

Geokodierung 6- Wochenbefragung Thurgau 2003

Working Paper

Author(s):

Machguth, Horst; Löchl, Michael

Publication date:

2004-12

Permanent link:

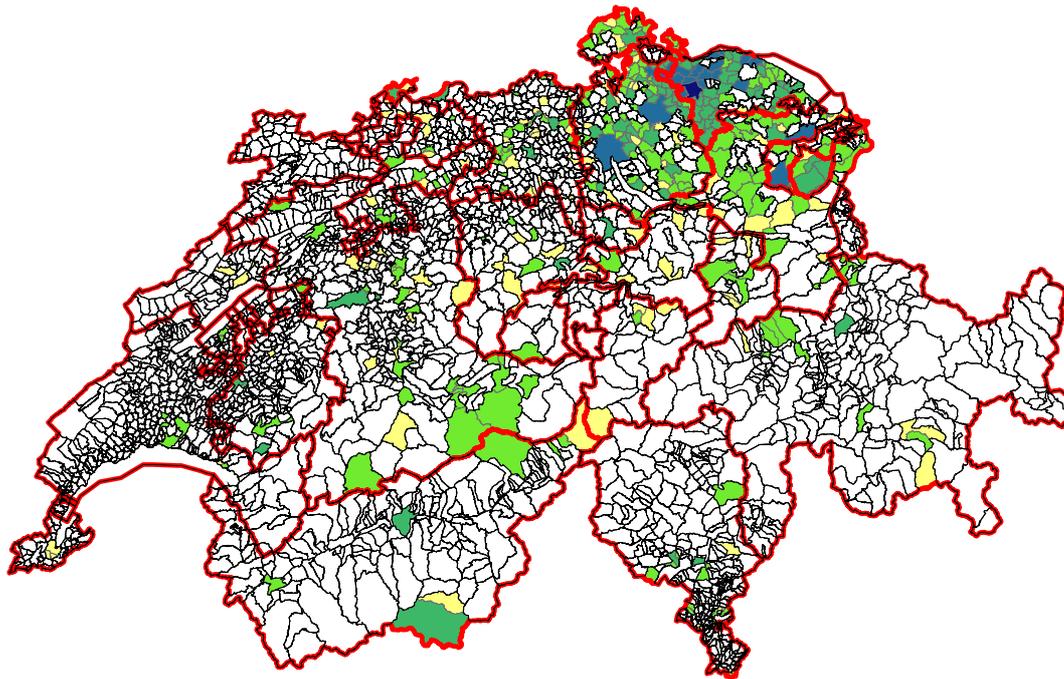
<https://doi.org/10.3929/ethz-b-000023594>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Originally published in:

Arbeitsberichte Verkehrs- und Raumplanung 219



Geokodierung 6-Wochenbefragung Thurgau 2003

Horst Machguth

Michael Löchl

Arbeitsbericht Verkehrs- und Raumplanung 219

December 2004

Arbeitsbericht Verkehrs- und Raumplanung 219

Geokodierung 6-Wochenbefragung Thurgau 2003

Michael Löchl
IVT / ETH Zürich
ETH Hönggerberg
CH-8093 Zürich

Telefon: +41-1-633 62 58
Telefax: +41-1-633-10-57
loechl@ivt.baug.ethz.ch

Dezember 2004

Kurzfassung

In der vorliegenden Arbeit wird das Vorgehen und die Erfahrungen bei der Geokodierung einer Verkehrserhebung im Kanton Thurgau beschrieben.

Im Herbst und Winter 2003 wurden 99 Haushalte über 6 Wochen zu ihrem Verkehrsverhalten befragt. Jedes Haushaltsmitglied hat dazu alle unternommenen Wege festgehalten, inkl. der Wegeziele. Mit Hilfe von Datenbank-Funktionalitäten wurden dabei den in der Erhebung angegebenen Wegezieladressen in der Schweiz und im Ausland Geodaten zugewiesen und die Koordinaten übertragen. Je nach Datenqualität konnte das *matching* automatisch oder durch verschiedene manuelle Schritte erfolgen. Entsprechend wurde ein *Rating* der Qualität der Geokodierung erstellt.

