritte Komponente fehlt noch die Benutzerschnittstelle, welche als ein Portal mittels Liferay [9] zur Verfügung gestellt wird. Sie regelt sämtliche Interaktion mit dem Benutzer. Alle 3 Bereiche können dabei physisch komplett voneinander getrennt werden, wie in einer Service-Orientierten-Architektur (SOA) üblich.

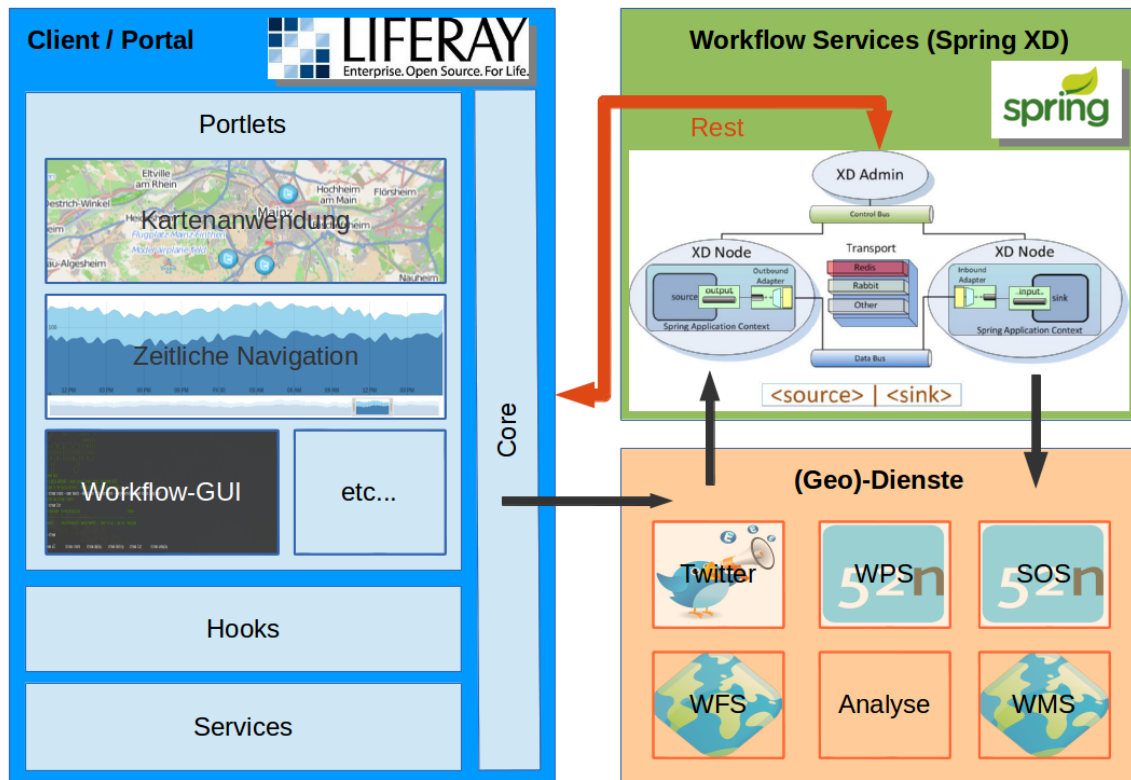


Abbildung 1: Systemkonzept des FlexSensor-Projektes

Das Workflow-System zur Verarbeitung von Twitter Analyse Informationen in einen SOS (Twitter2SOS) und das Portal sollen folgend näher beschrieben werden.

Twitter2SOS System

Um die Analyse-Daten von Social Media in einem Sensor-Netzwerk mit weiteren Daten verknüpfen zu können, müssen Sie als Sensorinformationen in einem SOS bereitgestellt werden. Hierzu sind mehrere Schritte nötig. Im FlexSensor-Projekt wurde dies zunächst mit Daten aus der Twitter API umgesetzt. Um allerdings bzgl. der Datenquellen, der Prozessschritte und des Datenspeichers ein flexibles System anzubieten, wurde entschieden, mit Datenströmen in Spring XD zu arbeiten. Spring XD steht für Extreme Data und ist somit auf Big Data ausgelegt. Spring XD ist eine Ausführungsumgebung zur einfacheren Kombination von anderen Spring Projekten, wie Spring Integration [10] und Spring Batch [11].

