

WIE GUT IST OPEN STREET MAP? ZUR METHODIK EINES AUTOMATISIERTEN OBJEKTBASIERTEN VERGLEICHES DER STRASSENNETZE VON OSM UND NAVTEQ IN DEUTSCHLAND

Ina Ludwig, Angi Voss, Maïke Krause-Traudes

Zusammenfassung: Im OpenStreetMap-Projekt erstellen Tausende von Freiwilligen eine freie Karte der ganzen Welt. Ziel ist es, die Abhängigkeit von kommerziellen Geodatenanbietern zu verringern. Immer neue Anwendungsfelder werden erschlossen, zunächst die Kartenproduktion, dann die Navigation und neuerdings das Geomarketing. Argumente sind die Aktualität der Daten, die kostenfreie Nutzung und die Kundennachfrage. Nur ist das OSM-Sträßennetz wirklich schon gut genug, um kommerzielle Produkte wie NAVTEQ oder TeleAtlas zu ersetzen? Und kann man Marktdaten für Potenzialanalysen, also soziodemographische, ökonomische und psychographische Attribute, die bisher nur für die kommerziellen Sträßennetze angeboten werden, auf OSM übertragen?

Dies war der Hintergrund für die Diplomarbeit von Ludwig (2010), die sie am Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAIS durchgeführt und an der Universität Bonn eingereicht hat. Dieser Beitrag präzisiert zunächst die Fragestellung (Abschnitt 1) und präsentiert einige interessante Ergebnisse auf verschiedenen räumlichen Bezugsebenen. Beispielsweise können die Vollständigkeit von Straßen und die Vollständigkeit bzw. Korrektheit von Attributen beachtlich auseinanderklaffen (Abschnitt 2). Die Qualität von OSM wird im Vergleich zum Sträßennetz von NAVTEQ bestimmt. Dazu wird ein Zuordnungsverfahren zwischen den jeweiligen Straßenobjekten entwickelt (Abschnitt 3). Damit wurden verschiedene Qualitätsmaße, wie die Objekt- und Attributvollständigkeit, Positions- und Attributdifferenzen definiert. Diese können für beliebige Zielregionen und Straßenkategorien auf Knopfdruck ausgewiesen werden (Abschnitt 4). Das Verfahren ist soweit automatisiert, dass es mit wenig Aufwand periodisch wiederholt werden kann. Dies und weitere Anknüpfungspunkte werden in Abschnitt 5 aufgezeigt. Im Fazit schließen wir, dass man zwar je nach Anwendungszweck genau hinschauen muss, aber die OSM-Daten in großen Städten für das Geomarketing durchaus nutzbar sind.

Schlüsselwörter: OpenStreetMap, NAVTEQ, Geomarketing, Navigation, Qualität

// HOW GOOD IS OSM? – CONCERNING THE METHOD OF AN AUTOMATED AND OBJECT-WISE COMPARISON OF OSM AND NAVTEQ STREET NETWORK IN GERMANY

// Abstract: OpenStreetMap (OSM) is a collaborative project in which volunteers are creating a free and editable map of the world to reduce the dependency on commercial data providers. By and by, new areas of application emerge: firstly mapping, followed by routing and newly Business Geographics. Data actuality, exemption from charges and customer requests are arguments for OSM data usage in that application area. But are these data already suitable for Business Geographics and for replacing commercial products like TeleAtlas or NAVTEQ? And is it possible to consign market data which is spatially referenced to TeleAtlas or NAVTEQ subsequently to OSM street network? To answer this question, Ina Ludwig analyzed the quality of OSM road data for Germany in her diploma theses at Fraunhofer IAIS and the University of Bonn.

The given article presents her main findings. After specifying the problem (section 1) we present some interesting results on different spatial aggregation levels. We find p. e. that the completeness of street objects or the completeness of attributes may diverge significantly (section 2). OSM data quality is analyzed by comparison with NAVTEQ's proprietary road network data. This approach requires a matching to be constructed between the objects in both data sets which is described in section 3. Based on the matching we define several quality measures (object- and attribute completeness, positional and attribute differences) which can be reported for arbitrary target areas or street categories (section 4). Future prospects like the periodical repetition of analysis are presented in section 5. Finally, we draw the conclusion that suitability of OSM street data for Business Geographics tasks depends on the actual question but that its completeness in densely populated areas of Germany is quiet adequate.

Keywords: OpenStreetMap, NAVTEQ, Geomarketing, Navigation, Quality

Anschrift der Autoren

Ina Ludwig
Bergstraße 27
D-54340 Riold
E: InaLudwig@gmx.de

Dr. Angi Voss, Maïke Krause-Traudes
Fraunhofer IAIS, Schloss Birlinghoven
D-53754 Sankt Augustin
E: angi.voss; maïke.krause-traudes@iais.fraunhofer.de

1. FRAGESTELLUNG

OpenStreetMap (OSM) sammelt weltweit Daten über Straßen und „alles andere, was gemeinhin auf Karten zu sehen ist“ (www.openstreetmap.de). Dazu zählt neben Straßen auch eine Vielzahl von Spezialzielen (Points of Interest), die nicht nur für den Tourismus, sondern auch für eine Wohnungs- oder Standortsuche interessant sind. Selbsterklärtes Ziel ist es, die Abhängigkeit von den Anbietern proprietärer Daten zu beenden. Nach der OSM-Lizenz (Creative Commons Attribution-Share Alike 2.0, demnächst vielleicht Open Database License) darf jeder die Daten kostenfrei einsetzen und beliebig verarbeiten. Es müssen nur abgeleitete Produkte unter dieselbe Lizenz gestellt werden. Wurde primär an digitale und druckbare Karten gedacht, so kamen bald freie Dienste für die Navigation hinzu (Neis, Zipf 2006; Neis et al. 2010). Die Verortung von Adressen, die Berechnung von Routen und Erreichbarkeitszonen sind allerdings auch wichtige Funktionen für das Geomarketing. Typische Aufgaben des Geomarketing sind die Auswahl von Gebieten für Plakate oder Prospektwerbung, von Vertriebsgebieten und Standorten.

Seit 2009 halten OSM-Daten in Geomarketing-Produkte Einzug. Infas GEOdaten bietet etwa einen ergänzenden Datenbestand OSM+ mit PLZ-Gebieten, Gemeindegrenzen und Hausnummernbereichen an. IVU stellt OSM-Daten in seiner Software Filialinfo und auch als Webdienst zur Verfügung. LOGIBALL bietet mit Quality Grid ein Werkzeug zur Bewertung der Vollständigkeit des OSM-Strassennetzes an. Als Argumente werden genannt: Nachfrage der Kunden, Einsparung der Lizenzgebühren für kommerzielle Straßennetze und die Aktualität der Daten. Dem wäre noch hinzuzufügen, dass jeder unzureichende OSM-Daten jederzeit nachbessern kann.

Für das Geomarketing mit OSM fehlen allerdings noch Marktdaten, also soziodemographische, ökonomische und psychographische Daten zur Schätzung von Potenzialen im Raum. Auf der Ebene von Kreisen und Kommunen sind solche Angaben oft zu grob, während sie auf Hausebene zu teuer und aus Datenschutzgründen statistisch verrechnet sind. Aus solchen Gründen sind Marktdaten auf Straßenebene attraktiv und werden für die Straßennetze von NAVTEQ™ oder Tele Atlas™ kommerziell angeboten.

Um OSM-Daten für das Geomarketing zu erschließen, wäre also ein „Marktdatenübertragungsdienst“ nützlich, der Attribute von den Straßenobjekten des NAVTEQ- oder TeleAtlas-Netzes auf die Straßensobjekte in OSM überträgt.

Vorher wäre noch die Frage der Datenqualität zu klären. Im OSM-Portal steht, der Detaillierungsgrad der Daten sei regional sehr unterschiedlich und in vielen Städten schon besser als die meisten proprietären Karten. Und überhaupt, was hieße schon „vollständig“? Verschiedene Qualitätsuntersuchungen von OSM haben regionale Unterschiede festgestellt, insbesondere ein Stadt-Land-Gefälle (Hakley 2009 in England, Zielstra und Zipf 2010 in Deutschland). Für das Geomarketing sollte genauer ermittelt werden, wie gut die Qualität im konkreten Zielgebiet ist.

Als Träger von Marktdaten interessieren vor allem bewohnte Straßen, weniger Autobahnen und Wanderwege.

