

Institut für Formale Methoden der Informatik

Universität Stuttgart  
Universitätsstraße 38  
D–70569 Stuttgart

Studienarbeit Nr. 2432

## **Hierarchisierung und Darstellung von Geodaten**

Filip Krumpe

**Studiengang:** Informatik

**Prüfer/in:** Prof. Dr. Stefan Funke

**Betreuer/in:** Prof. Dr. Stefan Funke

**Beginn am:** 10.07.2013

**Beendet am:** 09.01.2014

**CR-Nummer:** E1, H5.2, I.3.5





## **Kurzfassung**

Die vorliegende Arbeit beschäftigt sich mit der Darstellung von geographischen Basisdaten im Gebiet der Bundesrepublik Deutschland. Die Arbeit basiert auf zwei Datensätzen des Bundesamtes für Kartographie und Geodäsie und visualisiert sowohl Namen verschiedener geographischer Objekte (u.a. Ortsnamen, Namen von Verwaltungsgebieten) als auch den Verlauf der Grenzen von Verwaltungsgebieten (u.a. Staatsgrenzen, Bundeslandgrenzen, Gemeindegrenzen). Die Darstellung ermöglicht eine Navigation innerhalb der Daten durch Zoom und Verschieben des dargestellten Datenausschnitts und das Hinzufügen und Entfernen von Details.

Um die Übersichtlichkeit der Darstellung zu gewährleisten, wurde eine Hierarchisierung der zugrundeliegenden Datensätze entwickelt und eine Filterung der anzuzeigenden Details implementiert. Zudem wurde, zur Beschleunigung der interaktiven Operationen, eine Approximation der Verwaltungsgrenzen implementiert.

In der vorliegenden Arbeit werden die zugrundeliegenden Datensätze GN250 und VG250 und das UTM-Koordinatensystem, in welchem die geographischen Daten referenziert sind, beschrieben. Der Algorithmus für die Approximation der Grenzverläufe sowie die Hierarchisierung der Daten wird erläutert. Zudem wird der grundsätzliche Ablauf des Programms skizziert und zentrale Aspekte des Programms, wie der Aufbau der SHAPE-Datei, welche die Grenzverläufe enthält, und die interne Verwaltungsstruktur der geographischen Daten, beschrieben. Des Weiteren wird die Datenbank beschrieben, welche die initialen Daten bereitstellt, die beim Programmstart geladen werden. Eine Zusammenfassung und ein Ausblick auf weitere Forschungsfragen schließen die Arbeit ab.



# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>7</b>
<b>2. Datengrundlage</b>	<b>9</b>
2.1. Geographische Namen: GN250 . . . . .	9
2.2. Verwaltungsgebiete: VG250 . . . . .	11
2.3. Koordinatensystem UTM . . . . .	13
<b>3. Vereinfachung und Strukturierung der geographischen Daten</b>	<b>15</b>
3.1. Approximation der Grenzverläufe . . . . .	16
3.2. Der Douglas-Peucker-Algorithmus . . . . .	17
3.3. Hierarchisierung der Daten . . . . .	18
<b>4. Verschiedene Aspekte der Implementierung</b>	<b>23</b>
4.1. Programmablauf . . . . .	23
4.2. Zentrale Aspekte des Programms . . . . .	25
4.3. Beschreibung der genutzten Datenbanken . . . . .	28
<b>5. Zusammenfassung und Ausblick</b>	<b>31</b>
<b>A. Anhang</b>	<b>33</b>
A.1. ATKIS-Objekte im Datensatz GN250 . . . . .	33
<b>Literaturverzeichnis</b>	<b>37</b>

# Abbildungsverzeichnis

---

2.1. UTM-Zonen in Europa (Quelle: <a href="https://de.wikipedia.org/wiki/Datei:LA2-Europe-UTM-zones.png">https://de.wikipedia.org/wiki/Datei:LA2-Europe-UTM-zones.png</a> ). . . . .	14
3.1. Vergleich der Darstellung des Datensatzes mit verschiedenen Approximationsleveln (von links nach rechts: $\epsilon = 0$ , $\epsilon = 500m$ , $\epsilon = 1000m$ ). . . . .	15
3.2. Vergleich der Darstellung mit Ortsfilter Kreisebene (links) und Bundesebene (rechts). .	16
3.3. Vergleich der Darstellung mit Verwaltungsgebietsfilter Gemeinde (links) und Bundesland (rechts). . . . .	19
3.4. Angezeigte Ortsnamen bei Filtereinstellung Staatsebene (EWZ > 300.000) in Berlin. .	21
4.1. Aufbau SHAPE-File-Header (Quelle: [S-D98]). . . . .	26
4.2. Aufbau SHAPE-File Record-Header (Quelle: [S-D98]). . . . .	26
4.3. Aufbau SHAPE-File Polyline Record-Body (Quelle: [S-D98]). . . . .	27

# Tabellenverzeichnis

---

3.1. Anzahl der darzustellenden Punkte bei gesamtem Bundesgebiet und Darstellung der Verwaltungsgebiete inklusive Kreise, und benötigte Zeit zur Abfrage der Daten nach Verschieben (Durchschnitt aus 10 Operationen). . . . .	17
3.2. Einwohnerzahlen aus dem Datensatz GN250 und die zugeordneten Hierarchielevel. .	20

# Verzeichnis der Listings

---

3.1. Douglas-Peucker-Algorithmus . . . . .	18
--	----

# 1. Einleitung

Die vorliegende Arbeit beschäftigt sich mit der Auswertung, Hierarchisierung und Darstellung von geographischen Basisdaten (Ortsbezeichnungen und Grenzverläufe der Verwaltungsgebiete) im Gebiet der Bundesrepublik Deutschland. Das Ziel der Arbeit war die Auswertung der Basisdaten und eine Visualisierung dieser Daten zu entwickeln.

Als Datenbasis der Arbeit wurden zwei Datensätze (GN250 und VG250) des Bundesamtes für Kartographie und Geodäsie verwendet. Der GN250-Datensatz enthält unter anderem 45.037 Ortsnamen. Der VG250-Datensatz enthält Informationen zu 16.707 Verwaltungsgebieten der Bundesrepublik Deutschland. Die Informationen des VG250-Datensatzes umfassen unter anderem die Grenzverläufe sämtlicher Verwaltungsgebiete sowie verschiedene Attribute der Gebiete, wie Zugehörigkeit in der nächsthöheren Verwaltungsebene usw.

Das Ergebnis der Arbeit ist eine visuelle Darstellung der Daten, welche durch verschiedene Operationen, wie Verschieben und Zoom, ein interaktives Erforschen der Daten ermöglicht. Bei der Entwicklung lag der Fokus auf der Reaktionsgeschwindigkeit des Programms auf Nutzereingaben und auf der Übersichtlichkeit der Darstellung.

Um die Reaktionsgeschwindigkeit auf Nutzereingaben zu erhöhen, wurde der Douglas-Peucker-Algorithmus implementiert, durch den eine Approximation der Grenzverläufe berechnet und so die Anzahl der darzustellenden Punkte reduziert werden kann. Zusätzlich wurde eine Strukturierung der Daten entwickelt, durch die der Nutzer in der Lage ist, verschiedene Teile und Aspekte des Datensatzes in der Visualisierung darzustellen oder auszublenden. Durch diese selektive Darstellung wird zum einen die Reaktionsgeschwindigkeit des Programms auf Nutzereingaben und zum anderen die Übersichtlichkeit der Darstellung verbessert.

## Gliederung

Die vorliegende Arbeit gliedert sich in drei große Kapitel, welche wiederum in einzelne Abschnitte unterteilt sind.

**Kapitel 2 – Datengrundlage:** Hier werden die zugrunde liegenden Datensätze GN250 und VG250 beschrieben. Zudem wird das darin verwendete UTM-Koordinatensystem erläutert.

**Kapitel 3 – Vereinfachung und Strukturierung der geographischen Daten:** Dieses Kapitel enthält Erläuterungen zu der durchgeführten Strukturierung der Daten und eine Beschreibung des Douglas-Peucker-Algorithmus, durch den die Approximation der Grenzverläufe berechnet wird.