Bei FlexSensor kommen derzeit nur Bereiche von Spring Integration zum Einsatz. Spring Integration ist ein Projekt, welches die „Enterprise Integration Patterns“ [12] unterstützt. Hierfür wird mit Adaptern gearbeitet, welche über ein Messaging-System miteinander kommunizieren. Somit lassen sich die Komponenten sowohl programmiertechnisch als auch physikalisch stark entkoppeln. Allerdings wird der gesamte Workflow über eine XML-Konfiguration definiert, was dynamische Flexibilität erschwert. Dies vereinfacht Spring XD unter Anderem.

Social Media as Sensors

Folgend wird anhand von Abbildung 2 erläutert, wie der Workflow zur Verarbeitung der Twitter Informationen mit Hilfe der Spring Komponenten und selbst entwickelter Module bzw. Adapter umgesetzt wurde.

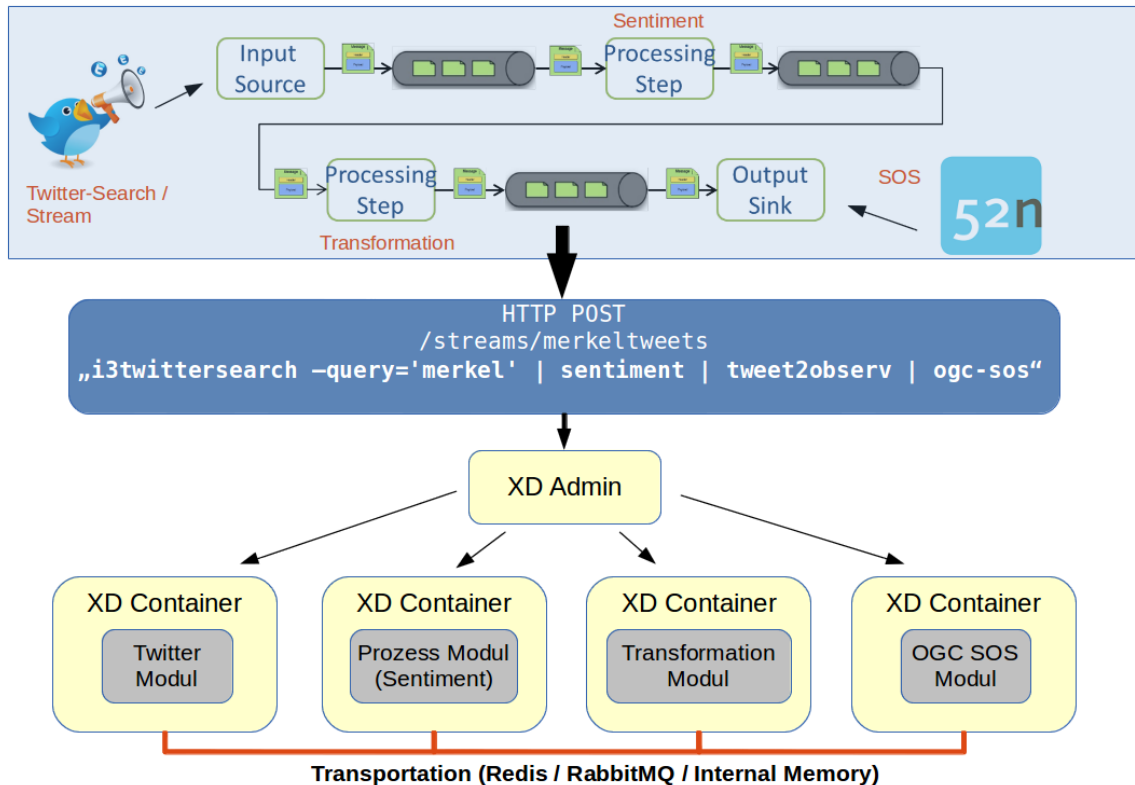


Abbildung 2: Von Twitter in den SOS - von Spring Integration nach Spring XD

Im oberen Teil ist eine vereinfachte Prozesskette des Workflows zu sehen. Die einzelnen Arbeitsschritte (Twitter Search/Stream Input Source, Sentiment, Transformation, SOS Output Sink) sind als Spring Integration Adapter umgesetzt. Die Kommunikation funktioniert über „Messages“, welche über „Channel“ geschickt werden. In Spring XD werden die einzelnen Schritte als Module zur Vereinfachung bereitgestellt. In einer aus der UNIX-Shell bekannten Schreibweise ist es so möglich, Datenströme zu definieren. Die Definition besteht nun nur noch aus den Modulen (z.B. i3twittersearch), deren Optionen (z.B. --query='merkel') und den „Pipes“, welche die Kanäle zwischen den Adaptern repräsentieren. Da es konfigurierbar ist, wie der Transport funktioniert (Interner Speicher, Redis [13] oder RabbitMQ [14]), können sämtliche Container, in denen die Module laufen, auch physikalisch voneinander getrennt sein. Die Steuerung übernimmt der Admin Service welcher über eine REST-Schnittstelle [15] bedient wird. Mit ihm lassen sich Workflows anlegen, starten, stoppen und löschen.

Um die Module definieren zu können, mussten zunächst die Spring Integration Adapter zur Verfügung gestellt werden. Das Spring Integration Projekt selbst liefert schon einige vorgefertigte Adapter, wie JDBC, File oder HTTP. Es gibt auch einen Adapter für die Twitter API, welche sich dem Twitter-Template aus dem Spring Social Projekt [16] bedient. Allerdings mussten hier einige Anpassungen gemacht werden, um auch die räumlichen Daten bei einer Suche zu erhalten.

Social Media as Sensors