2. AUSGEWÄHLTE ERGEBNISSE

Als Methode wurde der Vergleich mit einem Referenzbestand von NAVTEQ gewählt. Dazu muss bekannt sein, wie sich die verschiedenen Konzepte von OSM und NAVTEQ entsprechen. Trotz der Unterschiede soll möglichst eine 1:1-Zuordnung konstruiert werden. Als Ergebnis kann hier eine vergleichende Bewertung der OSM-Daten auf Strassenebene ausgewiesen und z. B. auf Kommunen oder Kreise aggregiert werden (vgl. Abb. 1 und 2). Es stellte sich heraus, dass die Vollständigkeit regional schwankt. Sie ist besser in besiedelten Gebieten (79,1 % als in unbesiedelten (64,2 % im 10 Kilometer-Puffer um

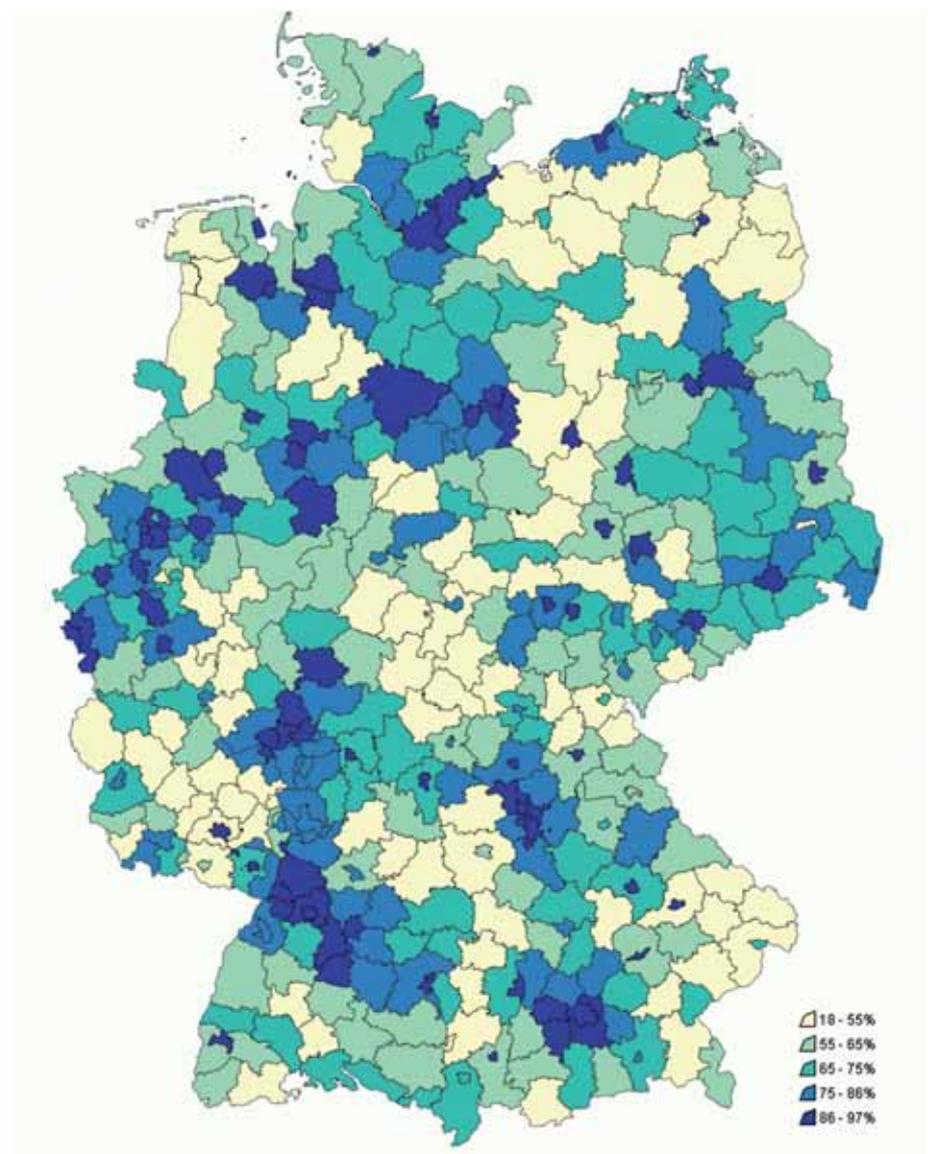


Abbildung 1: Relative Vollständigkeit von OSM für verschiedene Landkreise in Deutschland.

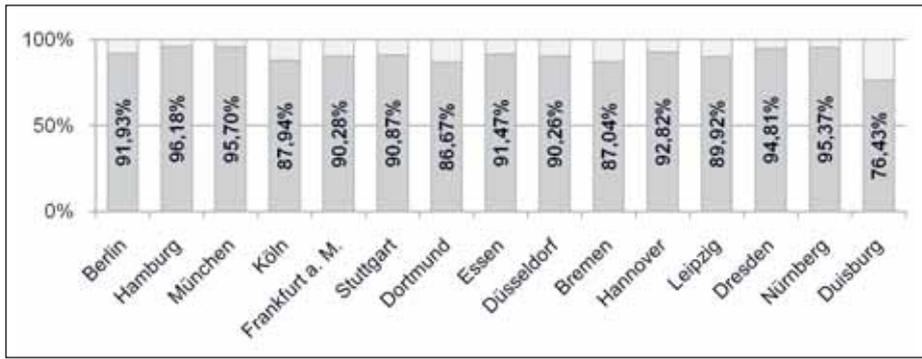


Abbildung 2: Relative Vollständigkeit von OSM in den 15 größten deutschen Städten.

Siedlungen, 50,8 % in Entfernung von mehr als 10 Kilometer) und bei stärker befahrenen Straßen (92,4 % für NAVTEQ Kategorie 4, 88,3 % für Kategorie 5 und 54,6 % für Kategorie 7).

Qualität erschöpft sich jedoch nicht in der Vollständigkeit. Weitere Kriterien laut ISO 19113 (2002) sind interne Konsistenz, Korrektheit und Aktualität, und zwar der Geometrie einerseits und der thematischen Attribute andererseits. Interessanterweise gibt es in OSM Städte mit hoher geometrischer Vollständigkeit, aber fehlenden

meisten die Höchstgeschwindigkeit (68 %) notiert, und sie unterscheidet sich auch nur gering von der in NAVTEQ (10,3 %). Dafür fehlt auffällig oft das Attribut Einbahnstraße (55 %).

- ▶ In Hamburg sind die untersuchten Attribute sehr vollständig, bis auf das Attribut Einbahnstraße (69 %).
- ▶ Die OSM-Straßen in Karlsruhe sind sehr vollständig (90 %) und größtenteils benannt (91 %), aber oft (15 %) fehlt die Höchstgeschwindigkeit oder weicht von der bei NAVTEQ ab (25 %).

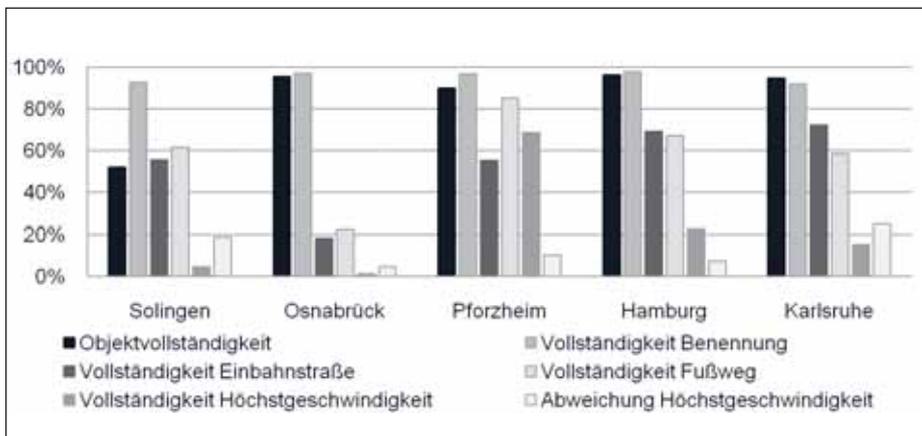


Abb. 3: Beispiele für Objektvollständigkeit, Vollständigkeit und Korrektheit von Attributen von OSM relativ zu NAVTEQ.

oder ungenauen thematischen Attributen, und umgekehrt.

- ▶ In Solingen z. B. sind die OSM-Straßen weniger vollständig (46 %), dafür aber größtenteils benannt (93 %).
- ▶ In Osnabrück hingegen sind viele Straßen (90 %) zu finden, aber auffällig selten wurden Einbahnstraßen (18 %), Fußwege (22 %) und die Höchstgeschwindigkeit (1,5 %) eingetragen.
- ▶ In Pforzheim wird von allen Städten am

3. DIE METHODIK: BEWERTUNG DURCH OBJEKTWEISEN VERGLEICH

Wie kommen wir zu diesen Ergebnissen? Die Untersuchung erstellt keine absoluten Aussagen, sondern zieht lediglich einen Vergleich zu den Straßendaten von NAVTEQ. Daher auch die relativen Prozentangaben in den Beispielen. In Osnabrück sind also für 90 % der NAVTEQ-Straßen entsprechende OSM-Straßen zu finden. Als kommerzielles Produkt, insbesondere

für die Fahrzeugnavigation, sollte das NAVTEQ-Straßennetz (oder alternativ das von TeleAtlas) eine hohe Qualität haben und sich als Referenzbestand eignen, wenn es um Standortanalysen und allgemeines Geomarketing geht. Für die Wahl einer Anwendung zwischen OSM und einem dieser proprietären Produkte sind Aussagen über die Unterschiede ohnehin interessant.

