

Ohsome – eine Plattform zur Analyse raumzeitlicher Entwicklungen von OpenStreetMap-Daten für intrinsische Qualitätsbewertungen

Ohsome – An OpenStreetMap History Analytics Platform for Intrinsic Quality Assessments

Michael Auer, Melanie Eckle, Sascha Fendrich, Fabian Kowatsch, Lukas Loos, Sabrina Marx, Martin Raifer, Moritz Schott, Rafael Troilo, Alexander Zipf

Geographisches Institut, Universität Heidelberg · marx@uni-heidelberg.de

Zusammenfassung: Die Ohsome-Plattform ermöglicht es, die vollständige Historie von OpenStreetMap (OSM) weltweit und in sekundlicher Auflösung zu analysieren. So kann die Qualität von OSM aus den Daten selbst, anhand ihrer historischen Entwicklung, bewertet werden. Um verschiedenen Nutzern gerecht zu werden, stehen Schnittstellen auf drei Abstraktionsebenen zur Verfügung. Anhand eines Web-Dashboards zur Analyse der OSM-Aktivitäten in Nepal und eines Analyse-Beispiels zur Vollständigkeit von Hausnummern in Afrika werden die Funktionalitäten und Möglichkeiten der Plattform beispielhaft vorgestellt.

Schlüsselwörter: OSM-Full-History, intrinsische Qualitätsindikatoren, Analyse-Plattform, verteilte Datenverarbeitung

Abstract: *The Ohsome-platform allows one to analyze the full-history of OpenStreetMap (OSM) worldwide on a temporal resolution of one second. Thus, it enables the quality assessment of OSM through the data itself based on its historical development. In order to serve different user groups, the platform provides interfaces at three different levels of abstraction. The functionalities and possibilities of the Ohsome-platform are illustrated by means of two examples: a web-dashboard to analyze the OSM-activities in Nepal and an analysis of the completeness of house numbers in Africa.*

Keywords: *OSM-Full-History, intrinsic quality indicators, analytics platform, distributed data processing*

1 Motivation und Stand der Technik

OpenStreetMap (OSM) stellt eine Datenbank mit einer stetig zunehmenden Menge an kollaborativ erhobenen Geodaten bereit, die jedoch hinsichtlich ihrer Qualität enorme Unterschiede aufweisen. Um die Nutzbarkeit von OSM-Daten für verschiedene Anwendungen beurteilen zu können, werden qualitätsbewertende Maße verwendet. Da für einige Gebiete wenige bis keine Referenzdatensätze für eine extrinsische Analyse zur Verfügung stehen, eignen sich im Fall von OSM insbesondere intrinsische Indikatoren zur Bewertung der Datenqualität (Anderson et al., 2016; Barron et al., 2014; Sehra et al., 2017). Diese werden aus den Daten selbst, das bedeutet z. B. aus deren historischer Entwicklung und der (attributiven) Vollständigkeit, abgeleitet. Vorhandene Softwaretools und Webservices zur intrinsischen Qualitätsbewertung, wie z. B. der iOSMANalyzer (Barron et al., 2013), OSM Analytics¹,

¹ <http://osm-analytics.org>

OSMatrix² oder EPIC-OSM (Anderson et al., 2016), waren bisher auf bestimmte Analysen und/oder auf ein begrenztes Gebiet limitiert.

In diesem Beitrag soll eine neue Analyse-Plattform vorgestellt werden, die es ermöglicht auf die vollständige Historie des OSM-Datensatzes zuzugreifen und dessen raumzeitliche Entwicklung weltweit und in sekundlicher Auflösung zu untersuchen. Dies liefert die Grundlage zur Berechnung von intrinsischen Qualitätsindikatoren von OSM-Daten auf globaler Ebene.

2 Die Ohsome-Plattform

Ziel der Ohsome-Plattform ist es, die OSM-Historie für verschiedene Benutzergruppen und Anwendungsgebiete auf globaler Ebene in sekundengenaue Auflösung leichter zugänglich zu machen. Der Zugriff auf historische OSM-Objekte kann auf verschiedenen Abstraktionsebenen erfolgen. Im Folgenden wird zunächst der Aufbau der Plattform vorgestellt, anschließend wird auf die abstrakteste Zugriffsebene, die Ohsome-API (engl. *Application Programming Interface*), eingegangen.