## 1. Einleitung

---

**Kapitel 4 – Verschiedene Aspekte der Implementierung:** In diesem Kapitel werden verschiedene Aspekte der entwickelten Software beschrieben und verschiedene zentrale Komponenten erläutert, um dem Leser einen guten Einstieg in das Verständnis des entwickelten Programmcodes zu ermöglichen. Zentrale Abschnitte in diesem Kapitel sind eine schematische Darstellung des Programmablaufs und eine Beschreibung von zentralen Komponenten des Programms, wie dem Aufbau der SHAPE-Datei, in der die Grenzverläufe der Verwaltungsgebiete dargestellt sind und der internen Datenstruktur, in der die Daten während des Programmablaufs verwaltet werden. Zudem erläutert ein Kapitel die Datenbank, in der die Informationen gespeichert werden solange das Programm nicht läuft.

## 2. Datengrundlage

Als mögliche Datengrundlage wurden verschiedene Optionen getestet. Schlussendlich wurde entschieden die Datensätze GN250 (vgl. Dokumentation in [GN-13]) und VG250 (vgl. Dokumentation in [VG-13]) zu nutzen. Beide Datensätze werden im Rahmen der OpenData Initiative des GeoDatenZentrum des Bundesamtes für Kartographie und Geodäsie (<http://www.bkg.bund.de>) als Open-Source veröffentlicht. Diese Datensätze wurden insbesondere deshalb ausgewählt, weil sie im Zusammenhang mit dem INSPIRE-Projekt der EU stehen. Dadurch besteht die Hoffnung, dass ähnliche Datensätze auch für andere europäische Länder bereitgestellt werden und somit eine relativ einfache Erweiterung des Programms auf weitere Länder möglich ist.

Der Datensatz GN250 enthält die Namen von geographischen Objekten wie Gemeinden, Landschaften, Gebirgen, Bergen, Inseln, Flüssen usw. Der Datensatz orientiert sich an einem Kartenmaßstab von 1:250 000m. Die Daten sind in verschiedenen Koordinatensystemen angegeben. In der vorliegenden Arbeit werden die UTM-Koordinaten verwendet (siehe Abschnitt 2.3). Für eine genauere Beschreibung der Daten des GN250-Datensatzes siehe Abschnitt 2.1.

Der Datensatz VG250 enthält Informationen zu den Verwaltungsgebieten im Gebiet der Bundesrepublik Deutschland, beispielsweise den Verlauf der Verwaltungsgebietsgrenzen, verschiedene Attribute der Gebiete und deren Name. Zusätzlich enthält der Datensatz VG250\_EW noch die Einwohnerzahlen der verschiedenen Verwaltungsgebiete. Für eine genauere Beschreibung der Daten siehe Abschnitt 2.2.

### 2.1. Geographische Namen: GN250

Der Datensatz enthält Namen von geographischen Objekten der Art (vgl. [GN-13]):

- Siedlung
- Verkehr
- Vegetation
- Gewässer
- Relief
- Gebiet

## 2. Datengrundlage

---

Sämtliche verzeichneten Objekte sind in verschiedene Klassen unterteilt, die mit einem ATKIS-Objektcode bezeichnet werden. So werden beispielsweise Ortslagen, Gewässer, Straßenverkehrsanlagen usw. unterschieden. Eine Auflistung aller vorhandenen Objektarten und deren Unterklassen sind im Anhang unter A.1 verzeichnet.

In der vorliegenden Arbeit liegt der Fokus auf den ATKIS-Objekten AX\_Ortslage und den Namen der verschiedenen Verwaltunggebiete (ATKIS-Objekte: AX\_Bundesland, AX\_Gemeinde, AX\_KreisRegion, AX\_Nationalstaat und AX\_Regierungsbezirk). Alle zusätzlich vorhandenen Objekte können bei der entwickelten Visualisierung zusätzlich eingeblendet werden.

### 2.1.1. Auszug aus den Attributen der geographischen Objekte

Unter anderem folgende Attribute werden durch den GN250-Datensatz bereitgestellt (vgl. [GN-13]):

**OBA:** ATKIS-Typ des Objekts (siehe Liste der genutzten Objektarten im Anhang A.1)

**OBA\_Wert:** ATKIS-Code des Objekts

**Name:** Name des Objekts

**Zusatz:** Zusatz zur genaueren Beschreibung des Objekttyps

**Hoehe:** Höhe des Objekts (bei Typen Ortslage und Besonderer Geländepunkt)

**Hoehe\_Ger:** Gerechnete Höhe des Objekts (für Typ Ortslage)

**EWZ:** Einwohnerzahl von Gemeinden

**EWZ\_Ger:** Gerechnete Einwohnerzahl für Typ Ortslage

**Gemeinde:** Name der zugehörigen Gemeinde

**Kreis:** Name des zugehörigen Kreises

**RegBezirk:** Name des zugehörigen Regierungsbezirks

**Bundesland:** Name des zugehörigen Bundeslands

**Staat:** Name des zugehörigen Staates

**UTMRE:** Rechtswert der UTM-Koordinaten

**UTMHO:** Hochwert der UTM-Koordinaten

**Box\_UTM:** Kleinstes umschließendes Rechteck des Objekts in UTM- Koordinaten

## 2.2. Verwaltungsgebiete: VG250

Der VG250-Datensatz beinhaltet sämtliche Verwaltungsgebiete der Deutschen Bundesrepublik (vgl. Dokumentation in [VG-13]). Der Datensatz stellt sowohl Attribute der einzelnen Gebiete als auch deren Grenzverläufe bereit. Die Daten unterteilen sich in sechs Ebenen: Staats-, Bundeslandes-, Regierungsbezirks-, Kreis-, Verwaltungsgemeinschaft und Gemeindeebene.

Die verfügbaren Daten werden in zwei verschiedenen Strukturen zur Verfügung gestellt (vgl. [VG-13]):

**VG250\_Kompakt:** Hier ist jedes Grenzsegment genau einmal gespeichert. Für jedes Grenzsegment ist über ein Attribut festgelegt, welches die höchste Ebene ist, der das Segment angehört.

**VG250\_Ebenen:** Hier sind für jede Verwaltungsebene eigene Datensätze vorhanden, die alle benötigten Grenzsegmente enthalten. Grenzverläufe werden hier redundant gespeichert, wenn sie in verschiedenen Verwaltungsebenen enthalten sind.

Der Datensatz beinhaltet folgende Informationen:

1. Informationen zu den Verwaltungsflächen (VG250\_F.xxx - Codiert als SHAPE- Objekte)
2. Informationen zu den Grenzlinien (VG250\_L.xxx - Codiert als SHAPE- Objekte)
3. Informationen zu den Attributen der Elemente (Grenzen) (VG250\_NAM.dbf)
4. Informationen zu der hierarchischen Struktur der Objekte (Verwaltungsstrukturen) (VG250\_ISN.dbf)

Die vorliegende Arbeit nutzt aus dem Datensatz lediglich die Informationen zu den Grenzverläufen. Die Flächen als solche werden nicht betrachtet.

### 2.2.1. Attribute der Flächen

Der Datensatz VG250 stellt für die gegebenen Flächen folgende Attribute bereit (vgl. [VG-13]):

**USE:** Beschreibt die unterste Verwaltungseinheit, der die Fläche zugehörig ist:

- 1: Staat
- 2: Bundesland
- 3: Regierungsbezirk
- 4: Kreis
- 5: Verwaltungsgemeinschaft
- 6: Gemeinde
- 11: Bodensee Deutschland
- 12: Bodensee Ausland

**RS:** Regionalschlüssel. Er stellt die Verbindungen zwischen den Flächen dar:

## **2. Datengrundlage**

---

- 1.-2. Stelle = Kennzahl des Bundeslandes
- 3. Stelle = Kennzahl des Regierungsbezirks
- 4.-5. Stelle = Kennzahl des Kreises
- 6.-9. Stelle = Verbandsschlüssel
- 10.-12. Stelle = Gemeindekennzahl

**GF:** Trennung von Land und Wasserfläche:

- 1: Gewässer Staats- und Landesebene
- 2: Gewässerfläche
- 3: Landfläche Staats- und Landesebene
- 4: Landfläche