Rund 99% der Ziele konnten aufgrund der vergleichsweise guten Qualität der Eingangsdaten geokodiert werden.

Schlagworte

Geokodierung; Verkehrsverhalten, Stabilität, 6-Wochen-Tagebuch, Thurgau

Zitierungsvorschlag

Machguth H. und M. Löchl (2004) Geokodierung 6-Wochenbefragung Thurgau 2003, *Arbeitsbericht Verkehrs- und Raumplanung*, **219**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich

Working paper Transport and Spatial Planning 219

Geocoding 6-weeks travel diary Thurgau 2003

Michael Löchl
IVT / ETH Zürich
ETH Hönggerberg
CH-8093 Zürich

Telefon: +41-1-633 62 58
Telefax: +41-1-633-10-57
loechl@ivt.baug.ethz.ch

Dezember 2004

Abstract

This paper describes the procedure and the experiences from the geocoding task of a transport survey in canton Thurgau.

In autumn and fall 2003, 99 households were surveyed concerning their travel behaviour for 6 weeks. Every household member recorded every single trip, incl. the trip destinations. With the use of database functionality, the stated trip destinations in Switzerland and abroad were assigned to geodata and referring coordinates. Depending on data quality, the matching could have been performed automatically or through manual steps. Accordingly, a rating of the quality of geocoding has been compiled.

Around 99% of the trip destinations were geocoded successfully due to the good quality of input data.

Keywords

Geocoding, travel behaviour, stability, 6-week travel diary, Thurgau

Preferred citation style

Machguth H. und M. Löchl (2004) Geokodierung 6-Wochenbefragung Thurgau 2003, *Arbeitsbericht Verkehrs- und Raumplanung*, **219**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich

1 Einleitung

Im Rahmen des vom Schweizer Verband der Ingenieure geförderten Projektes ‚Untersuchung der Stabilität des Verkehrsverhaltens‘ (SVI 2002/002) wurde im Herbst und Winter 2003 eine schriftliche Verkehrsbefragung von Personen im Kanton Thurgau vorgenommen. Dabei wurden die Personen aus 99 Haushalten über jeweils 6 Wochen zu ihren Verkehrswegen in Fragebögen befragt und die Wegzieladressen von ihnen festgehalten. Die knapp 37000 Wege konzentrieren sich auf die Heimat der Befragten im Thurgau und angrenzende Gebiete in der Schweiz und Deutschland, doch wurden auch Wege nach Frankreich, Österreich und Slowenien unternommen.

Die angegebenen Wegzieladressen wurden geokodiert. Darunter ist hier die Zuweisung der Landeskoordinaten der Schweiz für jeden Zielpunkt zu verstehen.

Im Kapitel 2 wird ein Überblick über die verwendeten Daten gegeben. Kapitel 3 zeigt das grundsätzliche Vorgehen der Geokodierung, Kapitel 4 die entsprechenden Resultate. Kapitel 5 geht auf die zu berücksichtigenden Restriktionen bzw. den möglichen Fehlerquellen ein, woraus in Kapitel 6 Empfehlungen für weitere Geokodierungsarbeiten abgeleitet werden.

2 Grundlagen

2.1 Umfang der zu geokodierenden Daten

Die vorliegenden Daten aus der Umfrage Thurgau gliedern sich folgendermassen:

- Befragte Personen: 231
- Anzahl Ziele (Wege): 36824
- Anzahl unterschiedliche Ziele: 4849

Die Menge der zu geokodierenden Ziele reduziert sich durch Mehrfachnennungen um 86 % von 36824 auf 4849. Im vorliegenden Bericht beziehen sich hingegen alle Angaben und Analysen immer auf die Gesamtanzahl der 36824 Wege (Wegpunkte).