2.1 Aufbau der Ohsome-Plattform

Ein Überblick über die Ohsome-Plattform und ihre Schnittstellen ist in Abbildung 1 dargestellt. Der zentrale Baustein zur Speicherung der OSM-Historie ist die *OpenStreetMap History Database* (OSHDB). Beim Import mit dem OSHDB-Tool werden die historischen OSM-Objekte anhand ihrer Objekt-ID zusammengefasst und mithilfe eines räumlichen Index je nach Objektgröße auf insgesamt 15 unterschiedliche Zoomstufen verteilt. Diese Partitionierung bietet die Grundlage für einen parallelen Datenzugriff. Die Daten werden in einem kompakten Format gespeichert. Die eigentliche Datenhaltung kann grundsätzlich in einer beliebigen Schlüssel-Werte-Datenbank stattfinden, wie z. B. Apache Ignite³, das die Daten automatisch auf einer Cluster-Infrastruktur verteilt.

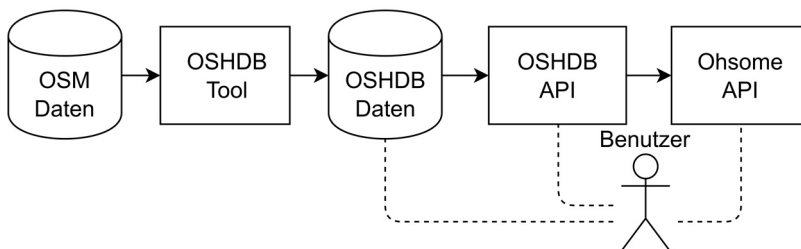


Abb. 1: Schematische Darstellung der Komponenten der Ohsome-Plattform

Um den Anforderungen verschiedener Endanwender gerecht zu werden, kann die Datenbankabfrage der OSM-Historie auf verschiedenen Abstraktionsebenen erfolgen. Der Benutzer

² <http://koenigstuhl.geog.uni-heidelberg.de/osmatrix>

³ <https://ignite.apache.org>

hat auf der tiefsten Abstraktionsebene über Java direkten Zugriff auf die OSHDB-Daten, welche die einzelnen historischen OSM-Elemente beinhalten. Auf dieser Ebene ist der Programmieraufwand für die Anwender am höchsten, sie bietet aber auch die größte Kontrolle. Auf der mittleren Abstraktionsebene steht eine weitere Java-Schnittstelle, die OSHDB-API, für den Zugriff auf diverse Aggregierungsfunktionen zur Verfügung. Sie basiert auf MapReduce, ein Programmiermodell für große Datenmengen (Dean & Ghemawat, 2004), das sich zur Parallelisierung der Datenverarbeitung eignet. Die höchste Abstraktionsebene ist eine Web-Schnittstelle namens Ohsome-API, die in Abschnitt 2.2 ausführlicher dargestellt ist.

Bei Verwendung eines bereitgestellten Webservices über die Ohsome-API ist kein Einrichtungsaufwand notwendig. Soll die OSHDB-API verwendet werden, lassen sich Regionen der Größe Nepals oder Österreichs mit wenig Aufwand lokal auf einem Laptop installieren. Das ist insbesondere aufgrund der kompakten Speicherung der Daten möglich. Bei größeren Regionen oder weltweitem Maßstab ist die Einrichtung eines Rechen-Clusters nötig.

2.2 Ohsome-API

Die Ohsome-API ist die Schnittstelle zwischen OSHDB und Webanwendungen (z. B. Nepal-Dashboard, siehe Kapitel 3.1). Sie kommuniziert über HTTP (GET oder POST) und ist nach dem Prinzip des *Representational State Transfer* (REST) gestaltet (Fielding, 2000). Die Ohsome-API wurde in Java mithilfe des Web-Frameworks Spring Boot⁴ realisiert.