**Gen:** Geographischer Name

**DES:** Amtliche Bezeichnung der Verwaltungseinheit

**ISN:** Numerische Kodierung des obigen DES' Attributs

**BEMERK:** Eventuell differenzierte Beschreibung für DES

**NAMBILD:** Binär. Beschreibt, ob DES dem Namen vorangestellt werden soll (beispielsweise 1 = "Kreis ..." oder 0 = "Salzlandkreis")

**AGS:** Amtlicher Gemeindeschlüssel:

- 1.-2. Stelle = Kennzahl des Bundeslandes
- 3. Stelle = Kennzahl des Regierungsbezirks
- 4.-5. Stelle = Kennzahl des Kreises
- 6.-8. Stelle = Gemeindekennzahl

**RS\_ALT:** Verbindungsschlüssel zu alten Datenbeständen

**WIRKSAMKEIT:** Beschreibt als Datum den Zeitpunkt, ab dem die letzte Änderung wirksam ist

**DEBKID:** Verknüpfung zu DLM250

## 2.2.2. Attribute der Grenzlinien

Für die Grenzlinien sind folgende Attribute vorhanden (vgl. [VG-13]):

**USE:** Die höchste Ebene, auf der die Grenze eine Verwaltungseinheit begrenzt:

- 1: Staatsgrenze
- 2: Bundeslandgrenze
- 3: Regierungsbezirksgrenze
- 4: Kreisgrenze
- 5: Verwaltungsgemeinschaftsgrenze
- 6: Gemeindegrenze
- 99: Küstenlinie: Trennt Land und Wasserfläche innerhalb einer Verwaltungseinheit

**LED:** Rechtliche Definition des Grenzabschnitts:

- 1: rechtlich festgelegte Grenze
- 2: rechtlich **nicht** festgelegte Grenze
- 9: Küstenlinie. Keine Bedeutung als Trennungsline für Verwaltungsgebiete

**GM5:** Entscheidet nur bei USE = 5:

- 8: Verwaltungsgrenze, die ausschließlich gemeinschaftsfreie Gemeinden umfasst
- 0: alle anderen Fälle

**EWZ:** Einwohnerzahl

## 2.3. Koordinatensystem UTM

Zur Referenzierung der geographischen Objekte werden UTM-Koordinaten verwendet, das Koordinatensystem wird unter anderem in [UTM09] und [WP-13] beschrieben. Es unterteilt die Erde vertikal in  $6^{\circ}$  breite Streifen. Die erstellten Zonen werden ab dem  $180^{\circ}$  Längengrad West in Östliche Richtung nummeriert. So ergibt sich für die Zone, in der der Großteil von Deutschland liegt, die Zone 32 ( $6^{\circ}$  bis  $12^{\circ}$  östlicher Länge) und zu kleineren Teilen die Zonen 31 und 33 (vgl. dazu auch 2.1).

In vertikaler Richtung werden die Zonen in  $8^{\circ}$  breite Zonenfelder unterteilt. Dabei wird bei  $80^{\circ}$  Süd begonnen. Den Zonenfeldern werden die Buchstaben des Alphabets zugeordnet, wobei bei C begonnen wird, die Buchstaben I und O ausgelassen werden, um eine Verwechslung mit den Ziffern 1 und 0 zu vermeiden. Das letzte Zonenfeld X ist mit  $12^{\circ}$  breiter als die restlichen Zonenfelder. So ergibt sich, dass Deutschland in den Zonen 31, 32 und 33 und in den Zonenfeldern T und U liegt (siehe Abbildung 2.1).

## 2. Datengrundlage

---



**Abbildung 2.1.:** UTM-Zonen in Europa (Quelle: <https://de.wikipedia.org/wiki/Datei:LA2-Europe-UTM-zones.png>).

Um Orte auf in der entsprechenden Zone zu referenzieren wird die Zone “ausgerollt” und das so entstandene Band mit einem karthesischen Koordinatensystem überzogen. Der Äquator bildet die X-Achse und der Mittelmeridian die Y-Achse des Koordinatensystems.

Um negative Werte bei der Beschreibung der Punkte zu vermeiden, werden von der X-Koordinate der Punkte 500.000m abgezogen, der Mittelmeridian erhält also die X-Koordinate 500.000m. Somit werden alle Orte in der Zone auf X-Werte zwischen 100.000 und 899.999 abgebildet.

Werden Orte auf der Nordhalbkugel beschrieben, so wird der Äquator mit dem Y-Wert 0 angenommen. Um negative Werte auf der Südhalbkugel zu vermeiden, werden den dort liegenden Punkten 10.000.000m von der Y-Koordinate abgezogen, wodurch sie ebenfalls ausschließlich positive Werte erhalten (vgl. [UTM09] und [WP-13]).

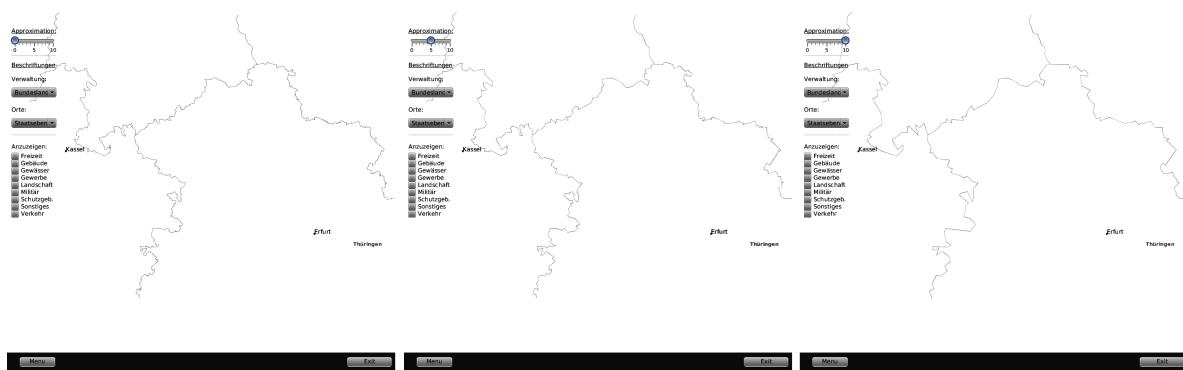
Auch Orte außerhalb einer bestimmten Zone können in dem Koordinatensystem der Zone dargestellt werden, allerdings wird dadurch eine zunehmende Verzerrung in Kauf genommen (vgl. [WP-13]). Die in dieser Arbeit verwendeten geographischen Objekte werden alle im Koordinatensystem der 32. Zone referenziert.

### 3. Vereinfachung und Strukturierung der geographischen Daten

Die in Kapitel 2 beschriebenen geographischen Basisdaten umfassen unter anderem 119.645 geographische Namen, davon 45.037 Ortsnamen, sowie 36.984 Segmente von Verwaltungsgebietsgrenzen. Diese Menge an Daten darzustellen und die Visualisierung interaktiv erkundbar zu machen, bringt verschiedene Probleme mit sich:

- Durch eine große Menge an Punkten, die bei der Darstellung des Datensatzes gezeichnet werden müssen, verlängert sich die Zeit, die benötigt wird, um eine Interaktion des Nutzers zu behandeln.
- Werden zu viele Details auf zu kleinem Raum dargestellt, so verringert Überdeckung die Wahrnehmbarkeit der Daten (vgl. Abb. 3.2 sowie Abb. 3.3).

Der VG250-Datensatz enthält die Grenzen der Verwaltungsgebiete der Bundesrepublik. Die Grenzen sind durch ca. 37.000 Grenzsegmente beschrieben. Der Verlauf der Grenzsegmente wird durch eine Menge an Punkten beschrieben, die auf dem Segment liegen (1.034.257 Punkte für alle Segmente). Vor allem auf hohen Zoomstufen, beispielsweise wenn die Darstellung den kompletten Datensatz umfasst, scheint eine Approximation der Grenzverläufe ein vielversprechender Ansatz zu sein, um die Menge der zu zeichnenden Punkte zu verringern ohne dabei sichtbare Details zu verlieren (siehe Abb. 3.1). Der Douglas-Peucker-Algorithmus, der in der vorliegenden Arbeit zur Approximation der Grenzverläufe verwendet wird, ist in Abschnitt 3.1 beschrieben.

