

Nominatim

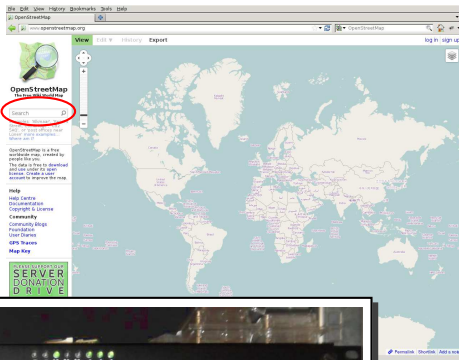
Wie die OSM-Suche intern funktioniert

Sarah Hoffmann
lonvia@denofr.de

14. Juli 2013

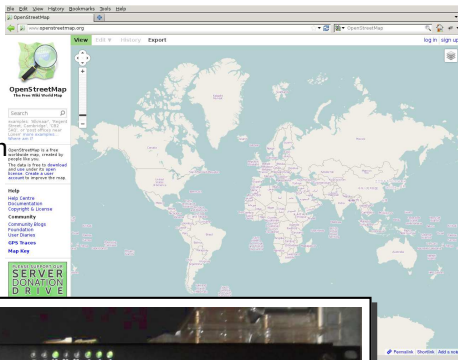


Nominatim



Überblick

- 1 Geocoder für OSM
- 2 Verarbeitung der OSM-Daten
- 3 Suchen & Finden



Kurze Geschichte

- 2007 erste Suchmaschine *Name Finder* von David Earl
- 2009 Nominatim geschrieben von Brian Quinion
- 2010 Hauptsuchmaschine auf osm.org
- 2012 Hardwarewechsel nach poldi mit minütlichen Updates

Anforderungen

- freie Textsuche
- heterogene Daten
- weltweit
- minütliche Updates
- hoher Durchsatz

Berlin

wien hütteldorfer str.

53.675231, 10.093289

Santander(Cantabria)

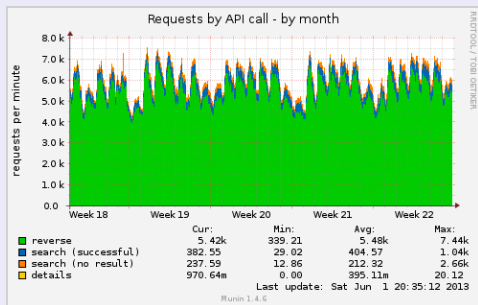
Deutschland, Karlsruhe, Leopoldstr. 9

Letchworth Lane, Letchworth, Hertfordshire, SG6 3NP, England

Anforderungen

- freie Textsuche
- heterogene Daten
- weltweit
- minütliche Updates
- hoher Durchsatz

- 120 Mio. Objekte
- 1 TB Daten
- ca. 100 Anfragen/s
10% Suchanfragen
90% Koordinatensuche



Vorverarbeitung

- 1 Objektimport
- 2 Adressbrechung
- 3 Suchbegriffe

Schritt 1: Objektimport

OpenStreetMap Nominatim x OpenStreetMap Nominatim x

← → ↻ nominatim.openstreetmap.org/details.php?osmtype=N&osmid=240062727

Rapperswil

Name:
Rapperswil (name)
Rapperschwil (name:gsw)
Type: place:town
Last Updated: 2012-10-10 10:42
Admin Level: 15
Rank: Village / Hamlet
Importance: 0.3 (estimated)
Coverage: Point
Centre Point: 47.2269198,8.8245459
OSM: node [240062727](#)
Extra Tags:
26177 (population)
<http://www.rapperswil-jona.ch> (website)

Address

Rapperswil (Type: place:town, node [240062727](#), 15, 0 [GOTO](#))
Rapperswil-Jona (Type: boundary:administrative, relation [1683921](#), 6, 0.0301679087093938 [GOTO](#))
Wahlkreis See-Gaster (Type: boundary:administrative, relation [1687207](#), 6, 0.201584383806251 [GOTO](#))