Der Großteil der derzeit verfügbaren REST-Ressourcen stellt aggregierte Daten bereit, welche sich auf Momentaufnahmen, hier als *Snapshot-Views* bezeichnet, der OSM-Historie beziehen. Über diese Ressourcen kann man Elemente zählen, wie auch ihre Länge, Fläche oder ihren Umfang berechnen. Die Ergebnisse werden nach den Zeitpunkten der Snapshots aggregiert, können aber auch nach räumlichen Eigenschaften oder Attributen gruppiert werden. Dem gegenüber stehen *Contribution-Views*, mit denen die Veränderungen der Daten in einer Zeitspanne betrachtet werden. Mit dieser Art von Anfrage kann man z. B. untersuchen, wie viele unterschiedliche OSM-Mitwirkende in einem gewissen Gebiet und in einer gewissen Zeitspanne aktiv waren. Die hier zurückgegebenen Resultate werden zeitlich nach Intervallen aggregiert und können ebenfalls räumlich oder attributiv gruppiert werden.

Zur zeitlichen Eingrenzung kann ein Zeitpunkt, eine Zeitspanne, oder eine Liste von Zeitpunkten übergeben werden. Als räumliche Filter stehen Bounding-Boxen, Punkte mit jeweiligem Radius, oder Polygone zur Verfügung. Attributive Filter sind die OSM-Elementtypen (*node*, *way*, *relation*), oder OSM-Attribute (z. B. *building=residential*), wobei sowohl nach Attribut-Schlüsseln, als auch nach Schlüssel-Wert-Paaren gefiltert werden kann. Das Resultat wird im JSON-Format zurückgegeben und beinhaltet neben den abgefragten Werten auch Metadaten, wie die Version der Ohsome-API und Copyright-Informationen. Zusätzliche Metadaten, wie die Dauer der Prozessierung, können auf Anfrage zurückgegeben werden.

⁴ <https://projects.spring.io/spring-boot>

3 Anwendungsbeispiele

Die Analyse von historischen OSM-Daten zu Zwecken der intrinsischen Qualitätsbewertung ist ein aktives Forschungsgebiet der Geoinformatik und auch von hohem praktischen Wert für OSM-Nutzer. Durch die Ohsome-Plattform werden Werkzeuge zur Verfügung gestellt, die eine einfache, flexible und schnelle Abfrage von Qualitätsindikatoren ermöglichen. Um die Funktionalität der Ohsome-Plattform zu veranschaulichen, werden im folgenden zwei Anwendungsbeispiele vorgestellt.

3.1 Ohsome-Nepal-Dashboard

Das *Humanitarian OpenStreetMap Team* (HOT) koordiniert und aktiviert ein globales Netzwerk von OSM-Mappern, die zur Unterstützung von Katastrophenmanagement und humanitären Projekten aktuelle Geodaten in die OSM-Datenbank einpflegen. Beispielsweise kartierten mehrere Tausend Freiwillige nach den Erdbeben in Nepal im April und Mai 2015 Gebäude, Straßen und weitere relevante Infrastrukturen auf der Basis von Satellitenbildern. Die Non-Profit-Organisation *Kathmandu Living Labs* (KLL) unterstützte HOT bei der Koordination der Freiwilligen und der Kommunikation mit nationalen und internationalen Einsatzkräften vor Ort. Das Ohsome-Nepal-Dashboard (Abb. 2), das in Absprache mit KLL entwickelt wird, ermöglicht es, die Aktivitäten in OSM im Rahmen der Katastrophenhilfe 2015 zu analysieren. Die prototypische Webanwendung demonstriert die Verwendung der Ohsome-API. Sie stellt die Historie von OSM-Objekten mit den Schlüsseln *highway* und *building* seit Oktober 2007⁵, aggregiert für die vom Benutzer angegebene Bounding-Box, dar. Zudem kann nach dem OSM-Schlüssel *landuse* gefiltert werden, da im Fall des Nepal-Erdbebens zerstörte Häuser bzw. Wohngebiete als *landuse=brownfield* kartiert wurden.