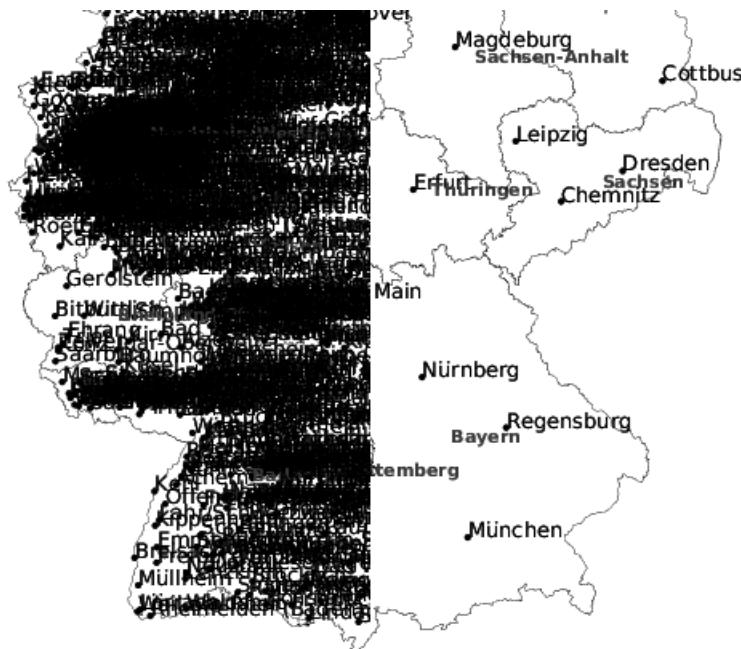


**Abbildung 3.1.:** Vergleich der Darstellung des Datensatzes mit verschiedenen Approximationsleveln (von links nach rechts:  $\epsilon = 0$ ,  $\epsilon = 500m$ ,  $\epsilon = 1000m$ ).

### 3. Vereinfachung und Strukturierung der geographischen Daten

---

Neben der Approximation der Grenzverläufe, die vor allem die Reaktionsgeschwindigkeit des Programms verbessert, verspricht eine weitere Maßnahme eine Reduktion der darzustellenden Daten: Die Hierarchisierung der darzustellenden Informationen. Im Zusammenhang mit einer Filtermöglichkeit für den Nutzer, wird zum einen wiederum die Anzahl der zu zeichnenden Punkte reduziert, zum anderen die Wahrnehmung der Daten erst möglich. Eine ungefilterte Darstellung aller Information würde durch Überdeckung sämtliche Information nicht lesbar machen (siehe Abb. 3.2). Die Hierarchisierung der Daten ist in Abschnitt 3.3 beschrieben.



**Abbildung 3.2.:** Vergleich der Darstellung mit Ortsfilter Kreisebene (links) und Bundesebene (rechts).

## 3.1. Approximation der Grenzverläufe

Wie in Abbildung 3.1 zu sehen ist, stellt die Approximation der Grenzverläufe eine gute Möglichkeiten dar, um die Anzahl der darzustellenden Punkte zu verringern. Dabei wird der Verlauf der Grenzsegmente nur in Details verändert, welche in großen Zoomstufen ohnehin nicht sichtbar wären (siehe Abbildung 3.1). Zugleich beschleunigt sich die Zeit zur Abfrage der Daten bei einer Interaktion des Nutzers (siehe Tabelle 3.1). Der verwendete Douglas-Peucker-Algorithmus zur Approximation der Liniensegmente wird im Folgenden erläutert.

Epsilon	Anzahl Punkte	Zeit in ms
0	330.171	244
100m	82.671	172
200m	52.647	165
300m	40.447	140
400m	34.179	137
500m	30.301	140
600m	27.545	131
700m	25.773	136
800m	24.372	135
900m	23.277	133
1km	22.482	135

**Tabelle 3.1.:** Anzahl der darzustellenden Punkte bei gesamtem Bundesgebiet und Darstellung der Verwaltungsgebiete inklusive Kreise, und benötigte Zeit zur Abfrage der Daten nach Verschieben (Durchschnitt aus 10 Operationen).

## 3.2. Der Douglas-Peucker-Algorithmus

Der Douglas-Peucker Algorithmus (vgl. [DP73]) ist ein Algorithmus zur Vereinfachung von Liniensegmenten, die durch eine Menge von Punkten definiert sind. Der Algorithmus (siehe Listing 3.1) entfernt Punkte, die zum Verlauf einer Linie einen Abstand kleiner einem gegebenen Epsilon besitzen, aus der Linie. Durch den Entwurf des Algorithmus wird gesichert, dass von der ursprünglichen Linie keine groben Details verlorengehen.

### Berechnung des Abstandes Punkt-Gerade

Um den Abstand eines Punktes  $t$  zur Geraden durch die Punkte  $p_1$  und  $p_2$  zu berechnen, wird folgende Abstandsformel (siehe Gleichung 3.1, vgl. [Wei13]) verwendet:

$$(3.1) \quad d = |v * r| = \frac{|(x_2 - x_1)(y_1 - y_t) - (x_1 - x_t)(y_2 - y_1)|}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$

### Verknüpfen der rekursiven Ergebnisse

Für den rekursiven Aufruf der Douglas-Peucker-Funktion wird die ursprüngliche Linie in zwei Segmente unterteilt: 0 bis  $i$  und  $i$  - Ende. Auf beiden wird rekursiv die Douglas-Peucker Funktion ausgeführt. Der Start und Endpunkt der Sequenz bleibt in jedem Fall erhalten, weshalb der Punkt  $i$  im Result doppelt vorhanden ist und entfernt werden muss.

### 3. Vereinfachung und Strukturierung der geographischen Daten

---

---

#### Listing 3.1 Douglas-Peucker-Algorithmus

---

```
funktion Douglas-Peucker(points, epsilon)
    maxDistance := 0;
    indexMaxDist := 0;
    for i := 1 to line.size - 2 do:
        if (distance(points[0], points[points.size - 1], points[i]) > maxDistance)
            maxDistance := distance(points[0], points[line.size - 1], points[i]);
            indexMaxDist := i;
        end if
    end do
    if (maxDistance < epsilon)
        return {points[0], points[points.size - 1]};
    else
        return removeDoublePoint( concat (
            Douglas-Peucker({points[0] ... points[i]}, epsilon),
            Douglas-Peucker({points[i] ... points[points.size -1]}, epsilon)));
    end if
end function

function distance(p1, p2, t)
    return (|(x_2 - x_1)(y_1 - y_t) - (x_1 - x_t)(y_2 - y_1)|) /
        (sqrt((x_2 - x_1)^2 + (y_2 - y_1)^2))

function concat(list1, list2)
    return {list1[0 .. length-1], list2[0..length]};
```

---

## 3.3. Hierarchisierung der Daten

Wie in Abbildung 3.2 und in Abbildung 3.3 zu sehen, ist eine Darstellung zu vieler Details zugleich nicht möglich. Für die darzustellenden Daten wurden deshalb Hierarchisierungen erstellt, die in diesem Abschnitt beschieben sind. Dem Nutzer wird die Möglichkeit gegeben durch eine Art Filterfunktion zusätzliche Details anzeigen zu lassen. Diese Filter wurden so implementiert, dass bei der Auswahl der Verwaltungsgebietsdaten ausschließlich die Daten der gewählten Ebene dargestellt werden. Bei der Auswahl einer Ebene für die Ortsnamen werden die ausgewählte Ebene und alle Daten der darüberliegenden Ebenen dargestellt. Zusätzlich wurden verschiedene Kategorien der geographischen Namen implementiert, die nicht hierarchisch aufgebaut sind und vom Benutzer einfach zur Darstellung hinzugefügt oder daraus entfernt werden können.

In Abschnitt 3.3.1 wird zuerst die Hierarchisierung der Verwaltungsgebietsdaten erläutert, bevor darauf aufbauend in Abschnitt 3.3.2 die Hierarchisierung der geographischen Namen vorgestellt wird.