Objektimport

Rapperswil

Name:

Rapperswil (name)

Rapperschwil (name:gsw)

Type: place:town

Last Updated: 2012-10-10 10:42

Admin Level: 15

Rank: Village / Hamlet

Importance: 0.3 (estimated)

Coverage: Point

Centre Point: 47.2269198,8.8245459

OSM: node [240062727](#)

Extra Tags:

26177 (population)

<http://www.rapperswil-jona.ch> (website)

Address

Rapperswil (Type: place:town, node [240062727](#))

Rapperswil-Jona (Type: boundary:administrative, relation [1683921](#), 8, 0.03016790870939)



Klassifizierung

- Haupttag des Objekts

highway, railway, amenity, boundary, place, ...

- mehrere Einträge pro OSM-Objekt möglich

Objektimport

Rapperswil

Name:

Rapperswil (name)

Rapperschwil (name:gsw)

Type: place:town

Last Updated: 2012-10-10 10:42

Admin Level: 15

Rank: Village / Hamlet

Importance: 0.3 (estimated)

Coverage: Point

Centre Point: 47.2269198,8.8245459

OSM: node [240062727](#)

Extra Tags:

26177 (population)

<http://www.rapperswil-jona.ch> (website)

Address

Rapperswil (Type: place:town, node [240062727](#))

Rapperswil-Jona (Type: boundary:administrative, node [1683921](#), 8, 0.03016790870939)



Namen

name	loc_name	official_name
int_name	nat_name	reg_name
common_name	place_name	short_name
loc_name	alt_name	old_name
ref	iata	icao
operator		

+ Sprachvarianten

Objektimport

Rapperswil

Name:

Rapperswil (name)

Rapperschwil (name:gsw)

Type: place:town

Last Updated: 2012-10-10 10:42

Admin Level: 15

Rank: Village / Hamlet

Importance: 0.3 (estimated)

Coverage: Point

Centre Point: 47.2269198,8.8245459

OSM: node [240062727](#)

Extra Tags:

26177 (population)

<http://www.rapperswil-jona.ch> (website)

Address

Rapperswil (Type: place:town, node [240062727](#))

Rapperswil-Jona (Type: boundary:administrative, relation [1683921](#), 8, 0.03016790870939)



Suchrang

- regionale Bedeutung
- `admin_level` bzw. Objektklasse

Objektimport

Rapperswil

Name:

Rapperswil (name)

Rapperschwil (name:gsw)

Type: place:town

Last Updated: 2012-10-10 10:42

Admin Level: 15

Rank: Village / Hamlet

Importance: 0.3 (estimated)

Coverage: Point

Centre Point: 47.2269198,8.8245459

OSM: node [240062727](#)

Extra Tags:

26177 (population)

<http://www.rapperswil-jona.ch> (website)

Address

Rapperswil (Type: place:town, node [240062727](#))

Rapperswil-Jona (Type: boundary:administrative, relation [1683921](#), 8, 0.03016790870939)



Wichtigkeit

- Bedeutung bzw. Bekanntheit des Objekts
- berechnet über zugehörigen Wikipedia-Eintrag

Objektimport

Rapperswil

Name:

Rapperswil (name)

Rapperschwil (name:gsw)

Type: place:town

Last Updated: 2012-10-10 10:42

Admin Level: 15

Rank: Village / Hamlet

Importance: 0.3 (estimated)

Coverage: Point

Centre Point: 47.2269198,8.8245459

OSM: node [240062727](#)

Extra Tags:

26177 (population)

<http://www.rapperswil-jona.ch> (website)

Address

Rapperswil (Type: place:town, node [240062727](#))

Rapperswil-Jona (Type: boundary:administrative, relation [1683921](#), 8, 0.03016790870939)



Mittelpunkt

- 1 label
- 2 admin_centre (mit gleichem Rang)
- 3 gleichnamiger place-Node
- 4 geometrischer Mittelpunkt