Es ist nicht neu, Referenzdaten für eine Qualitätsuntersuchung von Straßendaten zu verwenden. Joos (2000) beschreibt verschiedene Methoden.

Hakley (2008) verglich die Autobahnen des britischen Ordnance Survey Meridian 2 Datenbestands mit dem OSM-Straßennetz. Die Vollständigkeit prüfte er, indem er die Länge der Netze in identischen Gebieten verglich. Die Lagegenauigkeit bestimmte er, indem er die OSM-Daten mit Puffern um die Ordnance-Daten verschnitt.

Diese Methode wurde von Zielstra und Zipf (2009) für einen Vergleich mit TeleAtlas-Daten in Deutschland angepasst. Sie wurde in Neis et al. (2010) fortgeführt, um speziell die Eignung für Geocodierung, Routing, Einzugsgebiete und weitere freie Dienste zu untersuchen. Der Artikel vergleicht allerdings nicht die thematischen Attribute. Dazu wäre aber eine Zuordnung zwischen den Objekten der beiden Straßennetze erforderlich.

Eine solche Objektzuordnung konstruierte Eng (2009) für London teilautomatisch. Er verwendete die Integrated Transport Network Ebene des Ordnance Survey Master Map, und zwar für Autobahnen, A- und B-Straßen. Manuell musste er nicht passende Objekte aussortieren und die Namen einander angleichen. Anschließend benutzte er die Puffermethode, um die Lagen entsprechender Objekte zu vergleichen.

Ein solches Verfahren auf ein ganzes Land auszudehnen bedeutet erheblichen Aufwand. Die Herausforderung für uns bestand also darin, automatisch eine möglichst gute Zuordnung zu konstruieren. Ausserdem lag der Schwerpunkt auf bewohnten Straßen. Dafür wurde für den NAVTEQ-OSM-Vergleich eine Objektzuordnung nach Zhang, Meng (2007) konstruiert. Sie bezeichnen den Prozess des Zusammenbringens unterschiedlicher Datenquellen mit einer großen geographischen und/oder semantischen Ähnlichkeit als „Data Matching“. Derartige Verfahren werden häufig eingesetzt, um große

Datenbestände auch ohne räumliche Ähnlichkeit zu fusionieren.

Wir übersetzen im Folgenden den Begriff „Data Matching“ mit „Abbildung“. Eine Abbildung räumlicher Daten unterscheidet sich von einer mathematischen Funktion, die oft ebenso benannt wird. Dort wird in der Beziehung zwischen zwei Mengen jedem Element der einen Menge genau ein Element der anderen Menge zugeordnet (n:1-Beziehung). Zwischen Objekten räumlicher Datenbestände entstehen in der Verknüpfung hingegen auch n:m-Beziehungen (vergleichbar mit einer Relation in der Mathematik). Eine Abbildung kann sowohl zwischen Datenbeständen stattfinden, die auf dem gleichen Datenmodell aufbauen, als auch zwischen solchen, die unterschiedliche Modelle benutzen. Bei gleichen Datenmodellen, wie beispielsweise zwei Datenbeständen im GDF-Format, wird eine automatische Suche ähnlicher Objekte aufgrund identischer Datenstrukturen und Eigenschaften erleichtert. Datenbestände, die auf unterschiedlichen Modellen aufbauen, können dagegen nicht so einfach miteinander verknüpft werden, denn es ist möglich, dass sowohl die Eigenschaften als auch die geometrische Repräsentation von Objekten anders ausfallen.

Der verwendete Abbildungsprozess baut auf Devogele et al. (1998) sowie Walter, Fritsch (1999) auf und besteht aus den folgenden Schritten:

1 Vorbereitung der unterschiedlichen Datenschemata: Diese Phase umfasst alle vorbereitenden Aktivitäten für die Abbildung der unterschiedlichen Datenbestände.

- 2 Untersuchung von Konzept- und Objektkorrespondenzen:** Eine Verknüpfung zwischen zwei Straßendatenbeständen setzt korrespondierende Konzepte voraus. Mit der Untersuchung der Konzept- und Objektkorrespondenz wird festgestellt, wie die Objektklassen und ihre Attribute zueinander passen.
- 3 Abbildung und Nachbearbeitung:** Anschließend wird die Abbildung deutschlandweit mit Hilfe eines Algorithmus durchgeführt. Im Nachgang werden schlechte Abbildungen besonders betrachtet und gegebenenfalls nachbearbeitet.
- 4 Statistische Evaluation:** Mit diesem Abbildungsprozess kann die relative Objektivität des OSM-Datenbestandes ermittelt und die Qualität jeder einzelnen Straße hinsichtlich ihrer Geometrie und thematischen Attribute beurteilt werden. Ferner sind im Anschluss an die Abbildung auf NAVTEQ bezogene lizenzierbare Marktdaten auf die zugeordneten OSM-Straßen übertragbar.

3.1 VORBEREITUNG DER BEIDEN DATENSCHEMATA

Die Analyse basiert auf OSM-Daten vom April 2009, die ein Auszug aus der OSM-Datenbank von der Internetseite der Geofabrik (<http://download.geofabrik.de>) sind. Diese wurde zum Zeitpunkt des Downloads wöchentlich aktualisiert. Nach der Konvertierung der OSM-Daten von XML zu Oracle wurde der Datenbestand auf Geometrien ausschließlich aus Deutschland sowie auf zu vergleichende Attribute reduziert. Als NAVTEQ-Referenzbestand wurden die Straßendaten aus dem Straßendatenprodukt

„Digital Data Streets“ der Firma DDS vom Juli 2008 verwendet. Dieser Datenbestand musste zunächst verschiedentlich modifiziert werden: die Attribute „Straßenkategorie“, „Fußweg“, „Fußgängerzone“, „Länge“ und „Richtung“ wurden aus bereits existierenden NAVTEQ-Attributen abgeleitet oder neu berechnet. Die Straßenkategorie gibt die Signifikanz einer Straße auf einer ordinalen Skala von 1=Hauptstraße bis 7=„4th Class Road“ an. Die „Richtung“ basiert auf dem Originalattribut von NAVTEQ, enthält aber die zusätzliche Information, dass diese Straßen nicht ausschließlich für Spezialfahrzeuge freigegeben sind.

Da für die Untersuchung die Qualität von OSM speziell für das Geomarketing von Interesse war, erfolgte eine Konzentration auf bebaute Straßen. Deshalb wurden in beiden Datenbeständen von Beginn an Autobahnen und ihre Zubringer (NAVTEQ-Kategorie 1 und 2, OSM motorway, trunk, und deren Links) ausgeschlossen. Im NAVTEQ-Bestand fand gleichzeitig eine Vernachlässigung unbenannter Straßen statt. Die ausgewählten Attribute sind in beiden Beständen wesentlich, also vermutlich vollständig und korrekt, denn bei NAVTEQ wird der Straßennamen zur Verortung von Adressen und die Kategorie zum Finden von schnellen Routen benötigt. Auf der OSM-Karte werden nur Straßen mit Highway-Tag dargestellt. Insgesamt wurden 98,7 % der Straßenobjekte in OSM und 74,3 % von NAVTEQ verwendet. Die Abbildung 4 illustriert die verwendeten Untersuchungsmengen für Bonn: Straßen, die nicht in den Vergleich eingegangen sind, werden rot dargestellt, ausgewählte Straßen in schwarz.



Abbildung 4: Beispiele für Objektvollständigkeit, Vollständigkeit und Korrektheit von Attributen von OSM relativ zu NAVTEQ.

3.2 KONZEPTVERGLEICH

Eine Zuordnung zwischen zwei Straßennetzen setzt voraus, dass die Konzepte aufeinander abgebildet werden können. NAVTEQ 2008 richtet sich nach dem GDF-Format (ISO 14825 (2004)), die Konzepte von OSM hat Ramm (2009) beschrieben. Da OSM zunächst nicht für die Navigation konzipiert wurde, muss ein Straßenobjekt

nicht unbedingt an der nächsten Kreuzung oder Einmündung enden. Diese unterschiedliche Konzeption von Straßenobjekten ist der wesentliche Grund, weshalb es bei NAVTEQ etwa doppelt so viele Straßenobjekte gibt wie bei OSM. Sowohl bei OSM als auch bei NAVTEQ muss aber ein neues Objekt eingeführt werden, wenn sich ein thematisches Attribut, z. B. die Ge-

NAVTEQ		OSM	
Kategorie	Ordinal	Highway	Nominal, teilweise ordinal
Primary name, Secondary name	nominal (geschrieben mit „ss“)	Name, Ref	nominal (geschrieben mit „ß“)
Richtung	4 nominale Werte	One-way	Binär
Höchstgeschwindigkeit	9 ordinale Geschwindigkeitsklassen	Höchstgeschwindigkeit	Kontinuierliche Geschwindigkeitswerte
Fußweg, Fußgängerzone	Binär	Path, Footway, Steps, Track, Cycleway	Verschiedene qualitative Highway-Werte

Tabelle 1: Abbildungsattribute zwischen OSM und NAVTEQ.