### 3.3.1. Hierarchisierung der Verwaltungsgebietsdaten (VG250)

Die Hierarchisierung der Verwaltungsgebietsdaten erfolgt durch die Zugehörigkeit der verschiedenen Gebiete zu Verwaltungsebenen. Diese Ebenen sind in den Basisdaten bereits angelegt (vgl. [VG-13])

und benötigen keine weitere Änderung. Die Verwaltungsgebietsgrenzen sind demnach folgender Hierarchie zugeordnet:

- Staatsebene
- Bundeslandebene
- Regierungsbezirksebene
- Kreisebene
- Verwaltungsgemeinschaftsebene
- Gemeindeebene
- (Küstenlinie) - die Segmente dieser Kategorien werden immer angezeigt.



**Abbildung 3.3.:** Vergleich der Darstellung mit Verwaltungsgebietsfilter Gemeinde (links) und Bundesland (rechts).

#### 3.3.2. Hierarchisierung der geographischen Namen (GN250)

Wie in Abschnitt 2.1 bereits angedeutet, bezeichnen die geographischen Namen eine Vielzahl von unterschiedlichen Objekten. Für die vorliegende Arbeit waren vor allem zwei Objekte interessant, die Namen von Orten (ATKIS: AX\_Ortslage) und die Bezeichnungen der Verwaltungsgebiete (ATKIS: AX\_Nationalstaat, AX\_Bundesland, AX\_Regierungsbezirk, AX\_KreisRegion, AX\_Verwaltungsgemeinschaft\_ATKIS, AX\_Gemeinde).

Die Namen der **Verwaltungsgebiete** wurden nach ihrem ATKIS-Objekttyp und analog zur Hierarchisierung der Verwaltungsgebietsdaten in die Hierarchiestufen Staatsebene, Bundeslandebene usw. unterteilt. Um den Nutzer nicht zu überfordern, werden in der Darstellung ausschließlich Namen der gewählten Ebene und nicht zusätzlich die der darüberliegenden Ebenen angezeigt.

Ein erster Versuch zur Hierarchisierung der **Ortsnamen** wurde anhand der Einwohnerzahlen aus dem Datensatz unternommen. Dabei stellte sich jedoch heraus, dass das Attribut Einwohnerzahl (EWZ)

### 3. Vereinfachung und Strukturierung der geographischen Daten

---

Einwohnerzahl	Hierarchielevel
> 300.000	Staatsebene
> 100.000	Bundesebene
> 50.000	Regierungsbezirksebene
> 20.000	Kreisebene
> 10.000	Verwaltungsgemeinschaftsebene
< 10.000	Gemeindeebene

**Tabelle 3.2.:** Einwohnerzahlen aus dem Datensatz GN250 und die zugeordneten Hierarchielevel.

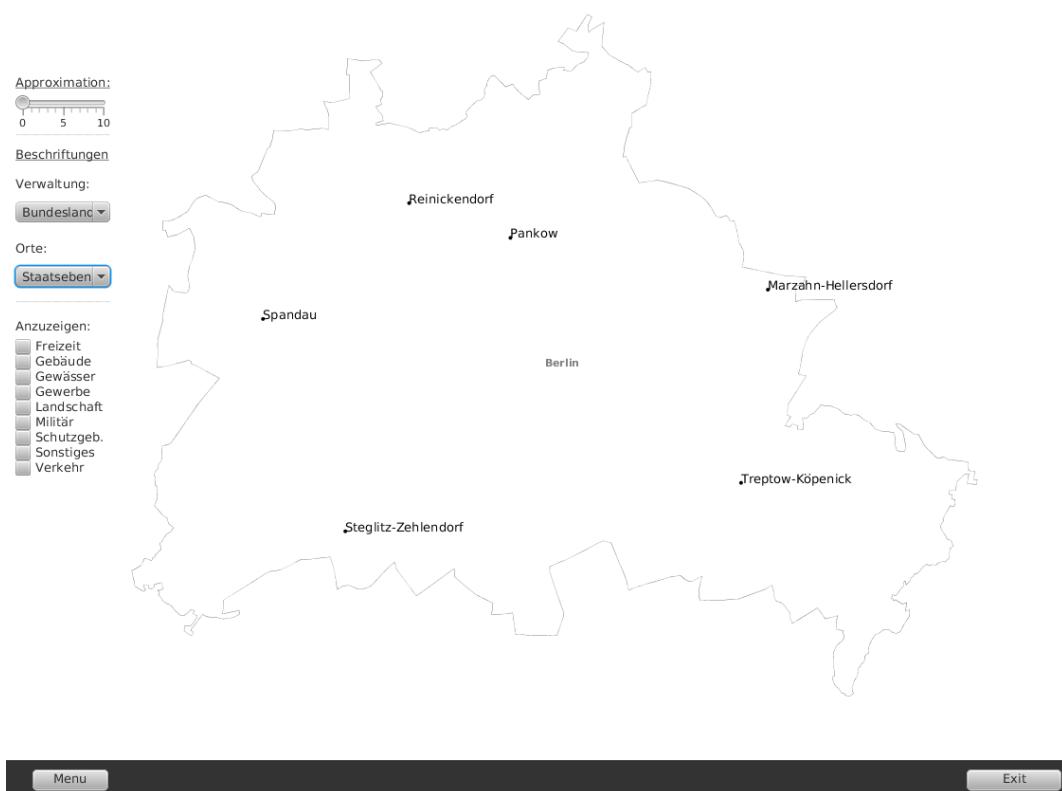
in den Daten zum Großteil nicht vorhanden war. Ein zweiter Versuch basierte auf der gerechneten Einwohnerzahl (EWZ\_GER), welche brauchbare Ergebnisse brachte. Die Hierarchielevel wurden analog zu den Verwaltungsleveln definiert. Die Zuordnung der gegebenen Einwohnerzahlen zu den entsprechenden Leveln ist in Tabelle 3.2 zu sehen.

Es ist zu beachten, dass die gerechnete Einwohnerzahl nicht der offiziellen Einwohnerzahl der jeweiligen Orte entspricht. So besitzt Stuttgart laut Datensatz beispielsweise eine Einwohnerzahl von 157.514, dafür sind die einzelnen Stadtteile extra in den Daten enthalten (beispielsweise Botnang: 13.633, Bad Cannstadt: 45.584, Degerloch: 28.091). Diese Besonderheit führt dazu, dass beispielsweise in Berlin die Ortsnamen von Marzahn-Hellersdorf, Pankow, Reinickendorf, Spandau und Steglitz-Zehlendorf der Staatsebene zugeordnet werden (siehe Abbildung 3.4), was dem intuitiven Verständnis widerspricht.

Die verbleibenden SHAPE-Typen wurden einer der folgenden Kategorien zugeordnet und können vom Nutzer zusätzlich ein- oder ausgeblendet werden.

- Schutzgebiete
- Gewässer
- Landschaft
- Gebäude
- Freizeit
- Gewerbefläche
- Verkehr
- Militär
- Sonstiges

### 3.3. Hierarchisierung der Daten



**Abbildung 3.4.:** Angezeigte Ortsnamen bei Filtereinstellung Staatsebene (EWZ > 300.000) in Berlin.



# 4. Verschiedene Aspekte der Implementierung

In diesem Kapitel wird der Leser in die Funktion des Programms eingeführt, um einen leichteren Einstieg in die Lektüre des Programmcodes zu ermöglichen. Um dieses Ziel zu erreichen, wird in Abschnitt 4.1 zuerst der gesamte Programmablauf skizziert, daraufhin werden in Abschnitt 4.2 zentrale Funktionen des Programms beschrieben und letztendlich in Abschnitt 4.3 ein Überblick über die Datenbanktabellen gegeben, welche für die Speicherung der Daten außerhalb des Programms verwendet werden.

## 4.1. Programmablauf

Der Programmablauf unterteilt sich grob in zwei Phasen:

- Initialisierungsphase
- Interaktionsphase

Beide Phasen werden gesondert beschrieben. Beim Start durchläuft das Programm immer zuerst die Initialisierung, an deren Ende eine Visualisierung des gesamten Datensatzes steht. Daran schließt die Interaktionsphase an, in der der User durch Zoom und Verschieben durch den Datensatz navigiert oder durch die vorhandenen Steuerungselemente Details hinzufügen- oder abschalten kann.