Geometrie

Diverse Kleinigkeiten

Schritt 2: Adresse

OpenStreetMap Nominatim x

nominatim.openstreetmap.org/details.php?place_id=632871

Rapperswil

Name:

- Rapperswil** (name)
- Rapperswil** (name:gs)

Type: place:town

Last Updated: 2012-10-10 10:42

Admin Level: 15

Rank: Village / Hamlet

Importance: 0.3 (estimated)

Coverage: Point

Centre Point: 47.2269198,8.8245459

OSM: node [240062727](#)

Extra Tags:

- 26177 (population)
- <http://www.rapperswil-jona.ch> (website)

Address

Rapperswil (Type: place:town, node [240062727](#), 15, 0 [GOTO](#))

Rapperswil-Jona (Type: boundary:administrative, relation [1683921](#), 8, 0.0301679087083838 [GOTO](#))

Wahlkreis See-Gaster (Type: boundary:administrative, relation [1687207](#), 6, 0.201584383806251 [GOTO](#))

St. Gallen (Type: boundary:administrative, relation [1687006](#), 4, 0.45010134413059 [GOTO](#))

Zug (Type: place:state, node [313875456](#), 15, 0.281710115336351 [GOTO](#))

Glarus (Type: place:state, node [313872884](#), 15, 0.376757884448571 [GOTO](#))

St. Gallen (Type: place:state, node [313873770](#), 15, 0.518634444355963 [GOTO](#))

Appenzell Ausserrhoden (Type: place:state, node [240086063](#), 15, 0.561335488426556 [GOTO](#))

Appenzell Innerrhoden (Type: place:state, node [240071290](#), 15, 0.579327506068 [GOTO](#))

Argovia (Type: place:state, node [240054861](#), 15, 0.656460882823099 [GOTO](#))

Luzern (Type: place:state, node [572184965](#), 15, 0.666625746736388 [GOTO](#))

Switzerland (Type: boundary:administrative, relation [51701](#), 2, 0.731926502442076 [GOTO](#))

Switzerland (Type: place:country, node [1504546320](#), 15, 0.731185334140426 [GOTO](#))

ch (Type: place:country_code, 0 [GOTO](#))

Für jedes Objekt unabhängig berechnet

Navigation icons: back, forward, home, search, zoom in, zoom out, full screen, print, etc.

Adressberechnung: Kandidaten

- 1 Polygone, die das Objekt einschliessen
- 2 Nodes, die sich in der Nähe befinden
- 3 benannte Objekte aus `is_in/addr:*`
auch in grösserer Entfernung

... mit kleinerem Suchrang

Adresse - Beispiel

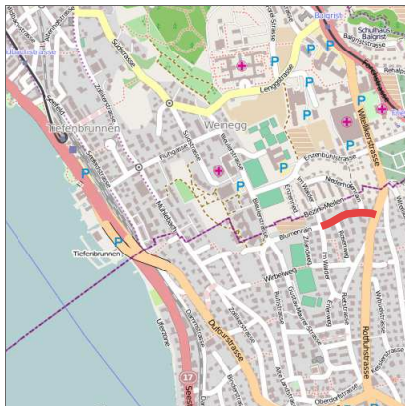
Blumenrain

Name:

Blumenrain (name)
 Type: highway:residential
 Last Updated: 2013-05-12 18:53
 Admin Level: 15
 Rank: Street / Major Landmark
 Importance: 0.1 (estimated)
 Coverage: Point
 Centre Point: 47.3477452,8.5760462
 OSM: way [8423413](#)
 Extra Tags:

Address

Blumenrain (Type: highway:residential, way [8423413](#), 15, 0 [GOTO](#))
 Tiefenbrunnen (Type: place:suburb, node [1079691186](#), 15, 0.014480708855923 [GOTO](#))
 Zollikon (Type: boundary:administrative, relation [1682246](#), 8, 0.0147999878189985 [GOTO](#))
 Bezirk Meilen (Type: boundary:administrative, relation [1691150](#), 6, 0.118611814897874 [GOTO](#))
 Zürich (Type: boundary:administrative, relation [1690227](#), 4, 0.102399064218626 [GOTO](#))
 8702 (Type: place:postcode, 0.00645521300830025 [GOTO](#))
 Schweiz (Type: boundary:administrative, relation [51701](#), 2, 0.647744225481821 [GOTO](#))
 Schweiz (Type: place:country, node [1504546320](#), 15, 0.648064581570847 [GOTO](#))
 ch (Type: place:country_code, 0 [GOTO](#))