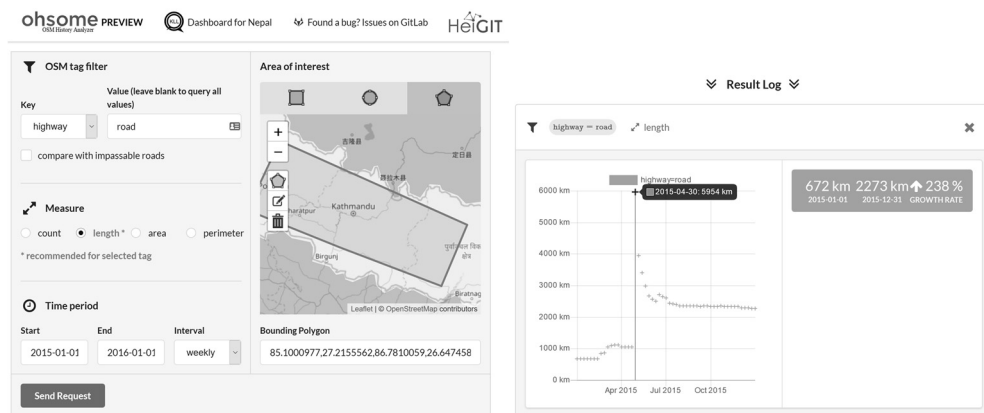


Abb. 2: Vorschau auf das Ohsome-Nepal-Dashboard, online unter: <https://ohsome.org/apps/kll-dashboard-preview>

⁵ Seit Beginn der öffentlichen Versionierung der OSM-Objekte: https://wiki.openstreetmap.org/w/index.php?title=API_v0.5&oldid=1551350.

Benutzer des Dashboards können den Analysezeitraum, das Zeitintervall (wöchentlich, monatlich, vierteljährlich oder jährlich) und ein beliebiges Gebiet innerhalb Nepals festlegen, sowie einen der beschriebenen OSM-Schlüssel auswählen, für den das Ergebnis dargestellt werden soll. Ein OSM-Attribut-Filter bietet zudem die Möglichkeit, die ausgewählten Objekte nach Attribut-Werten zu filtern. Beispielsweise können alle Straßen analysiert werden, die in der OSM-Datenbank das Attribut *highway=road* tragen, das für unbekannte oder nicht verifizierte Straßen verwendet wird⁶. Der Anteil dieser semantisch ungenauen Straßenkategorie ist nach dem Erdbeben zunächst angestiegen, nach einigen Wochen dann wieder gesunken (siehe Abb. 2, rechts). Zukünftig soll das Oshome-Nepal-Dashboard weitere Informationen zur OSM-Historie, z. B. über OSM-Benutzeraktivitäten, beinhalten.

3.2 Analyse-Beispiel mit der Oshome-API

Um die Funktionalität der Oshome-API zu veranschaulichen, wird nachfolgend ein Anwendungsbeispiel basierend auf Barron et al. (2014) dargestellt. Die Analyse untersucht die Vollständigkeit der Hausnummern aller Gebäudepolygone von Anfang 2010 bis Ende 2017 in einer wöchentlichen Auflösung. Dabei werden alle OSM-Elemente mit den OSM-Schlüsseln *building* (Elementtyp *way*), sowie *addr:housenumber* (Elementtypen *way* und *node*) berücksichtigt. Abbildung 3 stellt die zeitliche Entwicklung der Werte exemplarisch auf drei räumlichen Maßstabebenen dar: für den gesamten afrikanischen Kontinent, den Staat Tansania und die Stadt Daressalam. Die Anzahl der Gebäude wächst stetig, in Daressalam seit 2015. Dies lässt sich mit dem Start des gemeinschaftsbasierten Kartierungsprojektes Dar Ramani Huria⁷ in Verbindung bringen. Gegenläufig zu dem Zuwachs der Gebäude, nimmt der Anteil der kartierten Hausnummern ab.

Die Größe des der Analyse zugrunde liegenden OSM-Full-History Datensatzes beträgt 3,4 GB im *osh-pbf*-Format. Nach der Transformation in das OSHDB-Datenformat beträgt die Größe 21,5 GB. Die Analyse wurde auf einem Rechner mit 16 GB RAM und einem Intel Core i7-7500U-Prozessor mit vier logischen Kernen zu je 2,90 GHz durchgeführt und dauerte für Afrika ca. 11 Min., für Tansania 67 Sek. und für Daressalam 27 Sek.

