

Symbolische Suche in großen Geodatenbeständen

Themenstellung für eine Abschlussarbeit

Betreuer: Prof. Jörg Roth, Joerg.Roth@Ohm-hochschule.de

Aufgabenstellung

Mit Hilfe der bereits existierenden Geodatenbank soll eine textuelle Suche ermöglicht werden. Bei dieser Suche soll auf verschiedene Eingabemöglichkeiten eingegangen werden, die sich unterscheiden

- durch Reihenfolge und Anzahl der Hierarchiestufen von Ortsbegriffen (z.B. erst Stadt dann Straße, oder Stadt, Vorort, Straße, oder umgekehrte Reihenfolge),
- durch die Art der Trennzeichen zwischen Ortsbegriffen (z.B. Leerzeichen oder Komma)
- sowie durch weitere Trennworte (z.B. "in der Nähe", "am", "in", "bei").

Darüber hinaus können je nach Anfrage unterschiedliche Klassen von Ortsbegriffen verwendet werden, z.B. *Ortsname*, *Vorortsname*, *Straße*, *Objektyp* oder *Objektname*. Hieraus ergibt sich eine Vielzahl von Mustern.

Beispiel	Beschreibung
Nürnberg, Hohfederstraße	<i>Stadt, Straße</i> durch Komma getrennt
Laufamholz, Nürnberg	<i>Vorort, Stadt</i> durch Komma getrennt
Post, Laufamholz, Nürnberg	<i>Objektyp, Vorort, Stadt</i> durch Komma getrennt
Ammersee, Bayern	<i>Objektname, Bundesland</i> durch Komma getrennt
Ammersee, Deutschland	<i>Objektname, Staat</i> durch Komma getrennt
Klosterwirt, Fraueninsel, Chiemsee	<i>Objektname, Name des umgebenden Objektes, Name des umgebenden Objektes zweiter Stufe</i> jeweils durch Komma getrennt
München Parkplatz	<i>Stadt, Objektyp</i> durch Leerzeichen getrennt
Garten München	<i>Objektyp, Stadt</i> durch Leerzeichen getrennt
Hofbräuhaus München	<i>Objektname, Stadt</i> durch Leerzeichen getrennt
Post in Laufamholz	<i>Objektyp "in" Vorort</i>
Restaurant am Wolfgangsee	<i>Objektyp "am" Objektname</i>
Biergarten in der Nähe von München	<i>Objektype "in der Nähe von" Stadt</i>
Nürnberg, Post in Laufamholz	<i>Stadt, Objektyp "in" Vorort</i>
Gasthaus in der Nähe vom See, Bad Ischl	<i>Objektyp "in der Nähe von" Objektyp, Stadt</i>
Post in der Nähe Moritzbergstraße in Nürnberg	<i>Objektyp "in der Nähe von" Straße "in" Stadt</i>

Diese Tabelle gibt lediglich eine Übersicht über mögliche Beispiele. Weitere Muster, insbesondere viele Kombinationen von Mustern sind denkbar. In einem ersten Schritt soll in der Arbeit eine Übersicht über Muster erstellt und ein Satz geeigneter Muster identifiziert werden.

Die Datenbank erlaubt zurzeit keine direkte Zuordnung von Straßen oder Points of Interest (POIs) zu den dazugehörigen Städten, Dörfern, Landkreisen usw. Im Zuge dieser Arbeit sollen weitere Tabellen in dieser Datenbank automatisch durch ein Programm erstellt werden, um dies zu gewährleisten. Um eine schnelle Suche zu ermöglichen, müssen u.U. weitere Tabellen automatisch aus den existierenden Geodatenbeständen generiert werden.

Es soll ein Mechanismus zur Vervollständigung von Suchanfragen angeboten werden. Bei der Vervollständigung soll eine geordnete Liste möglicher Anfragen zurückgeliefert werden, die zum bisher eingegeben Text am besten passen. Die Vervollständigung soll dabei pro Treffer eine Sortierung der Suchbegriffe nach Hierarchiestufen durchführen, damit sich der Benutzer an eine einheitliche Eingabereihenfolge gewöhnt.

Neben der exakten Suche soll eine unscharfe Suche realisiert werden, die bei einigen fehlenden, falschen, vertauschten oder hinzugefügten Zeichen sinnvolle Treffer liefert.

Der Suchdienst soll als API bzw. als Klassenbibliothek gestaltet werden. Die Programmiersprache ist Java.

Die Aufgabe umfasst:

- Eine Recherche über
 - Suchverfahren in Datenbanken;
 - Mechanismen zur Autovervollständigung;
 - Klassifikation möglicher Eingabemuster, die Menschen für Orte verwenden. Hieraus soll ein geeigneter Satz unterstützter Eingabemuster abgeleitet werden.
- Die Konzeption eines Dienstes, der bei einer Suchanfrage ein mögliches Ergebnis zurückliefert. Dieser Dienst soll unterschiedliche Aufrufe unterstützen, insb. 1. Suche von Geobjekten bei vollständiger Eingabe, Rückgabe aller Treffer, die zu der Suche passen und 2. Eingabe eines Anfangstextes, Rückgabe aller sinnvollen Suchtexte, die dazu passen.
- Erstellen eines Programms, das aus den existierenden Tabellen weitere Tabellen erstellt, welche für die Suche benötigt werden.
- Die Implementierung und prototypische Integration in die vorhandene Datenbank.