Der zweite, wichtige zu entwickelnde Adapter ist für die Anbindung an den SOS zuständig. Hierzu wurde ein eigenes Spring Template entwickelt, welches als Schnittstelle zum SOS genutzt werden kann. Nun konnte ein Integration Adapter erstellt werden, welcher das SOS Template nutzt. Als SOS-Service wurde das Framework von 52°north [17] verwendet.

Zudem sind derzeit Lösungen zur Analyse der Daten in Arbeit. Erste Ansätze greifen auf externe Dienste zurück, welche bereits Methoden des NLP bereitstellen. Da diese Dienste hinsichtlich ihrer Nutzbarkeit allerdings Einschränkungen haben, ist geplant eigene Lösungen bereitzustellen. In FlexSensor wird beispielhaft eine Sentiment-Analyse, wie oben beschrieben, umgesetzt.

Portal

Zur Umsetzung der Benutzerschnittstelle wird die Portalsoftware Liferay eingesetzt. Mit Liferay lassen sich Portalsysteme aufbauen, welche über verschiedene Komponenten erweitert werden können. Als zentrales Element sind die Portlets zu sehen, welche nach der Portlet Specification 2.0 (JSR 286) [18] standardisiert sind. Dank der Portlets lassen sich die Benutzeroberflächen flexibel zusammenbauen. Weitere Elemente sind, wie in Abbildung 1 zu sehen ist, Hooks und Services. Mit Hooks greift man in der Regel in serverseitige Prozesse des Portals ein (z.B. „tue etwas, wenn ein neuer Benutzer angelegt wurde“). Services stellen Dienste bereit, welche Daten mit der Datenbank synchronisieren. Liferay verwendet als Persistenz-Schicht Hibernate und unterstützt somit zahlreiche Datenbankmanagementsysteme (DBMS).

Im Zuge der Arbeit mussten vor allem einige Portlets entwickelt werden. Hierzu gehört ein Portlet für die Kartendarstellung und -verwaltung. Als Framework wird OpenLayers 3 [19] verwendet. Die Kartenanwendung soll unterschiedliche Kartenlayer aus WMS, WFS und SOS-Datenquellen zeigen und verwalten können. Zugehörig ist ein Portlet für die Ebenenverwaltung. Ein weiteres Portlet ist für zeitliche Navigation in den Daten zuständig. Hierbei wird die horizontale Achse als Zeitachse mit einem Navigationsslider verwendet. Somit lässt sich gezielt auf Zeiträume zugreifen, wobei das in die Zeitachse integrierte Diagramm schon Einblicke in die Daten gibt. Das Portlet wird unter anderem mit dem Framework D3js [20] umgesetzt. In einem weiteren Portlet sollen die Workflows aus dem Spring XD System gesteuert werden. Da der Spring XD Admin Server über eine REST-Schnittstelle gesteuert werden kann, wird hierfür ein REST-Client verwendet. Das Portlet erfüllt zunächst die Hauptanforderungen zur Steuerung und wird dann individuell auf den Anwendungsfall (hier Twitter-Sensor) erweitert.