### 4.1.1. Initialisierungsphase

In der Initialisierungsphase werden folgende Schritte ausgeführt:

1. Laden der GUI-Komponenten und Initialisieren der GUI
2. Laden der Visualisierungsdaten:
  - a) Laden der geographischen Daten aus der Datenbank durch GN250Helper (geographische Namen) und VG250Helper (Verwaltungsgebiete)
  - b) Initialisieren des StorageHelpers mit den geladenen Daten:
    - Basierend auf der Boundingbox über alle Verwaltungsgebiete (also bisher deutschlandweit) wird eine Gridstruktur erzeugt, die 512x512 Gridzellen enthält.
    - Alle Grenzsegmente, die durch den VG250Helper geladen wurden, werden:

#### 4. Verschiedene Aspekte der Implementierung

---

- von UTM in lokale Koordinaten transformiert (Nullpunkt im nordwestlichen Eck der Boundingbox der Daten)
  - in einer interne Liste des Storage-Helpers gespeichert und
  - der Index des Segments in den Gridzellen gespeichert (sortiert nach BorderLevel), die die Boundingbox des Segments überdecken.
- Jedes Segment wird mit Hilfe des VGApproximators mit neun verschiedenen Genauigkeiten approximiert (Epsilon = 100m, 200m, ..., 900m, siehe Kapitel Der Douglas-Peucker-Algorithmus) und in separaten Listen gespeichert (BEACHTE: Der Index eines Segments ist immer der selbe, weshalb in den Gridzellen jeweils kein weiterer Speicheraufwand entsteht).
  - Alle geographischen Namen, die durch den GN250Helper geladen wurden, werden:
    - von UTM in lokale Koordinaten transformiert (Nullpunkt im nordwestlichen Eck der Boundingbox der Daten),
    - in einer internen Liste des Storage-Helpers gespeichert und
    - der Index des Segments in der Gridzelle gespeichert (sortiert nach LocationLevel), die den Punkt überdeckt.

#### 3. Setzen von GUI-Metadaten:

- a) Setzen der initialen Daten des Visualisierungsfensers (Breite und Höhe des Bereichs der Visualisierung in Pixeln)
- b) Setzen der Größe des Datensatzes (Ausdehnung in Meter in x-Richtung, Ausdehnung in Meter in y-Richtung)
- c) Aus den zuvor gesetzten Informationen werden die initialen Informationen (Größe der Boundingbox der aktuell dargestellten Daten und Maßstab der Visualisierung (Meter pro Pixel) und Mittelpunkt der Daten) berechnet, damit im ersten Bild der gesamte Datenbereich dargestellt wird.

#### 4. Anzeigen der Visualisierung

##### 4.1.2. Interaktionsphase

Verschiedene Interaktionen mit den visualisierten Daten sind möglich:

Zoom: Dabei wird der Maßstab der Visualisierung verändert. Ausgehend von dem neuen Maßstab wird die Boundingbox der aktuell dargestellten Daten neu berechnet und die Visualisierung mit den neuen Informationen aktualisiert.

Verschieben: Dabei wird der Mittelpunkt der Visualisierung verschoben und die Visualisierung mit der dadurch verschobenen Boundingbox aktualisiert.

Details hinzufügen oder abschalten: Verschiedene Optionen sind möglich:

1. Ändern des Levels der Verwaltungsgebiete: Die Visualisierung wird mit den Daten des angegebenen und allen darüber liegenden Level aktualisiert.
2. Ändern der Liste der zusätzlichen Level: Die Visualisierung wird aktualisiert, wobei alle Daten der zusätzlichen Level dargestellt werden.
3. Ändern des Approximationslevels: Aktualisieren der Daten mit den Grenzsegmenten der angegebenen Approximationsstufe.

Ändern der Fenstergröße: Hierbei ändert sich die Größe des Bereichs für die Visualisierung, dadurch muss die Größe der Datenbox entsprechend angepasst werden und die Visualisierung mit den neuen Daten aktualisiert werden.

## 4.2. Zentrale Aspekte des Programms

In diesem Kapitel werden zentrale Aspekte der Arbeit beschrieben. Beschrieben ist:

- Die Datenstruktur (sogenannte SHAPE-Dateien), in der die Grenzverläufe im ursprünglichen Datensatz codiert sind (siehe Abschnitt 3.2).
- Die Datenstruktur, in der die geographischen Daten intern verwaltet werden um einen schnellen Zugriff auf die Daten innerhalb des sichtbaren Bereichs zu ermöglichen (siehe Abschnitt 4.2.1).

### 4.2.1. Aufbau der SHAPE-Daten der Verwaltungsgebietsgrenzen

Die Datei vg250\_l.shp aus dem VG250-Datensatz enthält die Verläufe der Verwaltungsgebietsgrenzen in einer binären Kodierung. Der Aufbau dieser Datei wird im Folgenden beschrieben (vgl. [S-D98]). Dabei muss beachtet werden, dass SHAPE-Dateien im Allgemeinen verschiedene Typen von Geometrischen Objekten beschreiben können (z.B. Point, Multipoint, Polyline). Da das vorliegende SHAPE-File den Typ *Polyline* enthält, wird in diesem Kapitel nur auf diesen SHAPE-Typ eingegangen und der Aufbau der anderen Typen nicht weiter ausgeführt.

Die Datei besteht aus einem Header mit fester Länge, der in Kapitel 4.2.1 beschrieben ist. Darauf folgen beliebig viele Records, die jeweils ein SHAPE-Objekt beschreiben. Ihr Aufbau ist in Kapitel 4.2.1 beschrieben. Neben der Zuordnung der Bits zu den entsprechenden Informationen ist beim Parsen insbesondere die Byte Order der Kodierung zu beachten. Diese unterscheidet sich bei den unterschiedlichen Informationen.

#### Dateiheader

Der Dateiheader hat eine feste Länge von 100 Byte. Die Belegung der einzelnen Positionen und deren Byte-Order ist in Abbildung 4.1 dargestellt. Dabei ist die Dateilänge ab Byte 24 als Länge der gesamten Datei in 16-bit-Worten angegeben, sie umfasst auch den Header selbst. Der SHAPE-Type ab Byte 32 beschreibt den SHAPE-Type der folgenden Objekte, der in diesem Fall 3, also Polyline. Hier ist zu

#### 4. Verschiedene Aspekte der Implementierung

---

<b>Position</b>	<b>Field</b>	<b>Value</b>	<b>Type</b>	<b>Byte Order</b>
Byte 0	File Code	9994	Integer	Big
Byte 4	Unused	0	Integer	Big
Byte 8	Unused	0	Integer	Big
Byte 12	Unused	0	Integer	Big
Byte 16	Unused	0	Integer	Big
Byte 20	Unused	0	Integer	Big
Byte 24	File Length	File Length	Integer	Big
Byte 28	Version	1000	Integer	Little
Byte 32	Shape Type	Shape Type	Integer	Little
Byte 36	Bounding Box	Xmin	Double	Little
Byte 44	Bounding Box	Ymin	Double	Little
Byte 52	Bounding Box	Xmax	Double	Little
Byte 60	Bounding Box	Ymax	Double	Little
Byte 68*	Bounding Box	Zmin	Double	Little
Byte 76*	Bounding Box	Zmax	Double	Little
Byte 84*	Bounding Box	Mmin	Double	Little
Byte 92*	Bounding Box	Mmax	Double	Little

\* Unused, with value 0.0, if not Measured or Z type

**Abbildung 4.1.:** Aufbau SHAPE-File-Header (Quelle: [S-D98]).

beachten, dass das die SHAPE-Datei nur Objekte eines SHAPE-Types enthält. Die einzige Ausnahme bilden Null-Shapes, diese dürfen in jeder Datei eingefügt werden.

#### Record

Nach dem Dateiheader folgt eine beliebige Menge an Objekt-Records, die alle nach folgendem festen Muster aufgebaut sind. Sie beginnen mit einem 8 Byte langen RecordHeader (siehe Abbildung 4.2), der in Byte 0 bis 3 die laufende Nummer des Objekts und von Byte 4 bis 7 die Länge des Record-Bodys enthält. Die Länge ist hier wieder in 16 bit Wörtern angegeben, allerdings ohne den Record-Header. Die Gesamtlänge des Records ergibt sich also zu  $4 + \text{length of body}$ .