NAVTEQ \ OSM	1	2	4	5	7
Motorway, Motorway link					
Trunk, Trunk link					
Primary, Primary link					
Secondary, Secondary link					
Tertiary					
Unclassified					
Residential					
Living street					
Service					
Footway					
Path					
Pedestrian					
Track					
Cycle way					
Steps					

Tabelle 2: Entsprechende Straßenkategorien in OSM und NAVTEQ.

schwindigkeit, ändert. Das führt zu weiteren Unterschieden, da es bei OSM, anders als bei NAVTEQ, ein Attribut für die Nutzungsart gibt. Für ein Straßenobjekt mit separatem Fuß- oder Radweg bei NAVTEQ entsteht so bei OSM ein „Linienbündel“. Bei Plätzen und ihren Zufahrten gibt es weitere konzeptionelle Unterschiede, die ebenfalls weit weniger Objekte betreffen als die Kreuzungssituationen. Plätze können in NAVTEQ als ein oder mehrere Straßenobjekte dargestellt sein, in OSM als Straßenobjekt oder Fläche, Parkplätze sogar als Punkt. Zufahrten zu Plätzen können benannt (NAVTEQ) oder auch unbenannt (OSM) sein, Kreise als ein (OSM) oder mehrere Objekte (NAVTEQ) erfasst sein. Die Tabelle 1 zeigt die Entsprechungen zwischen den thematischen Attributen bei NAVTEQ und OSM. Die Wertebereiche konnten gut aufeinander abgebildet werden, nur bei den Kategorien gab es keine direkte Entsprechung. Tabelle 2 zeigt, wie die Straßenkategorien einander zugeordnet wurden. Insbesondere der niedrigen NAVTEQ-Kategorie 7 entsprechen mehrere unterschiedlich Highway-Ausprägungen bei OSM.

3.3 ABBILDUNG

Um eine Abbildung zwischen OSM und NAVTEQ herzustellen, verfolgen wir einen vierstufigen Ansatz:

- 1 Segmentierung der OSM-Straßen
- 2 Reduktion der Kandidatenmenge durch die Anwendung von thematischen Filtern
- 3 Abbildung und Nachbearbeitung bzw. Elimination von schlechten Abbildungen
- 4 Evaluation der Abbildung: Die Abbildung wird in Anlehnung an Walter, Fritsch (1999) konstruiert, indem zunächst für jedes NAVTEQ-Straßenobjekt räumlich nahe Kandidaten in OSM ermittelt werden und diese dann anhand mehrerer Ähnlichkeitskriterien reduziert werden. Die besten Kandidaten, wenn sie gewisse Mindestkriterien erfüllen, sind die Zuordnungspartner. Im Nachgang können Spezialfälle behandelt werden. Ein solches Vorgehen, bei dem nur die besten Kandidaten behalten werden, funktioniert, wenn im Wesentlichen eine 1:1-Zuordnung konstruiert werden muss.

3.3.1 SEGMENTIERUNG DER OSM-STRASSEN

Da auf ein NAVTEQ-Objekt häufig mehrere OSM-Objekte entfallen, müssen diese weiter zerlegt werden. Dies erfolgt durch ein Zerschneiden der OSM-Objekte, indem Puffer um die NAVTEQ-Objekte gelegt werden. Wenn das am besten passende OSM-Segment zu einem NAVTEQ-Objekt bestimmt wurde (unter Umständen auch mehrere gleich gute), sind auch die zugehörigen OSM-Objekte bekannt, aus denen die Segmente entstanden sind. Nach einigen visuellen Stichproben wurde ein Puffer in drei Größen festgelegt: 5 Meter, 10 Meter und 30 Meter. Dabei sind 5 Meter die bei NAVTEQ erlaubte Abweichung von der Mittellinie, der Abstand 30 Meter nimmt Rücksicht auf Straßen, die an Plätze angrenzen. Durch die unterschiedliche Konzeption in den beiden Datenmodellen kann vor allem in solchen Fällen eine große Positionsdivergenz entstehen.

Immer, wenn ein OSM-Objekt einen Puffer an einer Stelle verlässt und an anderer Stelle wieder eintritt, entsteht eine Multigeometrie. Bei einem 5 Meter-Puffer kann die Länge einer Multigeometrie stark von der Länge des NAVTEQ-Objekts abweichen. Im größeren 10 Meter-Puffer sind die Längen dagegen meist vergleichbar.

Ein einziges OSM-Objekt kann also mehrere Segmente erzeugen, nicht nur wegen der drei verschiedenen Puffergrößen, sondern auch weil es die Puffer mehrerer NAVTEQ-Objekte in seiner Umgebung schneiden kann. Wichtig ist, dass der richtige Partner überhaupt als Kandidat erfasst wird und überzählige Kandidaten dann durch Ähnlichkeitsprädikate wieder heraus sortiert werden. Mit Hilfe der geometrischen Pufferverschneidung wird so also die ursprüngliche Menge der möglichen OSM-Abbildungspartner auf diejenige Kandidatenmenge reduziert, die maximal 30 Meter von einem NAVTEQ-Segment entfernt liegen.

3.3.2 ANWENDUNG VON THEMATISCHEN FILTERN

Da es ansonsten zu einer Verfälschung der Qualitätsaussagen kommen kann, muss die Zuordnung möglichst korrekt sein. Nicht zugeordnete Straßen könnten so die Messung der Vollständigkeit und falsche Zuordnungen die Lagegenauigkeit beeinflussen und den Vergleich thematischer Attribute erschweren. Deshalb wurden für die Auswahl und Reduktion der ersten Kandidatenmenge nur die Geometrie (Puffergröße und Länge), sowie der Name und die Kategorie der Objekte verwendet. Name und Kategorie konnten bereits bei der Auswahl der Straßen als korrekt angenommen werden. Auch die Lage sollte bei einer Toleranz von ± 30 Meter hinreichend genau sein.

Die Kandidaten aus der Kandidatenmenge 1 müssen eine „passende“ Straßenkategorie haben (siehe Tab. 2). Um in Kandidatenmenge 2 zu gelangen werden z. B. Rad- und Fußwege aus den Linienbündeln aussortiert.

Längen- und Namensvergleich

Im nächsten Schritt wird die Kandidatenmenge 2 weiter reduziert, um den angestrebten 1:1-Vergleich zu approximieren. Dazu wird anhand mehrerer Prädikate ein Ähnlichkeitsrang gebildet. Die Kandidaten müssen einen Mindestrang erreichen, und davon werden nur die mit dem besten Rang behalten. Dies können ein oder mehrere OSM-Segmente sein. Die Ähnlichkeitsprädikate betreffen die drei Puffergrößen, Längen und Namen.

Bei einem Längenvergleich werden die Längen von NAVTEQ- mit OSM-Objekten verglichen. Der Vergleich basiert auf der Annahme, dass die OSM-Segmente mindestens gleich lang sind wie die NAVTEQ-Segmente. Bei einem Längenvergleich wird darüber hinaus berücksichtigt, dass ein Puffer über

die beiden Enden seines NAVTEQ-Objekts hinausragt. Deshalb wird die Originallänge des Objekts im 5 Meter-Puffer um 10 Meter, und im 10 Meter-Puffer um 20 Meter erhöht. Dies hat allerdings nur Sinn, wenn das NAVTEQ-Objekt selbst nicht zu klein ist. Für den 5 Meter-Puffer muss es mindestens 5 Meter lang sein, für den 10 Meter-Puffer mindestens 10 Meter. NAVTEQ-Objekt und OSM-Segment werden als längenähnlich definiert, wenn der Unterschied nicht mehr als 25 % beträgt. Da ein OSM-Objekt in der Regel an mehrere NAVTEQ-Objekte angrenzt, gibt es entsprechend viele OSM-Segmente in verschiedenen Puffergrößen. Durch einen Längenvergleich können hier aber gute Auswahlsergebnisse erzielt werden, da unerwünschte Segmente in der Regel zu kurz sind.

Namen können nur verglichen werden, wenn beide Partner über einen verfügen. Während unbenannte NAVTEQ-Straßen schon bei der Vorauswahl ausgeschlossen wurden, kann der Name bei OSM-Straßen fehlen. Existiert kein offizieller Straßename, so können Ähnlichkeiten über die Straßenummer gefunden werden (z. B. „B9“). Der Namensvergleich kann also zwischen dem Primärnamen und dem sekundären Namen von NAVTEQ einerseits und dem Namen und der Referenz von OSM andererseits durchgeführt werden. Der Vergleich des Sekundärnamens von NAVTEQ mit dem Namen von OSM lässt zu, dass auch Straßen, die an Plätzen verlaufen, verknüpft werden. Hier würde der Primärname bei NAVTEQ der Name des OSM-Platzes sein und der Sekundärname bei NAVTEQ der Name der OSM-Straße, die an den Platz angrenzt. Hat eine Straße keinen Namen, sondern nur eine Straßenummer, so kann es vorkommen, dass diese direkt als Primärname vergeben wird. Aus diesen Gründen wurden vier Namensvergleiche durchgeführt.