2.2 Datengrundlagen Punktdaten

Die Ziele werden basierend auf vorhandenen Punktdaten bzw. –koordinaten geokodiert. Diese wurden entnommen aus Gebäudedaten, Strassendaten, Bahnhofdaten und Postleitzahl-Daten.

Gebäudepunkte

Die Gebäudedaten bestehen aus den Gebäudekoordinaten des Schweizer Gebäuderegisters des Bundesamtes für Statistik. Für jedes Gebäude liegen die Variablen PLZ, Ortsname, Strassenbezeichnung und Hausnummer vor.

Strassenpunkte

Das gelegentliche Fehlen von Hausnummern bei den Zielangaben der Wege lies in vielen Fällen nur eine Geokodierung auf der Ebene der Strasse zu. Die Koordinaten der Strassen wurde auf der Basis der Gebäudedaten erstellt (siehe Kapitel 3.2.1).

Bahnhofspunkte

Die genaue Adresse von Bahnhöfen ist meist nicht bekannt. Entsprechend häufig findet sich als Zielangabe anstelle einer Strassenanschrift nur der Eintrag „Bahnhof“ oder „Bahnhof SBB“ sowie der Ortsname bzw. die Postleitzahl. Solche Einträge können am einfachsten unter Verwendung einer Datei mit den Koordinaten der Bahnhöfe referenziert werden.

Die verwendete Datei mit den Bahnhofdaten wurde im Rahmen der Geokodierung Mikrozensus erstellt (Jermann, 2003). Dabei wurde als Grundlage der Bahnhofsdatensatz des ARE verwendet, welches wiederum im Jahr 2000 aus der Datendank fester Anlagen (DfA) der SBB abgeleitet wurde. Dieser Datensatz stellt eine konsistente Punktdatenbasis der Bahnhöfe dar. Die Angaben zu den Koordinaten sind hier teilweise Zentimetergenau und wurden für die Georeferenzierung auf ganze Meter aufgerundet. Gemäss Jermann (2002) beziehen sich die Koordinaten auf den Schnittpunkt der Gebäudeachse des Aufnahmegebäudes mit der Achse von Gleis 1 oder auf eine zentrale Unterführung.

Postleitzahl-Punkte

Die Postleitzahl-Daten bestehen aus einem schweizweiten Poststellenfile MicroPost der Firma MicroGIS (St. Sulpice, VD). Das Poststellenfile MicroPost weist eine konsistente, flächendeckende PLZ-Angabe mit genauer Poststellenlage auf. Sie unterscheidet auch Untersektionen von PLZ-Regionen (Bsp. Lausanne 25), hingegen fehlen firmen- oder behödenspezifische Postleitzahlen. (Bsp 8092 für ETH Zentrum). Das Fehlen solcher Postleitzahlen ist nicht von Bedeutung weil die Befragten firmen- oder behödenspezifische Postleitzahlen nie angegeben haben. Das Poststellenfile erforderte keine Anpassungsarbeiten (vgl. Jermann, 2003).

2.3 Hilfsmittel

Gestützt auf die Erfahrungen aus der Geokodierung Mikrozensus (Jermann, 2003) wurde die Geokodierung mit der Datenbank Microsoft Access® durchgeführt. Vorherige Versuche bei der Geokodierung mit GIS-Software – es wurden ArcInfo, ArcView, ArcGIS (alle ESRI) und MapInfo (MapInfo Corporation) beigezogen – hatte keinen Erfolg, da bei diesen Produkten nur US-typische Adressformate und US-typische Strassentypen und -abkürzungen erkannt wurden.

