

## **Geokodierung mit Webkartendiensten – Möglichkeiten, Unterschiede und Grenzen**

Bennet SCHULTE, Frank LIPPMANN und Jürgen SCHWEIKART

### **1 Einführung**

Bedingt durch die sich stetig erweiternden Möglichkeiten von Geoinformationssystemen (GIS) besteht ein wachsender Bedarf an raumbezogenen Analysen. Dazu werden Koordinaten benötigt. In den zu analysierenden Datensätzen finden sich die Standortdaten jedoch in den meisten Fällen als Postanschrift. Der Vorgang der Lageverortung von Adressen ist einer der grundlegenden Vorgänge, Daten vorzubereiten. Es gilt daher Möglichkeiten aufzuzeigen, diesen Vorgang auch für eine große Anzahl von Adressen automatisiert durchzuführen, ohne von kostspieligen Daten oder Dienstleistern abzuhängen. Eine alternative Möglichkeit könnten kostenfrei verfügbare Geodaten von Webkartendiensten wie *Google Maps*, *Microsoft Bing Maps* usw. sein.

### **2 Georeferenzierung von Anschriften**

Georeferenzierung ist nach BILL (1999, 141) der Prozess, bei dem Objekten raumbezogene Lageangaben zugewiesen werden. Postanschriften in einen direkten Raumbezug zu überführen, kann daher als ein Teilbereich des Georeferenzierens betrachtet werden und wird als Adresskodierung bezeichnet.

Hierfür werden drei Komponenten benötigt: Postanschriften, ein „Geocoder“ und georeferenzierte Basisdaten. Der Geocoder ist eine Softwarekomponente bestehend aus miteinander verzahnten Operationen, Algorithmen und Datenquellen (GOLDBERG et al. 2008, 1 f.), die zwischen eingegebener Adresse und georeferenzierten Daten vermittelt.

Die Daten, die beim Prozess des Geokodierens angewendet werden, sind meist vektorbasiert (GOLDBERG et al. 2008, 2). Onlinedienstleister visualisieren Anschriften über georeferenzierte Polyliniennetzwerke. Diese bestehen aus topologisch verknüpften Kanten und Knoten, die die Straßenmittellinie bzw. die Kreuzungen repräsentieren. In den Kanten sind Straßennamen, Hausnummern sowie deren Zählweise attribuiert. Die Position der Hausnummer wird auf der Kante interpoliert. Die Genauigkeit ist abhängig von der Länge der Kanten und der realen Breite der Grundstücke bzw. der Lage der Gebäude (Abb. 1).

Eine nach GOLDBERG et al. (2008, 5) präzisere Methode ist die Verortung einer Anschrift durch ein Punkt- oder Polygonobjekt. Dies wird aus Kostengründen (TAPPERT 2007, 68) in der Wirtschaft nur selten und bei Onlinedienstleistern nur für herausragende Adressen, z. B. von Sehenswürdigkeiten gemacht. *Stadtplandienst.de* bieten mit amtlichen Daten diese Funktion auch für gewöhnliche postalische Anschriften an (EURO-CITIES 2010).

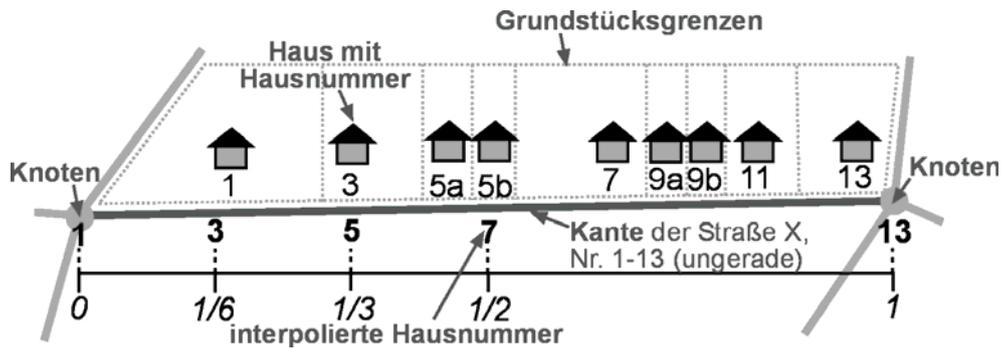


Abb. 1: Bestimmung der Lage durch Interpolation (modifiziert nach TABBERT 2007, 69)

