

Lokalisierung von Online-Karten aus Openstreetmap-Daten

Ausgangssituation

Online-Karten auf Basis von Openstreetmap (OSM) unterscheiden sich von anderen Karten auf Basis herkömmlicher Geodaten wie ATKIS in zwei wichtigen Aspekten. Zum einen ändern sich die zugrunde liegenden Vektordaten häufig, was bei deren Einsatz z.B. durch eine kontinuierlich aktualisierte Datenbank berücksichtigt werden muss. Zum anderen liegen diese Daten üblicherweise in mehreren Sprachen bezüglich des Namens von Objekten (Länder, Städte, Straßen usw.) vor.

Der vorliegende Beitrag zeigt Lösungsansätze zur Nutzung mehrsprachiger Daten für die Erzeugung von Online-Karten. Diese Beschränkung auf Online-Karten ist jedoch nicht technischer Natur. Die beschriebenen Lösungsansätze sind natürlich auch für die Erzeugung von Karten anwendbar, die für den Druck bestimmt sind.

Problemstellung

Im Gegensatz zu herkömmlichen Daten stellt Openstreetmap einen weltweiten Datensatz in einem einheitlichen Koordinatensystem (WGS84) zur Verfügung.

Dieser erfreut sich deshalb und insbesondere auch wegen seiner freien Verfügbarkeit zunehmender Beliebtheit, und zwar nicht nur bei Interessengruppen wie z.B. Radfahrern, Wanderern, Skifahrern usw., sondern auch bei international tätigen **Nichtregierungsorganisationen**, Behörden und sogar Militärs.

Openstreetmap verwendet für die Bezeichnung geografischer Objekte ein flexibles Schlüssel-Wert-System, das man aus modernen Programmiersprachen kennt (associative array).

Beispiel: Teil der Datenobjekte für Deutschland und Israel:

name	⇒	Deutschland	name	⇒	ישראל
..			..		
name:de	⇒	Deutschland	name:de	⇒	Israel
name:en	⇒	Germany	name:en	⇒	Israel
name:ar	⇒	ألمانيا	name:ar	⇒	إسرائيل
name:ja	⇒	ドイツ	name:ja	⇒	イスラエル
name:ru	⇒	Германия	name:ru	⇒	Израиль
name:he	⇒	גרמניה	name:he	⇒	ישראל

Wie man sieht, wird der gewöhnliche Name eines Objektes üblicherweise in der

Landessprache verfasst (Deutschland, ישראל usw.). Im Gegensatz dazu wünscht sich ein Kartenanwender jedoch eine Karte in seiner Sprache (oder einer Lingua franca wie Englisch) , bei der die Landessprache allenfalls als Bezeichnung in Klammern erscheint, wie z.B.: Moskau (Москва). Nicht-lateinische Schriftzeichen sollten dabei immer alternativ dargestellt werden.

Lösungsansatz und Stand der Technik

Geodaten aus Openstreetmap werden zur Kartenerzeugung üblicherweise in eine PostgreSQL/PostGIS¹ Datenbank importiert. Frei verfügbare Importwerkzeuge wie osm2pgsql² oder imposm³ halten Mechanismen bereit, um diese Datenbank kontinuierlich auf dem aktuellen Stand zu halten.

Die eigentlichen Karten werden dann aus dieser Quelle erzeugt. Mögliche freie Werkzeuge hierfür sind Mapnik, Mapserver oder Geoserver. Des Weiteren stehen auch kommerzielle Werkzeuge wie z.B. ESRI/ArcGIS zur Verfügung. Eine neue Entwicklung stellt die Darstellung von Kartenkacheln als Vektordaten dar. Hierbei werden in einem Zwischenschritt Daten aus der PostGIS-Datenbank kachelweise extrahiert und ggfs. geometrisch vereinfacht. Diese Kacheln werden dann bei Abruf durch den Server in Rasterkacheln umgewandelt oder sogar direkt im Client, z.B. in modernen Webbrowsern mit WebGL, gerendert.

Die Lokalisierung der auf der Karte dargestellten Beschriftungen ist zur Verwendung bei allen diesen Programmen und Techniken sinnvoll. Aus diesem Grund ist es naheliegend, solche Verfahren direkt in PostgreSQL als Datenbankfunktion (stored procedure) zu implementieren. Diesen Weg geht die Implementierung, die initial vom Autor dieses Beitrags für den Kartenstil der deutschen Openstreetmap-Webseite programmiert wurde. Die aktuelle Version enthält zudem einige Verbesserungen von Mitgliedern der OSM Community⁴.

Beispiel (SQL Abfrage mit der „stored procedure“ *get_localized_placename*):

```
SELECT get_localized_placename(tags->'name', tags->'name:de',
                                tags->'int_name', tags->'name:en', false, way) AS NAME
FROM planet_osm_point
WHERE osm_id=424315709;

      name
-----
Bulgarien (България)
(1 Zeile)
```

-
- 1 Das freie Datenbanksystem PostgreSQL und dessen Erweiterung PostGIS ist unter <http://postgis.net> verfügbar aber auch Teil gängiger Linux-Distributionen
 - 2 <http://wiki.openstreetmap.org/wiki/Osm2pgsql>
 - 3 <http://imposm.org/>
 - 4 <https://github.com/giggls/mapnik-german-l10n>

Probleme und Verbesserungsmöglichkeiten der aktuellen Implementierung

Die aktuell im deutschen Kartenstil verwendete Lokalisierung kann aus vielen Gründen niemals ganz perfekt sein. Das Projekt ist prinzipbedingt mit vielen technischen aber auch politischen Problemen behaftet. Diese werden im Folgenden beschrieben.