3 Vorgehen

Die vorgenommenen Arbeiten können in vier Schritte zusammengefasst werden:

1. Aufbereiten der zu referenzierenden Ziele und unterteilen in verschiedene Wege-Dateien
2. Sammeln und Aufbereiten der Punktdaten
3. Geokodierung der Ziele durch Zuweisung von Koordinaten aus den Punktdaten
4. Zusammenfügen der Wege-Dateien und Auswertung

3.1 Aufbereiten der Ziele

Die 36824 Ziele wurden aus technischen Gründen in zwei Dateien unterteilt: „Wegziele innerhalb der Schweiz“ und „Wegziele Ausland“. Die Einträge in den Wege-Dateien wurden auf die Punktdaten abgestimmt. Originaleinträge wurden nie verändert, sondern es wurden zusätzliche neue Spalten mit allfälligen angepassten Einträgen erstellt. Auf das Ersetzen von Umlauten konnte weitgehend verzichtet werden, weil sowohl die Wege-Dateien als auch die Gebäudedaten mit Umlauten in Access importiert werden konnten.

3.1.1 Anpassungsarbeiten

In den Rohdaten enthalten die Zieladressen in einer Spalte sowohl Strassennamen als auch Hausnummer. In Excel wurden diese beiden getrennt, auf zwei Spalten verteilt und anschliessend wieder nach Access importiert.

In einem zweiten Schritt wurden die Einträge in eine konsistente Form gebracht. So wurden beispielsweise alle Strassen abgekürzt mit „Str.“, „Rue“ immer mit „R.“ und „Avenue“ mit „Av.“ sowie „Gottlieb Kellerstr.“ umgeformt zu „Gottlieb-Keller-Str.“ etc.; siehe Tabelle 1.

Tabelle 1 Beispiele Austauschform bei unterschiedlicher Schreibweise

Ausprägung im Punktfile	Ausprägung in der Wege-Datei	Vereinheitlichung
Bahnhofstr.	Bahnhofstrasse	Bahnhofstr.
Av. des Pâquis	Avenue des Paquis	Av. des Paquis
Georg-Krämer-Strasse	Georg Kraemer Strasse	Georg-Kraemer-Strasse
Via San Gottardo	Via S. Gottardo	Via San Gottardo

Weiterhin lagen Adressen mit falscher Strassen- oder Namensbezeichnung vor. Diese Adressen mussten dann meist manuell auf eine entsprechende Austauschform gebracht werden, wie beispielhaft in der folgenden Tabelle 2 zu sehen ist.

Tabelle 2 Beispiele Austauschform bei unterschiedlicher Bezeichnung

Ausprägung im Punktfile	Ausprägung in der Wege-Datei	Austauschform
Route des Charoubiers	Rue des Charoubiers	Route des Charoubiers
Adlergasse	Adlergässli	Adlergasse
Chemin des écoles	Ch. de l'école	Chemin des écoles
Brühlstr.	Brüelstr.	Brühlstr.

3.2 Sammeln und Aufbereiten der Punktdaten

Zur Vorbereitung der Punktdaten war nur ein geringer Aufwand nötig, da die Postleitzahlen-Datei sowie die Bahnhoferdatei in benötigter Form vorlagen (betreffend der Aufbereitung dieser beiden Tabellen sei auf Jermann, 2003 verwiesen). Es mussten lediglich die Einträge in den Wegedateien auf die Form der Einträge in den Punktdaten-Dateien abgestimmt werden.

3.2.1 Erstellen einer Strassen - Datei

Bei einer ganzen Reihe von Einträgen in der Wegedatei fehlen Strassennummern. Solche Einträge können entsprechend nur auf der Ebene Strasse geokodiert werden. Eine Datei mit entsprechenden Koordinaten für die Strassen existiert nicht und musste daher auf der Basis der Gebäudedaten erstellt werden. Die Zuweisung von Koordination auf der Ebene Strasse bleibt problematisch. Die Genauigkeit der Geokodierung variiert mit der Länge der Strasse. Der mittlere Fehler hängt aber weiter von der Kurvigkeit der Strasse ab.