### 3 Adresskodierung im Internet

Geodaten sind bei raumbezogenen Analysen ein nicht zu vernachlässigender Kostenfaktor. Die Datenanbieter erlauben es daher nur unter Auflagen, ihre Geodaten zu nutzen (siehe Tabelle 1). *Google* z. B. erlaubt die Nutzung nur, wenn das Ergebnis in einer *Google Map* visualisiert wird (GOOGLE 2010a). *Microsoft* umschreibt dies in seinen Nutzungsbedingungen dergestalt, dass lediglich die Visualisierung zulässiges Ziel sein darf (MICROSOFT 2010).

Tabelle 1: Zugangsbedingungen (Quelle: GOOGLE 2010b/c, MICROSOFT 2010, YAHOO! 2010)

| Anbieter                           | Geodaten   | Limitierung   | Zugriff          |
|------------------------------------|------------|---|------------------|
| AOL MapQuest                       | Navteq     | unbekannt   | URL <sup>1</sup> |
| Google HTTP <sup>2</sup> -Geocoder | Tele Atlas | je IP <sup>3</sup> : 15.000 Anfragen je 24 Stunden, eine Anfrage je 1,73 Sekunden | URL              |
| Google Maps-API <sup>4</sup>       | Tele Atlas | je API: 50.000 Anfragen je 24 Stunden, eine Anfrage je 1,73 Sekunden              | API              |
| Microsoft Bing Maps                | Navteq     | je API: 125.000 Sessions oder 500.000 Transaktionen je 12 Monate                  | API              |
| OSM (Nomi.)                        | OSM        | unbekannt   | URL              |
| Yahoo! Maps                        | Navteq     | je API: 5.000 Anfragen je 24 Stunden  | API              |

<sup>1</sup> Uniform Resource Locator: Identifiziert und lokalisiert eine Quelle über das verwendete Netzwerkprotokoll (KOZIEROK 2005, 1327).

<sup>2</sup> Hypertext Transfer Protocol: Ein Protokoll zur Übertragung von Daten über ein Netzwerk (KOZIEROK 2005, 133).

<sup>3</sup> Internet Protocol: Netzwerkprotokoll in Computernetzwerken, das die Basis für Anwendungen im Internet bildet (KOZIEROK 2005, 92).

<sup>4</sup> Application Programming Interface: Englisch für Schnittstelle zur Programmierung von Anwendungen (KIRCHNER & BENS 2010, 124).

Die wichtigsten kommerziellen Anbieter von Straßendaten sind die *TomTom A.G.* mit *Tele Atlas* und die *Nokia A.G.* mit *Navteq*. Nach ESRI (2009) liegt für beide eine 95%ige Trefferquote für Adresskodierungen und eine Lagegenauigkeit im Bereich von 10 m in Europa vor. Das *OpenStreetMap* Projekt (OSM) hingegen basiert auf von Nutzern erzeugten Geodaten („*Usergenerated Content*“) und stellt diese Daten frei zur Verfügung. Es ist anzunehmen, dass die vollständigsten Daten für die Bundesrepublik Deutschland im kostenpflichtigen Amtlichen Topographisch-Kartographischen Informationssystem (ATKIS) zu finden sind.

## 4 Qualitätsunterschiede bei der Adresskodierung

### 4.1 Bedeutung der Präzision

Der Anspruch an die Lagegenauigkeit der produzierten Koordinaten hängt vom Arbeitsmaßstab und der Thematik ab. Für kleinmaßstäbige Analysen ist eine unscharfe Adresskodierung meist nicht relevant. Bei Analysen in großen Maßstäben ist eine möglichst geringe Abweichung zur realen Lage bei vielen Themen wünschenswert, manchmal notwendig.