- ▶ Name 1: Primärname (NAVTEQ) mit Name (OSM)
- ▶ Name 2: Sekundärname (NAVTEQ) mit Ref (OSM)
- ▶ Name 3: Sekundärname (NAVTEQ) mit Name (OSM)
- ▶ Name 4: Primärname (NAVTEQ) mit Ref (OSM)

Die Rubrik Name 2 erlaubt so die Abbildung von Straßen, wenn keine von beiden einen offiziellen Straßennamen hat. Name 3 umfasst Fälle, bei denen der Primärname von

NAVTEQ der Name eines Platzes ist und NAVTEQs Sekundärname der Name einer Straße an diesem Platz und damit der Name des Platzes bei OSM ist. Name 4 umfasst Fälle bei denen die NAVTEQ-Straße keinen Namen hat und ihre Straßenummer als Primärname zugewiesen wurde.

Namen werden mit der Levenshtein-Funktion verglichen, die den Buchstabenunterschied zweier Zeichenketten bestimmt. Groß- und Kleinschreibung, sowie Leerzeichen und Bindestriche werden nicht beachtet. Damit zwei Zuordnungspartner namensähnlich sind, muss die Differenz bei Straßennummern (Name 2 und Name 4) Null sein und darf bei Straßennamen (Name 1 und Name 3) vier Zeichen betragen. Vier Zeichen lassen zu, dass in einem Straßennamen maximal zwei „ß“ auftreten dürfen. Insgesamt bleiben nach dem Längen- und Namensvergleich nur noch solche Straßen in der Kandidatenmenge 3 übrig, die weniger als 30 Meter von einem NAVTEQ-Objekt entfernt liegen und zu diesem längen- oder namensähnlich sind.

3.3.3 RANKING

Durch die Kombination von Ähnlichkeitsprädikaten für Längen- und Namensähnlichkeit resultiert ein Ähnlichkeitsrang für alle potenziellen Abbildungskandidaten aus Kandidatenmenge 3. Die Tabelle 3 zeigt, wie verschiedene Ähnlichkeitsprädikate als Kriterien kombiniert wurden, um den Rang eines NAVTEQ-OSM-Paares zu bestimmen. Den höchsten Rang 1 bekommen z. B. nur OSM-Segmente, die in einem 5 Meter-Puffer längen- und namensähnlich zum NAVTEQ-Objekt sind. Ein Segment, das in einem 5 Meter-Puffer zwar namensähnlich, aber nicht längenähnlich ist und in einem 10 Meter-Puffer eine Ähnlichkeit in Name und Länge aufweist, ist ein Zuordnungskandidat mit Rang 2. Da der Ähnlichkeitsrang mehrere Bedingungen berücksichtigt und gemäß Tabelle 2 nicht nur identische, sondern auch „passende“ Straßenkategorien zugelassen wurden, ist die Zuordnung tolerant gegen Abweichungen in Position und Form der Straßen sowie in ihrer Kategorisierung und der Schreibweise des Namens. Außerdem bietet sich der Rang der Zuordnungspaare als Indikator für die Korrektheit der Zuordnung an.

Zusammengefasst definieren sich die Ränge folgendermaßen:

- ▶ Rang 1–3: Name und Länge sind im 5 Meter- oder 10 Meter-Puffer ähnlich

Rang	5LN	5N	10LN	10N	30N	5Lnull	10Lnull	5L	10L
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									

Table 3: Definition des Rangs als Kombination von Ähnlichkeitsprädikaten (5, 10, 30 steht für die Puffergröße, L für ähnliche Länge, N für ähnliche Namen, O für fehlende Namen).

- ▶ Rang 4–5: Längen sind ähnlich bis zu 10 Meter Entfernung und der OSM-Partner ist unbenannt
- ▶ Rang 6–8: es besteht nur Namensähnlichkeit
- ▶ Rang 9–10: Namen ähneln sich nicht, es besteht nur Längenähnlichkeit: Bei der Rangfolge wird dabei die geometrische Ähnlichkeit höher bewertet als eine Namensähnlichkeit, da hier auch unbenannte OSM-Straßen berücksichtigt wurden.

3.3.4 ABBILDUNG UND NACHBEARBEITUNG

Im weiteren Verlauf wählt der Abbildungsalgorithmus aus Kandidatenmenge 3 diejenigen OSM-Segmente als Abbildungspartner aus, welche den höchsten Rang haben. Wenn mehrere OSM-Partner den gleichen Rang haben, kommen alle als Abbildungspartner in Betracht. Daraufhin werden mit Hilfe eines GIS die Ränge der Zuordnungs-

partner visualisiert und stichprobenhaft auf ihre Richtigkeit überprüft (siehe Abb. 6). Auffällig viele falsche Zuordnungen wurden hier bei NAVTEQ-Objekten verzeichnet, die als Stichstraße in eine gleich benannte Straße einmünden und die in OSM fehlen. Ein Beispiel zeigt Abbildung 6 (NAVTEQ: blau, OSM: rot). Als Zuordnungspartner erhalten sie zwar ein namensähnliches, aber kein längenähnliches OSM-Segment. Im Rahmen einer Nachbearbeitung wurden diejenigen NAVTEQ-Objekte identifiziert, die sich in einem Endpunkt mit zwei oder mehreren gleichnamigen Straßenobjekten treffen und deren anderer Endpunkt mit keiner anders benannten Straße verbunden ist. Besitzen diese Objekte in der Zuordnung nur namensähnliche, aber keine längenähnliche OSM-Segmente (Rang 6 oder 7), so werden sie von der Zuordnungstabelle ausgeschlossen. In ganz Deutschland sind das rund 25.000 NAVTEQ-Objekte.

3.4 BEWERTUNG DER ZUORDNUNG

Eine Evaluation soll herausfinden, wie gut die Objekte zugeordnet werden. In vielen Ansätzen wird sie mit einer Zuordnungs- und Fehlerrate dargestellt, die durch den Vergleich mit einer manuellen Verknüpfung bestimmt werden. Während die Zuordnungsrate den Anteil der automatisch zugeordneten Objekte an dem Gesamtanteil der zu verknüpfenden Objekte aufzeigt, kann über die Fehlerrate der Anteil falsch zugeordneter Objekte bestimmt werden (vgl. Lüscher (2006)).

In der Diplomarbeit von Ludwig (2010) wird ein anderer Weg der Evaluation gewählt, der auf eine solche manuelle Zuordnung verzichtet. Die definierte Rangfolge ist hier ein Indiz zur Evaluation der Abbildung. Schlechte Zuordnungen können über einen schlechten Rang schnell aufgefunden und überprüft werden. In der beschriebenen Methode sind Objektpaare, die sowohl im Namen als auch in der Länge zum großen Teil übereinstimmen, die besten Zuordnungen. In Abbildung 7 sind die NAVTEQ-Objekte entsprechend dem Rang ihres zugewiesenen OSM-Partners eingefärbt: grüne NAVTEQ-Straßen weisen auf eine namens- und längenähnliche Abbildung in einer Entfernung von maximal 10 Metern an (Rang 1–3), gelbe Segmente auf unbenannte OSM-Partner (Rang 4–5), blaue Segmente auf OSM-Partner mit ähnlichen Namen, aber unterschiedlichen Längen (Rang 6–8), und rote Objekte haben einen Partner mit stark unterschiedlichem Namen (Rang 9–10).



Abbildung 6: Stichstraßen im NAVTEQ-Datensatz (blau).



Abbildung 7: NAVTEQ-Straßen in Bonn nach Rang der Zuordnung.

In der Qualitätskontrolle zeigt sich allerdings, dass viele Paare korrekt sind, auch wenn ihr Ähnlichkeitsrang schlecht ist. Das ist z. B. der Fall, wenn die OSM-Straße keinen Namen hat oder nicht nahe genug liegt. OSM-Straßen werden öfters per Rad oder zu Fuß erfasst, während sie bei NAVTEQ ja nur maximal 5 Meter von der Strassenmitte abweichen dürfen. NAVTEQ-Objekte mit unbenanntem OSM-Partner existieren eher bei niedrigen Straßenkategorien, oft handelt es sich dabei um in NAVTEQ benannte Wald- und Feldwege (4,67 % bei Kategorie 4, 13,71 % bei Kategorie 5, 13,87 % bei Kategorie 7). In die OSM-Qualitätsuntersuchung sind deswegen auch Verknüpfungen mit schlechten Rängen eingegangen. Insgesamt haben jedoch mehr als 60 % der Abbildungen einen Abbildungspartner mit ähnlichen Namen und ähnlicher Länge in einer maximalen Entfernung von 10 Metern gefunden (Rang 1–3), nur 3 % der Abbildungen fallen in die schlechtesten Klassen 9 und 10.