Fazit

Der Beitrag zeigt das FlexSensor-System, welches es ermöglicht, Analysen aus Tweets in einem Sensornetzwerk bereitzustellen. Dank der verwendeten Frameworks wurde darauf geachtet, dass das System flexibel auf andere Daten und Prozesse angepasst werden kann. Der derzeitig entwickelte Prototyp zeigt die Machbarkeit des Ansatzes und wird kontinuierlich weiter entwickelt. Hierbei wird vor allem der Funktionsumfang erweitert. Auch sind Tests im Hinblick auf größere Datenmengen durchzuführen.

Für die weitere Zukunft sind zum Einem bei der Verarbeitung der Social Media Daten noch viele Optimierungsschritte denkbar, welche das Analyseergebnis verbessern oder andere In-

Social Media as Sensors

formationen aus den Daten extrahieren. Auf der Seite der Visualisierung ist zeitlich-räumliche Darstellung der thematischen Information ein Thema, welches man noch stärker bearbeiten kann. Dies kann helfen, die gewonnenen Informationen mit weiteren Sensorinformationen zu verknüpfen und neue Einblicke zu gewinnen. Hinsichtlich der Workflow-Verwaltung sind verschiedene weitere Ansätze im Hinblick auf die unterschiedlichen Anforderungen denkbar.

Kontakt zu den Autoren:

Nikolai Bock
i3mainz – Institut für Raumbezogene Informations- und Messtechnik
Fachhochschule Mainz
Lucy-Hillebrand-Str. 2, 55128 Mainz
nikolai.bock@fh-mainz.de

Prof. Dr. Klaus Böhm
i3mainz – Institut für Raumbezogene Informations- und Messtechnik
Fachhochschule Mainz
Lucy-Hillebrand-Str. 2, 55128 Mainz
klaus.boehm@fh-mainz.de

Literatur

- [1] Resch, Bernd; Mittlböck, Manfred; Kranzer, Simon; Sagl, Günther; Heistracher, Thomas; Blaschke, Thomas: „People as Sensors“ mittels personalisierten Geo-Trackings, *Angewandte Geoinformatik* 2011, p. 683, 2011.
- [2] Atzmüller, Martin: Mining social media; *Informatik Spektrum*, 35(2); p. 132-135, 2012
- [3] Kaplan, Andreas M.; Haenlein, Michael: Users of the world, unite! The challenges and opportunities of social media; *Business Horizons*, 53(1); p. 59-68, 2010
- [4] Goodchild, Michael F.: Citizen as sensors: the world of volunteered geography. *GeoJournal*, 69(4); p. 211-221, 2007
- [5] Forrest, Brady: Humans As Sensors. *LBxJournal*, <http://www.lbxjournal.com/articles/humans-sensors/260057>, 2010 (28.01.2014)
- [6] Resch, Bernd: People as Sensors and Collective Sensing-Contextual Observations Completing Geo-Sensor Network Measurements, *Progress in Location-Based Services – Lecture Notes in Geoinformation and Cartography*, 2013, p. 391-406
- [7] Sagl, Günther; Resch, Bernd; Hawelka, Bartosz; Beinat, Euro: From Social Sensor Data to Collective Human Behavior Patterns – Analysing and Visualising Spatio-Temporal Dynamics in Urban Environments; *GI_Forum 2012: Geovisualization, Society and Learning*, 2012, p. 54-63
- [8] Spring XD: <http://projects.spring.io/spring-xd>
- [9] Liferay Enterprise open source portal and collaboration software: <http://liferay.com>
- [10] Spring Integration: <http://projects.spring.io/spring-integration>
- [11] Spring Batch: <http://projects.spring.io/spring-batch>
- [12] Enterprise Integration Pattern (Hohpe, Gregor; Woolf, Bobby) : <http://www.eaipatterns.com>

Social Media as Sensors

[13] Redis <http://redis.io>

[14] RabbitMQ <http://www.rabbitmq.com>

[15] Wikipedia zu REST: http://de.wikipedia.org/wiki/Representational_State_Transfer

[16] Spring Social: <http://projects.spring.io/spring-social>

[17] 52°north SOS: <http://52north.org/sos>

[18] JSR 286: <https://jcp.org/aboutJava/communityprocess/final/jsr286>

[19] OpenLayers3: <http://ol3js.org>

[20] D3js: <http://d3js.org>