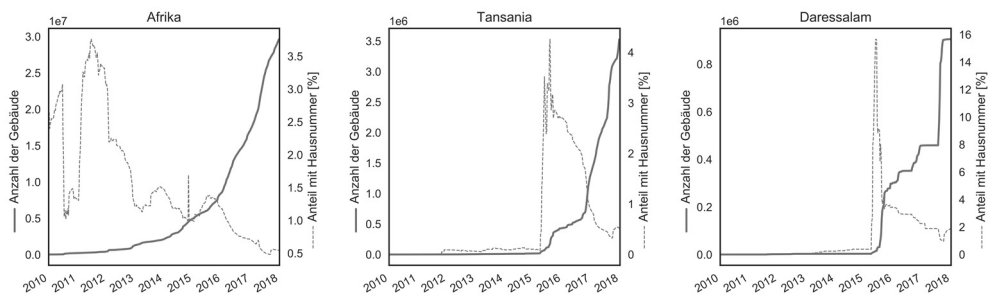


Abb. 3: Entwicklung von OSM-Gebäuden (mit einer Hausnummer)

⁶ OSM-Wiki. Tag: *highway=road*:
<https://wiki.openstreetmap.org/w/index.php?title=Tag:highway%3Droad&oldid=1439311>

⁷ <http://ramanihuria.org>

4 Ausblick

Die Ohsome-Plattform ermöglicht den Zugriff auf die gesamte OSM-Historie und bietet so die Grundlage zur intrinsischen Qualitätsbewertung von OSM-Daten. Einige aus der Literatur bekannte Qualitätsindikatoren, wie *conceptual compliance* (Ballatore & Zipf, 2015), sind bereits implementiert. Neben der Abfrage von vorgegebenen Parametern, können über die drei Schnittstellen beliebige, benutzerdefinierte Analysen implementiert werden. Die Ohsome-Plattform soll noch im Jahr 2018 als Open-Source-Software veröffentlicht werden.

Neben OSM-Qualitätsanalysen aus der Retrospektive, bietet die Ohsome-Plattform die Grundlage für zeitnahe Qualitätsbewertungen. Das ermöglicht beispielsweise, Mechanismen zur Bewertung von im Rahmen von Katastrophenhilfe erhobenen OSM-Daten einzuführen, sodass die Community noch während des Mapping-Prozesses darauf reagieren kann. Die Abfrage der OSHDB ist nicht auf einzelne Länder oder Kontinente beschränkt, sondern weltweit in sekundengenauer Auflösung möglich. Zukünftig soll dafür ein Webservice bereitgestellt werden.

Danksagung

Diese Arbeit wurde durch die *Klaus Tschira Stiftung* (KTS) Heidelberg unterstützt und im Rahmen des Projekts *Heidelberg Institute for Geoinformation Technology* (HeiGIT) durchgeführt.

Literatur

- Anderson, J., Soden, R., Anderson, K. M., Kogan, M., & Palen, L. (2016). EPIC-OSM: A Software Framework for OpenStreetMap Data Analytics. *49th Hawaii International Conference on System Sciences (HICSS)* (pp. 5468–5477).
- Ballatore A., & Zipf A. (2015). A Conceptual Quality Framework for Volunteered Geographic Information. *Spatial Information Theory – COSIT 2015* (pp. 89–107).
- Barron, C., Neis, P., & Zipf, A. (2013). iOSMANalyzer – ein umfassendes Werkzeug für intrinsische OSM Qualitätsuntersuchungen. In: J. Strobl et al. (Eds.), *Angewandte Geoinformatik 2013, Beiträge zum 25. AGIT-Symposium*. Berlin/Offenbach: Wichmann.
- Barron, C., Neis, P., & Zipf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, 18(6), 877–895.
- Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. *4th Symposium on Operating System Design and Implementation (OSDI)*.
- Fielding, R., T. (2000). *Architectural Styles and the Design of Network-based Software Architectures* (Dissertation). University of California, Irvine.
- Neis, P., & Zipf, A., (2008). OpenRouteService.org – Combining Open Standards and Open Geodata. *The State of the Map, 2nd OpenStreetMap Conference*.
- Sehra, S. S., Singh, J., & Rai, H. S. (2017). Assessing OpenStreetMap Data Using Intrinsic Quality Indicators: An Extension to the QGIS Processing Toolbox. *Future Internet*, 9(2), 15.