Adresse - Beispiel

Blumenrain

Name:

Blumenrain (name)
 Type: highway:residential
 Last Updated: 2013-05-12 18:53
 Admin Level: 15
 Rank: Street / Major Landmark
 Importance: 0.1 (estimated)
 Coverage: Point
 Centre Point: 47.3477452,8.5760462
 OSM: way [8423413](#)
 Extra Tags:

Address

Blumenrain (Type: highway:residential, way [8423413](#), 15, 0 [GOTO](#))
 Tiefenbrunnen (Type: place:suburb, node [1079691186](#), 15, 0.014480708855923 [GOTO](#))
 Zollikon (Type: boundary:administrative, relation [1682246](#), 8, 0.0147999878189985 [GOTO](#))
 Bezirk Meilen (Type: boundary:administrative, relation [1691150](#), 6, 0.118611814897874 [GOTO](#))
 Zürich (Type: boundary:administrative, relation [1690227](#), 4, 0.102399064218626 [GOTO](#))
 8702 (Type: place:postcode, 0.00645521300830025 [GOTO](#))
 Schweiz (Type: boundary:administrative, relation [51701](#), 2, 0.647744225481821 [GOTO](#))
 Schweiz (Type: place:country, node [1504546320](#), 15, 0.648064581570847 [GOTO](#))
 ch (Type: place:country_code, 0 [GOTO](#))



Adresse - Beispiel

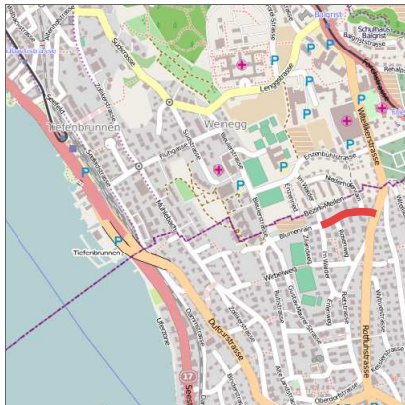
Blumenrain

Name:

Blumenrain (name)
 Type: highway:residential
 Last Updated: 2013-05-12 18:53
 Admin Level: 15
 Rank: Street / Major Landmark
 Importance: 0.1 (estimated)
 Coverage: Point
 Centre Point: 47.3477452,8.5760462
 OSM: way [8423413](#)
 Extra Tags:

Address

Blumenrain (Type: highway:residential, way [8423413](#), 15, 0 [GOTO](#))
 Tiefenbrunnen (Type: place:suburb, node [1079691186](#), 15, 0.014480708855923 [GOTO](#))
 Zollikon (Type: boundary:administrative, relation [1682246](#), 8, 0.0147999878189985 [GOTO](#))
 Bezirk Meilen (Type: boundary:administrative, relation [1691150](#), 6, 0.118611814897874 [GOTO](#))
 Zürich (Type: boundary:administrative, relation [1690227](#), 4, 0.102399064218626 [GOTO](#))
 8702 (Type: place:postcode, 0.00645521300830025 [GOTO](#))
 Schweiz (Type: boundary:administrative, relation [51701](#), 2, 0.647744225481821 [GOTO](#))
 Schweiz (Type: place:country, node [1504546320](#), 15, 0.648064581570847 [GOTO](#))
 ch (Type: place:country_code, 0 [GOTO](#))



Finale Adresse

- bester Kandidat aus jedem Rang: Polygon, is_in, nächster Node
- Place-Nodes müssen in die Hierarchie passen