4. QUALITÄTSMASSE

Nach Joss (2000) kann die Qualität auf drei Bezugsgrößen eingehen: Individuelle Objekte, einzelne Attributwerte oder Aggregationen von Objekten auf räumlicher, thematischer oder zeitlicher Ebene. Entsprechend haben wir in Abschnitt 2 z. B. Unterschiede bei stärker und weniger stark befahrenen Straßen festgestellt, in besiedelten und unbesiedelten Gebieten, sowie die Qualität auf der Ebene von Kreisen und Kommunen verglichen. Für die Aggregation haben wir den amtlichen Gemeindegemeinschaftsschlüssel verwendet, eine achtstellige Ziffer, die auch die Kreis- und Länderzugehörigkeit kodiert. Für die (un)besiedelten Gebiete haben wir die Flächen von ca. 350 städtischen Siedlungen verwendet. Als Übergangszone wurde ein Puffer von 10 Kilometern um diese Siedlungspolygone deklariert, der Rest wurde als ländlicher Raum klassifiziert. Diese Bezugsgröße wird in Abbildung 7 weiter unten herangezogen.

Um die Qualität zu messen, werden für die ausgesuchten Qualitätskriterien „Vollständigkeit“, „Positionsdivergenz“ und „Attributdivergenz“ Maße benötigt. Die Maße sollten verschiedene Bezugsgrößen (eine Strassenkategorie, eine Stadt, ein Rang, ...) zulassen. Unsere Maße bilden für jedes Qualitätskriterium eine Grundmenge, die von den Bezugsgrößen abhängt. Davon

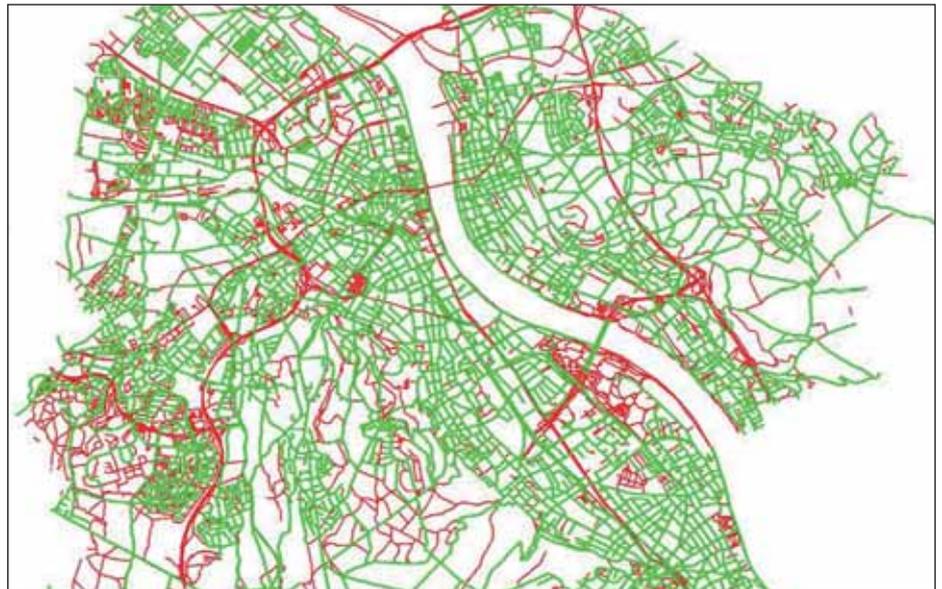


Abbildung 8: Zugeordnete (grün) und nicht zugeordnete OSM-Straßenobjekte in Bonn.

wird der Prozentsatz von Objekten bestimmt, die den Qualitätsbedingungen entsprechen. Für die Grundmenge kommen NAVTEQ-Objekte oder OSM-Objekte in Frage.

4.1 RELATIVE OBJEKTVOLLSTÄNDIGKEIT

Die Objektivvollständigkeit der OSM-Straßen drücken wir als Prozentsatz der NAVTEQ-Objekte aus, für die ein OSM-Partner gefunden wurde. „Relativ vollständig“ bedeutet also, „hat einen Partner“. Die Gesamtmenge bilden z. B. in Abbildung 1 alle NAVTEQ-Objekte eines Kreises, in Abbildung 2 alle in einer Kommune. Das Maß für die relative Objektivvollständigkeit ist definiert als Prozentanteil der NAVTEQ-Objekte mit mindestens einem zugeordneten OSM-Segment, gemessen an der Gesamtanzahl der NAVTEQ-Objekte. Umgekehrt können wir jedoch nicht systematisch erfassen, wo eine Übervollständigkeit der OSM-Straßen vorliegt. Denn im OSM-Bestand können sich Straßen befinden, deren potenzielle Partner vom Vergleich ausgeschlossen wurden, weil sie zur NAVTEQ-Kategorie 1 oder 2 gehören.

Abbildung 8 kann jedoch einen Eindruck von der „Übervollständigkeit“ von OSM geben. Gezeigt wird das Ergebnis der Verknüpfung für die Stadt Bonn. Nicht verknüpfte OSM-Straßen sind hier rot markiert. Es ist zu erkennen, dass viele kleine Wege in OSM nicht zugeordnet wurden, z. B. in den Rheinauen, westlich des Rheinkniees, oder im Kottenforst, im Südwesten.

4.2 RELATIVE ATTRIBUTVOLLSTÄNDIGKEIT

Für die Attribute Name, Einbahnstraße, Fußwege, Fußgängerzonen und Höchstgeschwindigkeit bestimmen wir die relative Attributvollständigkeit. Auch diese drücken wir als Prozentsatz von NAVTEQ-Objekten aus. In die Gesamtmenge kommen nun aber alle NAVTEQ-Objekte, die einen Zuordnungspartner haben und das Attribut besitzen. Sie entsprechen den Qualitätsanforderungen, wenn ihr OSM-Partner das entsprechende Attribut ebenfalls besitzt. Strenger hätten wir fordern können, dass all ihre OSM-Partner das entsprechende Attribut besitzen. Das Maß für die Quantifizierung der relativen Attributvollständigkeit ist definiert als Prozentanteil der NAVTEQ-Objekte mit mindestens einem zugeordneten OSM-Segment, das ein entsprechendes Attribut hat, gemessen an der Gesamtanzahl der NAVTEQ-Objekte, die erstens einen Zuordnungspartner und zweitens dieses Attribut haben. Dem Maß zugrunde liegt der Umstand, dass sowohl NAVTEQ als auch OSM für eine Straße ein neues Objekt einführen, wenn sich eines dieser Attribute ändert. NAVTEQ führt zwar bei Kreuzungen weitere Objekte ein, aber dafür werden OSM-Objekte entsprechend segmentiert.

Im OSM-Portal gibt es bereits einige Instrumente, die überprüfen, ob eine Straße einen Namen trägt oder nicht (z. B. Maplint). Solche Programme können die interne Konsistenz prüfen, aber sie können nicht berücksichtigen, ob eine Straße in Wirklichkeit keinen Namen hat. Die relative Attribut-

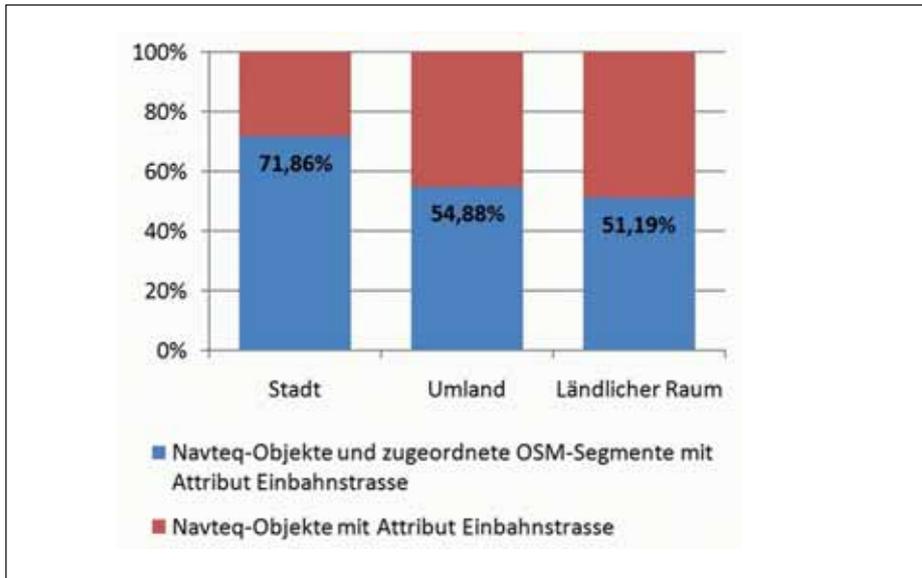


Abbildung 9: Relative Vollständigkeit des Attributs „Einbahnstraße“ in städtischen Gebieten, Umland und ländlichem Raum.

vollständigkeit des Namens, die in dieser Arbeit über eine Objektzuordnung untersucht wird, kann mit großer Wahrscheinlichkeit unbenannte Straßen in OSM erkennen, die in dem Referenzdatenbestand von NAVTEQ einen Namen haben. Eine Voraussetzung für das Erkennen unbenannter Strassen ist eine Längenähnlichkeit in einem Puffer von maximal 5 Metern oder 10 Metern um ein NAVTEQ-Objekt. Unbenannte Straßen, die in einem Puffer liegen, aber nicht längenähnlich sind, werden nicht erkannt, da diese von der Zuordnung ausgeschlossen wurden. Über die Zuordnung können Straßen, die in OSM keinen Namen besitzen, leicht erkannt werden. In den meisten Fällen handelt es sich um kleinere Wege oder Pfade, wie in Bonn beispielsweise Wege am Rheinufer oder Wanderwege. In einer manuellen Überprüfung einiger unbenannter OSM-Straßen (Daten aus April 2009) mit einer aktuelleren Karte in dem Portal (Oktober 2009) konnte festgestellt werden, dass viele Straßen in dem Zeitraum der Qualitätsuntersuchungen bereits mit einem Namen ergänzt wurden. In innerstädtischen Bereichen fallen aber vor allem wieder Straßen auf, die an einen Platz angrenzen. Diese sind in den NAVTEQ-Daten mit dem Namen des Platzes gekennzeichnet und haben in OSM meist keinen Namen.