Automatische Erzeugung lateinischer Bezeichnungen

Viele unbedeutende Objekte in der Openstreetmap-Datenbank haben gar keine internationale Bezeichnung. Hier stellt sich insbesondere bei nicht-lateinischen Alphabeten die Frage nach der Vorgehensweise. In vielen Fällen ist eine lateinische Transliteration sinnvoll. Im Falle von Straßen wird dieses Verfahren derzeit angewendet.

Beispiel:

Eine Straße in Moskau heißt „Новоостаповская улица“. Nach Transliteration ergibt sich „Novoostapovskaâ ulica“, was in der Karte dargestellt wird.

Transliterationen sind jedoch mit einer Reihe von Problemen behaftet. Einige davon sind technisch lösbar.

Nachfolgend werden bekannte Probleme dieser Art beschrieben. Es darf als sicher gelten, dass weitere existieren, die dem Autor bisher unbekannt sind.

- In manchen Sprachen (z.B. arabisch und hebräisch) werden keine Vokale geschrieben.
Beispiel: Die Transliteration von تهران (Teheran) liefert „thrān“
Ein prinzipielles, kaum technisch lösbares Problem.
- Derzeit kommt zur Transliteration die ICU-Bibliothek (International Components for Unicode) von IBM zum Einsatz. Diese Bibliothek verwendet eine in der thailändischen Sprache unübliche Transliteration nach ISO11940, die nicht derjenigen lokaler Straßenschilder entspricht
Beispiel: ถนนข้าวสาร, ICU: t̄hnn k̄hāw̄sār, üblich: Thanon Khao San
Dieses Problem ist prinzipiell durch die Verwendung einer anderen Bibliothek lösbar, sofern eine solche verfügbar ist. Man kann diese dann nur für die Zeichen der Thai Sprache verwenden.
- Die Transliteration chinesischer Zeichen ist nur in China selbst sinnvoll anwendbar. Die selben Zeichen werden in Japan unter der Bezeichnung Kanji verwendet und völlig anders ausgesprochen.
Beispiel: Tokio, jap. 東京 Transliteration (chin): dōng jīng

Dieses Problem ist prinzipiell dadurch lösbar, dass man den Ort der geografischen Objekte in die Transliteration einfließen lässt. Dieser Lösungsansatz wurde vom Autor implementiert und in die Software⁵ eingebaut.

Darstellung verschiedener Schriftzeichen innerhalb von Beschriftungen

Es sind nur wenige Schriftfonts mit Schriftzeichen aller Sprachen der Unicode-Tabelle in gleichbleibend guter Qualität verfügbar. Dies ist insbesondere bei Karten ein Problem, bei denen der lokale Name in nicht-lateinischer Schrift in Klammern erscheinen soll, denn die verfügbaren Renderer können die Schriftart oft nicht innerhalb eines Bezeichners ändern.

Die Implementierung des deutschen Kartenstils nutzt daher derzeit nur bei lateinischen, griechischen und kyrillischen Zeichensätzen den Kompromiss, den lokalen Namen in Klammern zu schreiben.

Dieses Problem kann technisch gelöst werden, und zwar entweder durch Weiterentwicklung der Renderer oder durch Verwendung eines Schriftfonts mit besserer Qualität.

Eventuell wäre es auch möglich, einen speziellen Kartenschriftfont zu erzeugen, der aus mehreren frei verfügbaren Einzelfonts zusammengesetzt wird.

Politische Probleme bei der Internationalisierung von Karten

Viele Regionen dieser Welt gehörten in der Vergangenheit zum Einflussbereich anderer Staaten als heute. Bekannte Beispiele sind die ehemaligen Kolonien europäischer Staaten oder deutsche Siedlungsgebiete in den heutigen Ländern Polen und Frankreich.

Wie bei vielen politischen Problemen ist eine technische Lösung in solchen Fällen schwierig. Aufgrund der Geschichte haben beispielsweise selbst kleinste Dörfer im heutigen Polen, im Elsass und in Lothringen einen deutschen Namen. Ob diese heute noch gebräuchlich sind oder nicht, lässt sich oft auch nicht mit Sicherheit sagen.

Ein Beispiel eines heute nicht mehr üblichen Namens ist die Bezeichnung „Nanzig“ für die Stadt Nancy in Lothringen. Die Einstufung solcher Namen in der OSM-Datenbank wahlweise als „name“ oder „old_name“ sollte den lokal aktiven Mappern überlassen bleiben, die es hoffentlich am besten wissen. Hier wurde name:de als „Nancy“ erfasst, weil der deutsche Name „Nanzig“ heute nicht mehr gebräuchlich ist.

Aus diesen Gründen wird bei Ortsnamen stets der aktuelle lokale Name in Klammern verwendet, wie z.B. *Stettin (Szczecin)*. Bei Straßennamen wird derzeit der deutsche Name abgekürzt in Klammern geschrieben, weil Straßennamen vor Ort selten mehrsprachig beschildert sind. So wird bei der Karlsgasse in Prag z.B. *Karlova*

⁵ <https://github.com/giggls/mapnik-german-l10n>

(*Karls*g.) verwendet.

Autoreninfo:

Sven Geggus betreibt am Fraunhofer IOSB mehrere Geodaten- und Kartenserver, die in Forschungsprojekten zum Einsatz kommen. In seiner Freizeit arbeitet er beim Projekt Openstreetmap mit und hat die Lokalisierung des deutschen Kartenstils programmiert.