Sonderfall: Hausnummer und POIs

Adresse vom Elternobjekt:

- 1 associatedStreet-Relation
- 2 addr:interpolation-Wege
- 3 Strasse von addr:street bzw. Ort von addr:place
- 4 nächste Strasse

Schritt 3: Suchbegriffe

Suchindex

Namenstoken

name-Tags des Objekts

Adresstoken

name-Tags der Adresskandidaten
`is_in, addr:*`
Postleitzahl

Beispiel Blumenrain:
Blumenrain

Tiefenbrunnen, Zollikon, Bezirk Meilen, Zürich,
8702, Schweiz, Switzerland, ...

Teilwörter und Normalisierung

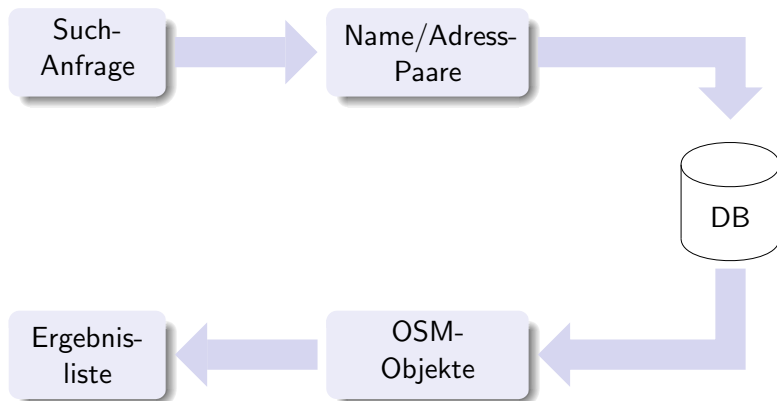
Normalisierung der Schreibweise

- Gross/Kleinschreibung
- Umlaute, Lateinisierung
- häufige Abkürzungen (str, pl, ...)

Zerlegung mehrteiliger Namen

- Teilung bei Leerzeichen, Komma, Klammern
- jeder Teil wird zusätzlich zum ganzen Wort gespeichert

Die Suche



Aufbereitung der Anfrage

Suchanfrage

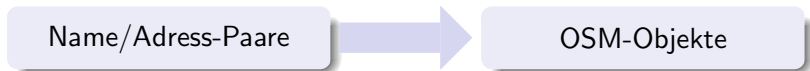
Bsp: Etzelstr. 32, Pfäffikon, SZ

Name/Adress-Paare

Etzel / st, 32, pfaffikon, sz
Etzel st / 32, pfaffikon, sz
sz / etzel, st, 32, pfaffikon
sz / etzel st, 32, pfaffikon

- Komma trennt Adressbestandteile
- Name am Anfang oder Ende
- Sonderbehandlung für Hausnummern, Ländernamen, etc.

DB-Suche



- Anfragen nach verschiedenen Heuristiken geordnet (wenig Zerlegung, volle Übereinstimmung, ...)
- einzeln an DB gesendet bis Resultate gefunden werden oder Limit erreicht

Partielle Matches

Suche kann abgebrochen werden, bevor die richtige Anfrage in der Liste gefunden wurde.

Übliche kurze Versionen mit `short_name` oder `loc_name` taggen.

Aber ich meinte doch...



Resultate müssen noch sortiert werden nach

- Wichtigkeit
- wie nah an der Suchanfrage
- (im Kartenausschnitt)

und aufbereitet

- Objektinformationen ergänzt
- Namen in gewünschte Sprache übersetzt

Die Zukunft

- 1 Postleitzahlen
- 2 tolerantere Schreibweise
- 3 Adressvorschläge
- 4 Dokumentation

Fragen

Website <http://nominatim.osm.org>

Source Code <http://github.com/twain47/Nominatim>

Mehr Informationen <http://wiki.osm.org/wiki/Nominatim>



Mit freundlichen Dank an Brian Quinion und Grant Slater (Serverfoto).

Folien by Sarah Hoffmann unter CC-by-SA 3.0