### 4.2 Praxistests

Um eine Aussage über die Genauigkeit von Adresskodierungen treffen zu können, wurden in zwei Testgebieten in Berlin ein Testdatensatz bestehend aus insgesamt 624 Anschriften sowohl mit *Microsoft Mappoint 2009 (Navteq)* als auch dem *Google HTTP-Geocoder (Tele Atlas)* geokodiert. Als Referenzdaten dienen die Koordinaten, die aus der deutschen Grundkarte 1:5000 entnommen sind. Die Abweichungen dazu wurden statistisch ausgewertet. Aus technischen Gründen wurde statt auf *Microsofts* Webkartendienst auf ein Anwendungsprogramm zurückgegriffen. Durch Unterschiede im Geocoder ist eine 100%ige Vergleichbarkeit mit dem Onlineangebot nicht gegeben.

Bei *Microsoft* differieren die Ergebnisse von der tatsächlichen Lage der Hauseingänge um das arithmetische Mittel von 43,71 m mit einer Standardabweichung von 109,95 m, bei *Google* hingegen um 38,22 m mit einer Standardabweichung von 81,22 m. Ausreißer mit einer Differenz von mehr als 100 m treten dabei in beiden Datensätzen bei unterschiedlichen Adressen auf. Von diesen Ausreißern gab es 20 bei *Navteq*-Daten mit bis zu 1125,8 m und 32 bei *Tele Atlas*-Daten mit bis zu 1746,3 m (siehe Abb. 2).

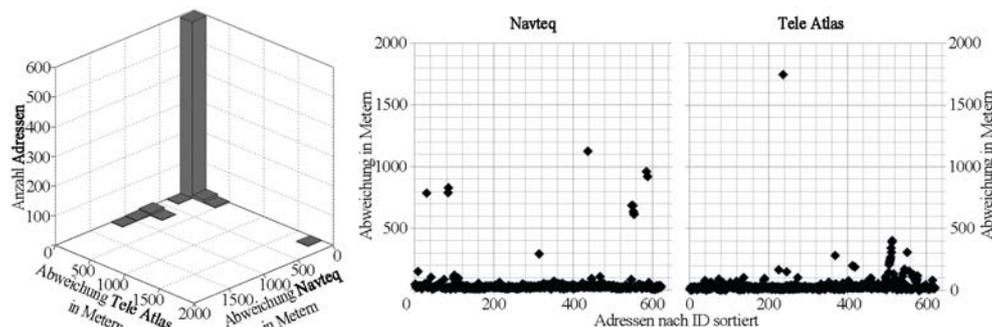


Abb. 2: Statistische Auswertung der Genauigkeit der Adresskodierung

Wie nahe die Koordinaten an der realen Lage liegen, ist abhängig von den Geodaten. Ob die Adresse den möglichst kleinräumigsten Geodaten zugeordnet werden kann, hängt wiederum von den Algorithmen des Geocoders ab. *Google* z. B. geokodiert bei Verwendung der Abkürzung „Str.“ in einer Anschrift auf einen gleichnamigen Platz im Postleitzahlengebiet. Diese Schwäche des Geocoders macht es notwendig, die Anschriften vollständig einzugeben und dabei Regeln sehr genau zu beachten. Anschriften, die nicht auf dem detailliertesten Adresslevel, z. B. hausnummerngenau, verortet werden können, werden auf dem nächsthöheren eindeutig zuordenbaren Level, z. B. auf dem Mittelpunkt des Straßenabschnittes, verortet.

### 4.3 Inadäquate Adresskodierung

Jede der drei Komponenten eines Adresskodierungsvorgangs, die Adresseingabe, der Geocoder und die Geodaten, bergen Gefahren in sich, die zu einem inadäquaten Ergebnis führen können.

Nach GRASEKAMP (2007, 533) sind Ursachen bei der Angabe der Anschriften maßgeblich Tippfehler, phonetische Verwechslung und abweichende Schreibweisen. Ebenso schwerwiegend kann es sein, wenn nur einzelne Teile einer Anschrift unvollständig oder fehlerhaft sind. Tests zeigen, dass die Interpretation und der Umgang mit falschen oder fehlenden Angaben von den Algorithmen der Geocoder abhängig sind. Nicht zuletzt können auch die Referenzgeodaten fehlerhaft oder unvollständig sein.