<b>Position</b>	<b>Field</b>	<b>Value</b>	<b>Type</b>	<b>Byte Order</b>
Byte 0	Record Number	Record Number	Integer	Big
Byte 4	Content Length	Content Length	Integer	Big

**Abbildung 4.2.:** Aufbau SHAPE-File Record-Header (Quelle: [S-D98]).

Eine Polyline beschreibt eine Menge von Punkten, die verbunden einen Linienzug bilden. Ein SHAPE-Polyline-Objekt kann aus mehreren Teilen bestehen, wobei ein Teil immer aus einer Kette von ver-

bundenen Punkten besteht. Der Aufbau des Record-Bodys ist in Abbildung 4.3 dargestellt. Der Record-Body umfasst:

- SHAPE-Typ: Der Typ des SHAPE-Objekts
- Box: Die Kodierung der Boundingbox des Objekts (kodiert als 4 Double- Werte: Xmin, Ymin, Xmax, Ymax)
- NumParts: Die Anzahl der Teile, aus denen das Objekt besteht
- NumPoints: Die Anzahl der Punkte, die das Objekt definieren
- Parts: Ein Integer-Array der Länge NumParts, das für jeden Teil den Index des Startpunktes in Points enthält
- Points: Ein Array bestehend aus 2 Double-Variablen pro Punkt. Zwischen den einzelnen Punkten ist kein Trennzeichen oder ähnliches vorhanden.

Position	Field	Value	Type	Number	Byte Order
Byte 0	Shape Type	3	Integer	1	Little
Byte 4	Box	Box	Double	4	Little
Byte 36	NumParts	NumParts	Integer	1	Little
Byte 40	NumPoints	NumPoints	Integer	1	Little
Byte 44	Parts	Parts	Integer	NumParts	Little
Byte X	Points	Points	Point	NumPoints	Little

Note:  $X = 44 + 4 * \text{NumParts}$

Abbildung 4.3.: Aufbau SHAPE-File Polyline Record-Body (Quelle: [S-D98]).

#### 4.2.2. Beschreibung der internen Datenstruktur

Die interne Datenstruktur der geographischen Daten muss eine einfache Abfrage der sichtbaren Elemente in einem rechteckigen Abfragefenster ermöglichen. Als Datenstruktur wird daher ein rechteckiges Grid verwendet, wobei in jeder Zelle die sichtbaren Grenzsegmente gespeichert werden. Die Abfrage beschränkt sich in diesem Fall auf die Bestimmung der Grid-Zellen, die vom Sichtfenster geschnitten werden und der Abfrage der Segmente dieser und der dazwischen liegenden Zellen.

Die Größe des Grids wird auf 512x512 festgelegt, somit definiert sich die Größe der Gridzellen aus Datensatz. Das Grid wird in ein Array mit  $512 * 512$  Einträgen abgebildet, die Zählung beginnt mit der nordwestlichsten Zelle des Grids und wird dann zweilenweise fortgesetzt. Die oberste Reihe der Zellen bekommt also die Indizes 0 bis 255, die zweite Reihe die Indizes 256-511 usw.

Jede Gridzelle verwaltet eine Menge von geographischen Namen, die sie überdeckt, und eine Menge von Grenzsegmenten, die vollständig oder teilweise in ihr liegen. Um die redundante Speicherung von Daten zu verhindern wird für jedes Objekt lediglich ein Index gespeichert, über den das Objekt zu finden ist.

### 4.3. Beschreibung der genutzten Datenbanken

Nach dem Start des Programms werden alle Daten in einer internen Datenstruktur verwaltet. Um die Daten jedoch nicht bei jedem Programmstart erneut aus den ursprünglichen Datenquellen einzulesen und um einen besseren Überblick über die verwendeten Daten zu erhalten, werden die Daten extern in einer sqlite-Datenbank verwaltet. Die hier vorhandenen Datentabellen werden im folgenden Kapitel beschrieben.

Die Datenbank besteht aus 6 Tabellen (eine für geographische Namen, 5 für Verwaltungsgebiete). Die Tabellen stellen alle geographischen Daten bereit, die für die Ausführung des Programms benötigt werden.

Die Daten werden im Rahmen der Open-Data-Initiative der Bundesregierung durch das Geodatenzentrum bereitgestellt und können zum Teil direkt (bei Datensätzen, die im CSV-Format vorliegen) oder indirekt durch Funktionen, die im Programm implementiert sind, in die Datenbank importiert werden.

Im Folgenden wird zuerst die Tabelle des gn250-Datensatzes beschrieben (siehe Kapitel 4.3.1), darauf folgend in Kapitel 4.3.2 die Tabellen des VG250-Datensatzes.

#### 4.3.1. Tabellen für GN250

Die geographischen Namen wurden direkt in die Datenbank übernommen (Tabelle gn250), lediglich die Werte der Spalte Hierarchylevel wurden im Nachhinein berechnet (siehe Kapitel 3.3.2). Die Beschreibung der Datenbanktabelle gleicht daher der Beschreibung in Kapitel 2.1 zuzüglich der Spalte Hierarchylevel.

Für die Visualisierung wird lediglich auf die Attribute *OBA*, *NAME* und die Koordinatenfelder *UTMRE* und *UTMHO* zugegriffen.

#### 4.3.2. Tabellen für VG250

Zur Vollständigkeit wurden alle Tabellen des VG250-Datensatzes in die sqlite-Datenbank übernommen (Tabellen vg250\_f, vg250\_isn, vg250\_l, vg250\_l\_shape und vg250\_nam). Alle Tabellen, außer vg250\_l\_shape, wurden direkt aus den CSV-formatierten Originaldateien übernommen. In der Tabelle vg250\_l wurden die Spaltennamen verkürzt.

Tatsächlich verwendet werden für die Visualisierung lediglich die Grenzsegmente und deren Attribute aus der Tabelle vg250\_l. Der Verlauf jedes Grenzsegments ist in vg250\_l\_shape wie folgt gespeichert:

- *id*: Laufende Nummer des Objekts, stimmt mit den Nummern in vg\_250\_l überein.
- *numPoints*: Die Anzahl der Punkte, die das Segment definieren.
- *points*: Binäre Kodierung der Punkte des Segments als Folge von X- und Y- Koordinate der Punkte in Little\_Endian Ordnung.

#### 4.3. Beschreibung der genutzten Datenbanken

---

- boxXMin: Koordinate der Boundingbox
- boxYMin: Koordinate der Boundingbox
- boxXMax: Koordinate der Boundingbox
- boxYMax: Koordinate der Boundingbox

Zu jedem SHAPE-Objekt werden in der Tabelle vg250\_l die entsprechenden Eigenschaften bereitgestellt (siehe Kapitel 2.2.2). Für den Ablauf des Programms ist hier lediglich der *USE*-Eintrag notwendig.



## **5. Zusammenfassung und Ausblick**

Im Rahmen der vorliegenden Arbeit wurde eine Visualisierung von geographischen Basisdaten, insbesondere Ortsnamen und Grenzen von Verwaltungsgebieten, entwickelt.

Als Datenbasis dienten die Datensätze GN250 (Geographische Namen) und VG250-EW (Daten der Verwaltungsgebiete) des Bundesamtes für Kartographie und Geodäsie. Die Daten des GN250-Datensatzes wurden in der Arbeit nach dem Objekttyp und bei Ortsnamen zusätzlich nach der Einwohnerzahl in verschiedene Hierarchiestufen (von Staats-, über Landes- bis hin zur Gemeindeebene) unterteilt. Für die Grenzverläufe der Verwaltungsgrenzen wurde der Douglas-Peuker-Algorithmus implementiert, um den Detailgrad der einzelnen Grenzsegmente, vor allem bei hohen Zoomstufen verringern zu können und so die Reaktionszeiten der Visualisierung zu verbessern.

Für die Visualisierung der Daten wurde eine gridartige Datenstruktur entwickelt die eine schnelle Abfrage der Elemente in einem sichtbaren Bereich ermöglicht. Für die Visualisierung wurden verschiedene Interaktionsmöglichkeiten wie Zoom, Verschieben und Hinzufügen und Abschalten der Daten von verschiedenen Detailstufen implementiert.