Weiterhin sind oft kleinere Straßen, die zu einem Parkplatz führen, in OSM unbenannt, während sie in NAVTEQ den Namen der größeren Straße tragen, von denen sie abgehen. In der relativen Attributvollständigkeit

müssen solche Fälle beachtet werden. Sind in Städten viele der beschriebenen Fälle vorzufinden, so wird die Vollständigkeit, die über den Anteil der benannten OSM-Zuordnungspartner von NAVTEQ-Objekten gemessen wird, beeinflusst.

Bei den Einbahnstraßen nimmt die Rate der relativen Vollständigkeit ähnlich wie bei den Namen auch mit zunehmender Entfernung von städtischen Siedlungen ab. Zusätzlich zu den auf fünf ausgewählte Städte bezogenen Informationen aus Abbildung 2 zeigt 9 die Vollständigkeit des Attributs „Einbahnstraße“.

4.3 UNTERSCHIEDE IN DER HÖCHSTGESCHWINDIGKEIT

Attributdifferenzen können nur für quantitative Attribute bestimmt werden, also hier nur für die Höchstgeschwindigkeit. Da NAVTEQ nur elf Geschwindigkeitswerte vergibt und OSM beliebige Werte zulässt, wurde der Wertebereich zunächst in fünf Klassen diskretisiert:

0 – < 30 km/h, < 50 km/h, < 70 km/h, < 100 km/h, ≥ 100 km/h.

Wir berechnen nicht die durchschnittliche Geschwindigkeitsdifferenz, sondern den Anteil der Objekte, bei denen die Geschwindigkeitsklassen nicht übereinstimmen. Genauer bilden NAVTEQ-Objekte, die einen Zuordnungspartner und eine Geschwindigkeitsangabe haben, die Grundmenge. Davon bestimmen wir den Prozentsatz der Objekte, die einen OSM-Partner haben, der entweder keine Höchstgeschwindigkeitsangabe hat oder eine mit abweichender Wertklasse. Das Maß für die Unterschiede in der Höchstgeschwindigkeit wird definiert als Prozentanteil der NAVTEQ-Objekte mit mindestens einem zugeordneten OSM-Segment, das keine Höchstgeschwindigkeitsangabe oder eine Höchstgeschwindigkeitsangabe in einer anderen Klasse hat, gemessen an der Gesamtanzahl der NAVTEQ-Objekte die erstens einen Zuordnungspartner und zweitens eine Höchstgeschwindigkeitsangabe haben.

Die Attributdifferenz kann in einer Konfusionsmatrix weiter aufgeschlüsselt werden. Die Tabelle 12 zeigt für ganz Deutschland,

NAVTEQ \ OSM	1	2	3	4	5
1	2,74	0,28	0,09	0,01	0,00
2	6,37	18,01	3,36	0,47	0,08
3	0,28	0,82	6,72	1,90	0,97
4	0,01				
5	0,01				
No speed limit	0,26				
	2,35				
	1,23				
	0,00	0,01	0,24	1,57	2,7
	90,60	80,86	89,34	93,7	95,02

Tabelle 4: Konfusionsmatrix für Klassen der Höchstgeschwindigkeiten in Deutschland.

in welchen Klassen NAVTEQ und OSM wie oft voneinander abweichen. Am häufigsten haben Partner von NAVTEQ-Objekten der Geschwindigkeitsklasse 2 überhaupt eine Geschwindigkeitsangabe (Spalte der Klasse 2). Interessant ist die Geschwindigkeitsklasse 1 von NAVTEQ (Spalte der Klasse 1). Hier haben 6,37 % einen OSM-Partner mit der abweichenden Geschwindigkeitsklasse 2, hingegen nur 2,74 % in der gleichen Geschwindigkeitsklasse 1.

4.4 POSITIONSDIFFERENZ

Für die Positionsdifferenz nutzen wir die Puffergröße der Zuordnungspartner. Da wir die genauen Abstände nicht kennen, weisen wir nur die Anteile der Objekte aus, deren Partner alle im 5 Meter, 10 Meter bzw. 30 Meter Puffer liegen. Die Grundmenge bilden diesmal die OSM-Objekte. Das Qualitäts-

Puffer fallen, aber in keinen kleineren Puffer, gemessen an der Gesamtanzahl der OSM-Objekte. Berechnet für ganz Deutschland liegen 73 % der OSM-Objekte in einer maximalen Entfernung von 5 Metern von einem NAVTEQ-Objekt, 21 % liegen zwischen 5 und 10 Meter entfernt und nur 6 % liegen zwischen 10 und 30 Meter entfernt. Erstaunlicherweise weichen vor allem kleinere Wege in OSM von der Position der NAVTEQ-Straßenobjekte ab (vgl. Abb. 10). Denkbar wären hier Ungenauigkeiten der Positionsaufnahme, die über die Digitalisierung der GPS-Tracks im Editor entstehen. Bei dem dichten Straßennetz einer Stadt existieren weitaus mehr Referenzpunkte, auf die sich in einer genauen Nachzeichnung der Tracks gestützt werden kann, als bei einem wenig erfassten Waldgebiet. Der geringe Prozentwert bei der Strassenkategorie

wird laufend verbessert. So wurden noch während unserer Untersuchung einige Einbahnstraßen in Heidelberg im Sinne von NAVTEQ korrigiert.

Es ist also wichtig, unsere Methode leicht wiederholen zu können. Ein Datenvergleich zu verschiedenen Zeitpunkten würde auch Aussagen zur Aktualität ermöglichen. Als Voraussetzung müssen die beiden Datenbestände in eine Oracle Datenbank importiert werden. Der NAVTEQ-Bestand wurde automatisch vorverarbeitet, um insbesondere mehrere Attribute für Namen und Fußwege zusammenzuführen. Auch die Zuordnung und die Berechnung der Qualität pro Straßenobjekt geschehen vollautomatisch. Das Ergebnis ist eine Tabelle der NAVTEQ-Objekte mit zugeordneten OSM-Segmenten und -objekten, sowie den zugehörigen Ähnlichkeitskriterien und dem Vergleich der thematischen Attribute.