*Google* gibt die Genauigkeit für die Ergebnisse der Adresskodierung vorbildlich an. Über die Variable *GeoAddressAccuracy* wird mittels einer hierarchischen Skala von 0 bis 9 beziffert, in welchem Bezug zur Anschrift sich die ausgegebenen Koordinaten befinden. Z. B. bedeutet „5“, dass sich die Koordinaten im Zentrum des Postleitzahlengebiets, „1“ hingegen in der Mitte des Landes befindet (GOOGLE 2010e).

Es existiert jedoch auch die Möglichkeit, dass die Variable *GeoAddressAccuracy* eine genaue Geokodierung bestätigt, obwohl dies nicht der Fall ist. Das tritt ein, wenn der Geocoder eine Anschrift falsch interpretiert und dazu eine passende Übereinstimmung in den Geodaten findet.

## 5 Geokodieren mit Online-Kartendiensten

Mit dem Problem konfrontiert, Adresskodierungen für raumbezogene Forschungen durchzuführen, wurde es durch steigende Komplexität und Masse der Daten nötig, nach vollautomatischen und kostenfreien Werkzeugen zu suchen. Ziel der Machbarkeitsstudie „*AdressZilla*“ ([www.adresszilla.org](http://www.adresszilla.org)) ist es, eine auf den deutschsprachigen Raum angepasste kostenlose Applikation zu entwickeln, die für kleinräumige Analysen möglichst genaue Adresskodierungen erzeugt und verschiedene Onlinedienstleister nutzt.

Als Pilot wird ein pragmatischer Weg beschritten, um das Geokodierungsangebot zu nutzen. In einer Tabelle abgelegte Daten wurden als CSV-Dateien an *Googles HTTP-Geocoder* gesendet. Um eine sequenzielle Adresskodierung zu ermöglichen, wurde *AdressZilla* als ferngesteuerter Browser angelegt. Es wird über die Liste der Postanschriften ite-

riert und jede Adresse wird bei *Google* per GET (HTTP-Methode) abgefragt. Das Ergebnis wird ausgelesen und wieder in einer CSV-Datei abgespeichert.

Die Adressen werden vor der Anfrage bei *Google* auf die Bedingungen einer GET-Abfrage und des Geocoders angepasst. Alle Sonderzeichen werden normalisiert und Abkürzungen ausgeschrieben. Da die Anzahl der Abfragen zeitabhängig limitiert ist (vgl. Tab. 1), wird die Sequenz angepasst.

Das Programm kann in seiner aktuellen Version auch *Yahoo!* und *Bing Maps* per GET ansprechen und verwerten. *AdressZilla* kann jedoch in seiner jetzigen Form nicht auf [www.adresszilla.org](http://www.adresszilla.org) allgemein zugänglich gemacht werden, da es den Nutzungsbedingungen widerspricht. Insbesondere die fehlende Visualisierung in den Karten des jeweiligen Anbieters der Ergebnisse ist problematisch.

## 6 Fazit

Die automatisierte Adresskodierung größerer Datensätze mit Webkartendiensten ist möglich. Besonders bei *Google*, *Microsoft* und *Yahoo!* bestehen über hausinterne APIs einfache Möglichkeiten, auf Geokodierungsfunktionen zuzugreifen. Mit ihnen muss zwingend gearbeitet werden, will man sich an die eng gesetzten Grenzen der Nutzungsbedingungen halten.

Es wurde gezeigt, dass die Genauigkeit der erzielten Ergebnisse unterschiedlich ist. Trotz einiger Mängel erwies sich *Googles* Geocoder in Kombination mit *Tele Atlas* Daten gegenüber *Microsoft* in Kombination mit *Navteq* Daten als leicht überlegen. Zwar wies *Google* die meisten Ausreißer und dazu die größte Abweichung im Einzelfall auf, erzielte jedoch im Durchschnitt eine kleinere Lageabweichung als *Microsoft*. Es fiel ebenfalls auf, dass die Dienste nicht bei den gleichen Adressen Fehler erzeugten, sondern jeder – je nach Kombination aus Geocoder und Geodaten – eigene Stärken und Schwächen aufwies.