### **Ausblick**

Das entwickelte Programm verwaltet die geographischen Basisdaten in einer sqlite-Datenbank, die alle Daten der zugrundeliegenden Datensätze beinhaltet. Um die Lösung auch für mobile Endgeräte attraktiv zu machen, wäre eine Reduktion der Daten auf die tatsächlich benötigten Informationen wünschenswert, um die Datenmenge zu verringern. Die Nutzung eines anderen Datenformats könnte die Datenmenge der Basisdaten weiter verringern und zugleich die Abhängigkeit des Programmcodes von externen Bibliotheken (sqlite-Bibliothek) reduzieren, wodurch der Platzverbrauch der Anwendung weiter minimiert würde.

Neben dem Platzverbrauch wäre eine weitere Optimierung der Anwendung denkbar, um die Geschwindigkeit der Anwendung auf mobilen Geräten zu beschleunigen. Ansätze für eine Optimierung könnten eine weitere Parallelisierung des Codes zur Behandlung der Interaktionsmöglichkeiten sowie eine Anpassung der Struktur der aktuell angezeigten Daten, um bei der Interaktion nur die tatsächlich notwendigen Änderungen in der Visualisierung durchzuführen (also das Hinzufügen neuer Bereiche und Entfernen alter, nicht sichtbarer Bereiche), liefern.

Eine Erweiterung der Arbeit auf zusätzliche Länder wäre interessant, wenn zu diesen Ländern entsprechende Datensätze vorhanden sind. Dabei sind verschiedene Probleme zu erwarten, wie ein

## 5. Zusammenfassung und Ausblick

---

größerer Ressourcenverbrauch wegen der zusätzlichen Datenmengen und das Problem der UTM-Koordinaten, da die verwalteten Daten dann vermutlich nicht komplett im Sektor 32 des UTM-Koordinatensystems verwaltet werden können.

Verschiedene Charakteristiken der Datensätze wurden erst durch die entwickelte Visualisierung sichtbar, beispielsweise die Belegung des EWZ\_GER-Attributs und die Besonderheit bei der Verteilung der Regierungsbezirke. Das führt dazu, dass die Hierarchisierung und die Auswahl der anzuzeigenden Namen überarbeitet werden sollte, da sonst beispielsweise bei der Filterauswahl Regierungsbezirksebene ausschließlich Namen der Regierungsbezirke in Bayern, Baden-Württemberg, Hessen, Nordrhein-Westfalen und Sachsen angezeigt werden. Diese Besonderheiten zu behandeln wäre eine weitere Aufgabestellung für weitere Forschung.

# **A. Anhang**

## **A.1. ATKIS-Objekte im Datensatz GN250**

### **A.1.1. Bauwerke**

AX\_Bahnstrecke: Eisenbahn, Güterzugbahn, Magnetschwebebahn, Museumsbahn, S-Bahn, Standseilbahn, Zahnradbahn

AX\_Bahnverkehrsanlage: Bahnhof, Haltepunkt

AX\_BauwerkImGewaesserbereich: Siel, Sperrwerk, Staudamm, Staumauer, Wehr

AX\_BauwerkImVerkehrsbereich: Brücke, Tunnel/ Unterführung

AX\_BauwerkOderAnlageFuerSportFreizeitUndErholung: Sprungschanze (Anlauf), Stadion

AX\_Flugverkehr: Internationaler Flughafen, Regionalflughafen

AX\_Flugverkehrsanlage: Hubschrauberlandeplatz, Sonderlandeplatz, Verkehrslandeplatz

AX\_Gebaeude: Burg/ Festung, Hütte (mit Übernachtungsmöglichkeit), Kirche, Krankenhaus, Museum, Parlament, Schloss

AX\_Grenzuebergang:

AX\_IndustrieUndGewerbeflaeche: Deponie (oberirdisch), Kraftwerk

AX\_Kanal: Binnenwasserstraße

AX\_Schleuse: Kammerschleuse, Schiffshebewerk

AX\_SchiffahrtslinieFaehrverkehr: Autofährverkehr, Personenfähverkehr

AX\_SeilbahnSchwebebahn: Kabinenbahn/ Umlaufseilbahn, Luftseilbahn/ Großkabinenbahn, Materialseilbahn, Sessellift, Ski-/ Schlepplift

AX\_SonstigesBauwerkOderSonstigeEinrichtung: Gedenkstätte/ Denkmal/ Denkstein/ Standbild

AX\_SportFreizeitUndErholungsflaeche:

AX\_Strasse: Bundesautobahn, Bundesstraße, Kreisstraße, Landesstraße/ Staatsstraße, sonstiges

AX\_Strassenverkehrsanlage: Autobahnknoten, Raststätte

AX\_TagebauGrubeSteinbruch: Erden/ Lockergestein, Industriemineral/ Salze, Steine/ Gestein/ Festgestein, Torf, Treib- und Brennstoffe

## A. Anhang

---

AX\_Turm:

AX\_Wasserlauf: Seewasserstraße

AX\_WegPfadSteig:

### **A.1.2. Natur**

AX\_DammWallDeich:

AX\_Fliessgewaesser:

AX\_Gewaessermerkmal:

AX\_Heide:

AX\_Hoehleneingang:

AX\_Insel:

AX\_Landschaft: (Tief-) Ebene/ Flachland, Becken/ Senke, Berg/ Berge, Gebirge/ Bergland/ Hügelland, Inselgruppe, Küstenlandschaft, Moorlandschaft, Plateau/ Hochfläche, Seenlandschaft, Siedlungs-/ Wirtschaftslandschaft, Tal/ Niederung, Wald- Heideland

AX\_Meer:

AX\_Moor:

AX\_Sumpf:

AX\_Wald:

### **A.1.3. Gebiete**

AX\_NaturUmweltOderBodenschutzrecht: Naturpark

AX\_Ortslage:

AX\_SchutzgebietNachNaturUmweltOderBodenschutzrecht: Biosphärenreservat, Nationalpark

AX\_Siedlungsflaeche:

AX\_SonstigesRecht: Truppenübungsplatz/ Standortübungsplatz

Key: AX\_StehendesGewaesser: Landesgewässer mit Verkehrsordnung

Key: AX\_UnlandVegetationsloseFlaeche: Eis/ Firn, Fels, Sand

Key: AX\_Verwaltungsgemeinschaft\_ATKIS:

#### **A.1.4. Verwaltungsgebiete**

AX\_Bundesland:

AX\_Gemeinde:

AX\_KreisRegion:

AX\_Nationalstaat:

AX\_Regierungsbezirk:



# Literaturverzeichnis

- [DP73] D. Douglas, T. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10/2:112–122, 1973. (Zitiert auf Seite 17)
- [GN-13] Geographische Namen 1:250 000: GN250, 2013. URL [http://sg.geodatenzentrum.de/web\\_download/gn/gn250/gn250.pdf](http://sg.geodatenzentrum.de/web_download/gn/gn250/gn250.pdf). (Zitiert auf den Seiten 9 und 10)
- [S-D98] ESRI Shapefile Technical Description, 1998. URL <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>. (Zitiert auf den Seiten 6, 25, 26 und 27)
- [UTM09] UTM-Abbildung und UTM-Koordinaten, 2009. URL <http://vermessung.bayern.de/file/pdf/1910/UTM%20Abbildung%20und%20Koordinaten.pdf>. (Zitiert auf den Seiten 13 und 14)
- [VG-13] Verwaltungsgebiete 1:250 000: VG250 und VG250-EW, 2013. URL [http://sg.geodatenzentrum.de/web\\_download/vg/vg250-ew\\_3112/vg250-ew\\_3112.pdf](http://sg.geodatenzentrum.de/web_download/vg/vg250-ew_3112/vg250-ew_3112.pdf). (Zitiert auf den Seiten 9, 11, 13 und 18)
- [Wei13] E. W. Weisstein. Point-Line Distance–2-Dimensional, 2013. URL <http://mathworld.wolfram.com/Point-LineDistance2-Dimensional.html>. (Zitiert auf Seite 17)
- [WP-13] UTM-Koordinatensystem, 2013. URL <https://de.wikipedia.org/wiki/UTM-Koordinatensystem>. (Zitiert auf den Seiten 13 und 14)

Alle URLs wurden zuletzt am 6.01.2014 geprüft.



### **Erklärung**

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Exemplaren überein.

---

Ort, Datum, Unterschrift