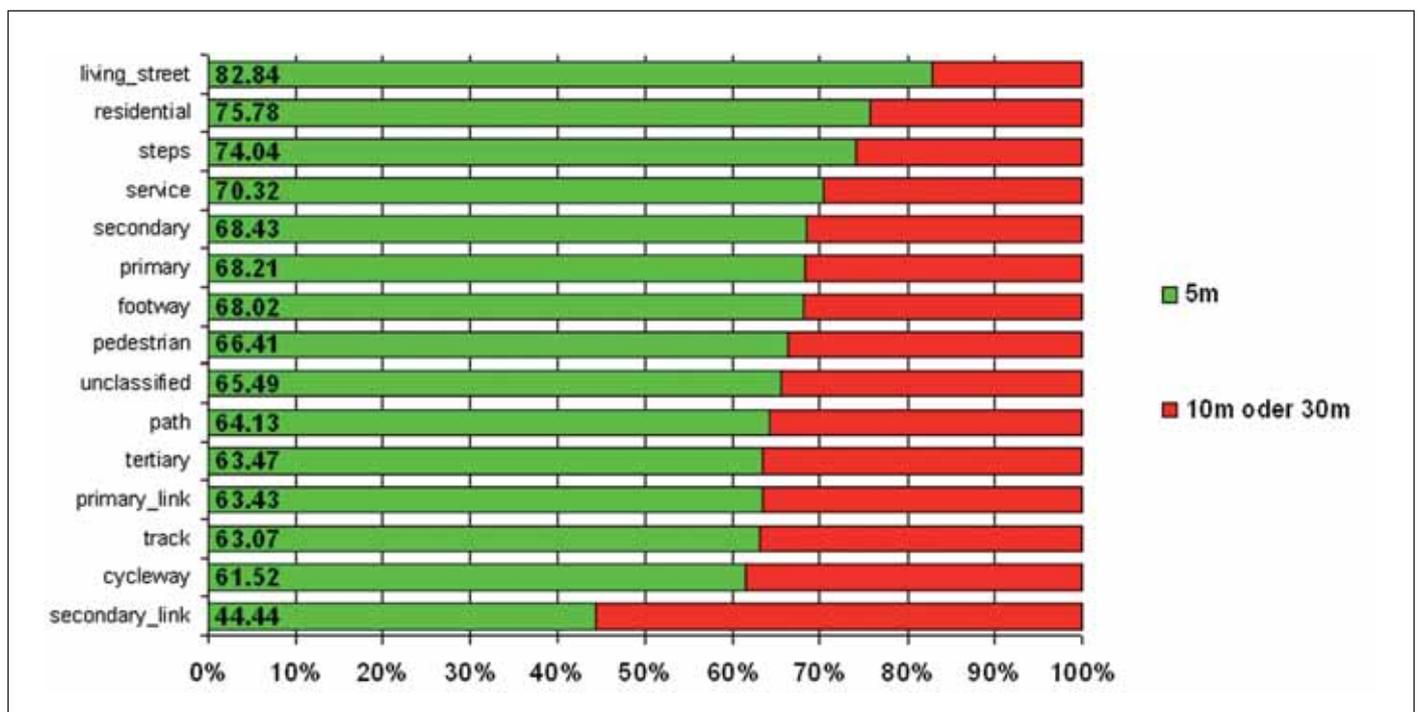


Abbildung 10: Anteil der Zuordnungen im 5 Meter-Puffer nach Straßenkategorien von OSM.

prädikat bedeutet, dass das OSM-Objekt mindestens im 5 Meter, 10 Meter bzw. 30 Meter Puffer all seiner Zuordnungspartner liegt. Die Puffergröße der Zuordnungspartner ist im Rang der Zuordnung (nach Tabelle 3) verschlüsselt: 5 Meter-Puffer = Rang 1, 2, 4, 6, 9, 10 Meter-Puffer = Rang 3, 5, 7, 30 Meter-Puffer = Rang 8. Das Maß für die Positionsdifferenz ist hier der Prozentanteil der OSM-Objekte, deren Zuordnungspartner alle in den 5 Meter, 10 Meter bzw. 30 Meter

„Secondary link“ ist auf die geringe Anzahl zugeordneter Objekte zurückzuführen (in ganz Deutschland nur 9 Objekte).

5. ZUSAMMENFASSUNG UND AUSBLICK

5.1 WIEDERHOLBARKEIT

Die hier präsentierten Ergebnisse betreffen die NAVTEQ-Daten vom dritten Quartal 2008 und die OSM-Daten von April 2009. Der OSM-Datenbestand wächst weiter und

5.2 AUSBAU

Die Arbeit von Ludwig bietet viele Fortsetzungsmöglichkeiten. Die Untersuchung kann periodisch wiederholt werden, um die Qualitätsänderung in OSM zu dokumentieren. Werden die relativen Qualitätswerte steigen und ließe sich daraus schließen, dass der NAVTEQ-Datenbestand als Referenzbestand noch geeignet ist?

Unsere Methode kann verbessert werden, indem die Zuordnung durch Nach-

behandlung in weiteren Spezialfällen korrigiert wird. Die Zuordnung kann aber auch jetzt schon genutzt werden, um Marktdaten von NAVTEQ-Objekten auf OSM-Objekte zu übertragen. Zuordnung und Qualitätsbeurteilung können auf die gesamten Datenbestände ausgebaut werden, beginnend mit Deutschland. Die Methode sollte ferner auf die Topologie ausgedehnt werden. Ihre Qualität ist wichtig, wenn das Straßennetz von OSM für Navigationsdienste genutzt werden soll. Interessant für die Standortanalyse wäre schließlich eine Zuordnung der Spezialziele, die ja eine besondere Stärke von OSM sind.

5.3 FAZIT

Unsere Untersuchung hat frühere Arbeiten bestätigt: Das OSM-Straßennetz ist im Allgemeinen recht lagegenau (72,64 % im 5 Meter-Puffer und 20,89 % im 10 Meter-Puffer), aber es ist regional unterschiedlich vollständig und weist ein Stadt-Land-Gefälle auf. Differenzierter als frühere Analysen konnte festgestellt werden, dass hohe Vollständigkeit

der Straßen längst nicht bedeutet, dass auch sonstige Attribute vollständig und korrekt sind und umgekehrt. Es muss also immer genau hingeschaut werden.

Speziell für Geomarketing-Anwendungen werden Antworten erwartet auf die Fragen, wie vollständig das OSM-Netz im anvisierten Zielgebiet ist und wie vollständig und korrekt die „relevanten“ thematischen Attribute, und zwar für die „relevanten“ Straßenkategorien.

Der Straßenname ist ein relevantes Attribut, wenn Adressen verortet werden müssen. Die anderen untersuchten Attribute sind relevant, wenn Routen und Einzugsgebiete berechnet werden sollen, wobei auch die topologische Konsistenz bewertet werden müsste. Relevante Straßenkategorien sind bewohnte Straßen, wenn Potenziale bestimmt werden sollen. Autobahnen sind relevant, wenn Einzugsgebiete bestimmt werden sollen.

Unsere objektweise Vergleichsmethode ermöglicht erstmalig solche differenzierte

Qualitätsaussagen für beliebige Zielgebiete und Straßenkategorien auf Knopfdruck. Nur über die topologische Konsistenz und Autobahnen können keine Aussagen gemacht werden, aber die anderen Untersuchungen lassen auf eine hohe Vollständigkeit bei Autobahnen schließen.

Unsere Untersuchung legt nahe, dass die Lage und Benennung der OSM-Straßen gut genug für die Konstruktion unserer Zuordnung zu den NAVTEQ-Straßen ist. Sie legt auch nahe, dass in den größeren Kommunen die Vollständigkeit (87 % im Durchschnitt, 89 % im Median) für Geomarketing-Zwecke in vielen Fällen ausreicht. Da die Menge und Qualität der OSM-Daten weiterhin beständig zunimmt (Zipf 2010) ist es gerechtfertigt, OSM-Straßendaten in Geomarketing-Produkte einzubauen. Wir erwarten, dass neue Dienste neuen Bedarf in der OSM-Community stimulieren, die dann die OSM-Datenbestände für die neuen Zwecke geeignet verbessern wird. ◀

Literatur

Devoegele, T.; Parent, C.; Spaccapietra, S. (1998): On spatial database integration. INT J GEOGR INF SCI, 12, Nr. 3, S. 1-20.

Eng, M. (2009): A quality analysis of OpenStreetMap data. Dissertation. University College London.

Hakley, M. (2009): How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England. (online: http://www.ucl.ac.uk/~ucfamha/OSM%20data%20analysis%20070808_web.pdf, Zugriff 12/2009).

ISO 14825 (2004): Intelligent transport systems-Geographic Data Files (GDF) – Overall data specification.

ISO 19113 (2002): Geographic information – quality principles.

Joos, G. (2000): Zur Qualität von objektstrukturierten Geodaten. München, Universität der Bundeswehr München, Dissertation.

Ludwig, I. (2010): Abbildung von Straßendaten für Qualitätsuntersuchungen – Ein Vergleich von OpenStreetMap mit Navteq, Universität Bonn, Diplomarbeit.

Lüscher, P. (2006): Matching von Straßendaten stark unterschiedlicher Masstäbe und Aufbau einer Multirepräsentationsdatenbank. Universität Zürich, Dissertation. (online: http://www.geo.unizh.ch/~luescher/publications/luescher_diplomarbeit.pdf, Zugriff 09/2010).

NAVTEQ (2008): NAVTEQ's NAVSTREETS Street Data: Reference Manual v3.0. (online: http://faculty.unlv.edu/jensen/gisdata/navteq/Technical-Reference/NAVSTREETS_Reference_Manual_v3.pdf, Zugriff 12/2009).

Ramm, F.; Topf, J. (2009): OpenStreetMap: Die freie Weltkarte nutzen und mitgestalten. Lehmanns Media.

Neis, P.; Zielstra, D.; Zipf, A.; Strunck, A. (2010): Empirische Untersuchungen zur Datenqualität von OpenStreetMap – Erfahrungen aus zwei Jahren

Betrieb mehrerer OSM-Online-Dienste, AGIT 2010 Symposium für Angewandte Geoinformatik. Salzburg.

Neis, P.; Zipf, A. (2006): Realisierung der Tourenplanung für ein Emergency-Route-Service auf Basis der OpenLS Spezifikation. AGIT 2006 Symposium für Angewandte Geoinformatik. Salzburg.

Walter, V.; Fritsch, D. (1999): Matching spatial data sets: a statistical approach, INT J GEOGR INF SCI, 13, Nr. 5, S. 445-473.

Zielstra, A, D.; Zipf, A. (2010): A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. AGILE 2010. The 13th AGILE International Conference on Geographic Information Science. Guimarães,

Zhang, M.; Meng, L. (2007): An iterative road-matching approach for the integration of postal data. Computer, Environment and Urban Systems 31, S. 597-615.