Werden mehrere Dienste kombiniert, wäre dies ein Weg, das Ergebnis zu optimieren. Auffallend ist, dass identische Datenabfragen beim gleichen Anbieter zu unterschiedlichen Ergebnissen führen können. So gibt *Google Maps* andere Positionen aus als die *Google Maps*-API, da es sich um unterschiedliche Geocoder und modifizierte Geodaten von *Tele Atlas* handelt (GOOGLE 2010d). Gleiches gilt für verschiedene *Microsoft* Produkte, die auf *Navteq* Daten zugreifen. Die Daten von *OpenStreetMaps* sind noch nicht ausreichend auf Adresskodierung vorbereitet. Sie könnten jedoch in wenigen Jahren, unter aktiver, gewissenhafter und flächendeckender Implementierung von Adressen durch die Nutzer, eine echte und – so bleibt zu hoffen – präzisere Alternative werden, als die gewerblichen Anbieter es heute sind.

Es könnte mit *AdressZilla* ein Programm geschaffen werden, das Adresskodierungen kostenfrei unter Nutzung verschiedener Anbieter durchführen kann. Würden die Nutzungsbedingungen es zulassen, könnte *AdressZilla* von einer breiten Öffentlichkeit genutzt werden. *AdressZilla* könnte zu einer API-basierten Anwendung entwickelt werden. Der Übergang könnte es ermöglichen, den Nutzungsbedingungen der Anbieter zu entsprechen, da die APIs dem Programm die Möglichkeit geben, die Ergebnisse kartographisch zu visualisieren. Dennoch bestünde weiterer Entwicklungsbedarf. Am Ende könnte ein benutzerfreundliches und präzises Werkzeug zur kostenfreien Adresskodierung stehen.

## Literatur

- BILL, R. (1999): Grundlagen der Geo-Informationssysteme. Band 1. 4. Aufl. Wichmann, Heidelberg.
- ESRI (2009): ESRI StreetMap Premium – Comparison Matrix. – <http://www.esri.com/data/streetmap/comparison-matrix.pdf> (20.04.2010).
- EURO-CITIES AG (2010): Impressum – stadtplandienst.de. – <http://www.stadtplandienst.de> (10.04.2010).
- GOLDBERG, D. W., SWIFT, J. N. & WILSON, J. P. (2008): Geocoding Best Practices: Reference Data, Input Data and Feature Matching. Technical Report No 8. – University of Southern California GIS Research Laboratory, Los Angeles. – <http://gislab.usc.edu/i/publications/gislabtr8.pdf> (10.04.2010).
- GOOGLE (2010a): The Google Geocoding Web Service – What is Geocoding? – <http://code.google.com/intl/de-DE/apis/maps/documentation/geocoding/index.html#Geocoding> (12.04.2010).
- GOOGLE (2010b): Google Mobile – Google Maps Navigation. – <http://www.google.com/mobile/navigation/> (17.04.2010).
- GOOGLE (2010c): Anmelden für Google Maps-API. – <http://www.google.com/apis/maps/signup.html> (10.04.2010).
- GOOGLE (2010d): Häufig gestellte Fragen – Warum geben die Geokodierer der Google Maps-APIs andere Positionen an als Google Maps? – [http://code.google.com/intl/de-DE/apis/maps/faq.html#geocoder\\_differences](http://code.google.com/intl/de-DE/apis/maps/faq.html#geocoder_differences) (16.04.2010).
- GOOGLE (2010e): Google Maps-API – Referenz. – <http://code.google.com/intl/de-DE/apis/maps/documentation/reference.html> (12.04.2010).
- GRASEKAMP, D., FRECH, I. & FRECH, K. (2007): Geomarketing mit Microsoft MapPoint. Microsoft Press Deutschland, Unterschleißheim.
- KIRCHNER K. & BENS, P. (2010): Google Maps. Webkarten einsetzen und erweitern. dpunkt.verlag, Heidelberg.
- KOZIEROK, C. M. (2005): The TCP/IP guide: a comprehensive, illustrated Internet protocols reference. No Starch Press, San Francisco.
- MICROSOFT (2010): Microsoft Bing Maps Platform Api's Terms Of Use (03.2010). – <http://www.microsoft.com/maps/product/terms.html> (19.04.2010).
- TAPPERT, W. (2007): Geomarketing in der Praxis. Bernhard Harzer, Karlsruhe.
- YAHOO! (2010): Yahoo! Maps Web Services – Geocoding API. – <http://developer.yahoo.com/maps/rest/V1/geocode.html> (02.04.2010).