Zunächst wurde in Access eine Strassendatei erstellt. Jeder Strasse wurden die Koordinaten des Medians der Hausnummern zugeteilt. In einigen Fällen fehlen auch in der Gebäudedatei Hausnummern. Diese Einträge wurden ignoriert solange für dieselbe Strasse Einträge mit Hausnummern vorlagen. Sobald jedoch für eine Strasse nur Einträge ohne Hausnummern vorlagen, wurden der Strasse die Koordinaten eines von Access automatisch ausgewählten Eintrags zugewiesen. Dabei handelt es sich nicht um einen nach Hausnummer oder Koordinaten minimalen oder maximalen Eintrag. Von 826 Strassen, die für das Strassen-genaue Matching verwendet wurden, war dies bei 71 der Fall. Es handelt sich meistens um kleine Stras-

sen, durchschnittlich verfügt jede der 71 Strassen über vier Gebäude. Es wurde eine Tabelle der 71 auf beliebige Gebäude gematchten Strassen erzeugt.

Die Zuweisung einer mittleren X- und Y-Koordinate ist nicht sinnvoll, da diese mittleren Koordinaten einer Strasse in den meisten Fällen nicht mit dem Verlauf der Strasse zusammenfallen.

3.3 Geokodierung Ziele

Die Zielfdateien wurden basierend auf den vier Punktdaten-Dateien geokodiert. Dabei wurde folgendermassen vorgegangen:

Tabelle 3 Strukturierung der Geokodierung

Ziele ...

... liegen vor als

... werden geokodiert über

In der Form:

PLZ + Ortschaft + Str. + HNr.

Gebäudepunkte:

- genaue Adresse, oder wenn nicht möglich
- nächstliegende H-Nr.

PLZ + Ortschaft + Str.

Strassenpunkte: Koordinaten der Median – Hausnummer; sonst zufällige ausgewähltes Gebäude

PLZ + Ortschaft

Koordinaten des zugehörigen Postamts

PLZ + Ortschaft + „Bahnhof“

Bahnhofskordinaten

Deutsches Ausland

Häusernummern attribuiert auf dem NavTech-Strassennetz

Sonstiges Ausland

Gemeindemittelpunkte

Die grundsätzliche Methodik der Georeferenzierung in Microsoft Access kann an folgendem Beispiel schematisch erläutert werden.

Wenn folgende Bedingungen in der Punktdatensatz und in der Wegedatensatz erfüllt sind:

[Punktdatensatz].[Ortschaft] = [Wegedatensatz].[Ortschaft] AND
 [Punktdatensatz].[Strasse] = [Wegedatensatz].[Strasse] AND
 [Punktdatensatz].[Hausnummer] = [Wegedatensatz].[Hausnummer]

...dann werden die Koordinaten des Eintrages aus der Punktdatensatz in den entsprechenden Eintrag in der Wegedatensatz übertragen.

3.3.1 Matching

Das Vergleichen zweier Tabellen und das Zuweisen von Koordinaten von einer in die andere Tabelle nach oben aufgezeigtem Beispiel wird *matching* genannt.

Das *matching* erfolgt in drei Stufen. Als erste Stufe wird mit allen Einträgen ein automatisierter Abgleich durchgeführt, bei dem die Übereinstimmung von Wegedatensatz und Punktdaten über die Datenbankfunktionen geprüft wird (vgl. Tabelle 4). Wo ein erster Durchlauf keine Übereinstimmung findet, wird über mehrere automatische Zwischenschritte versucht, möglichst viele Zuordnungen zu nahe gelegenen Alternativpunkte zu finden. Zu diesem Zweck wird die Hausnummer in folgender Reihenfolge verändert: +2, -2, +1, -1, +4, -4, +3, -3. Nach jeder Veränderung wird erneut versucht noch ausstehende Einträge abzugleichen.

Tabelle 4 Ablauf der automatischen Geokodierung

Ziele ...

<i>Übereinstimmung in den Feldern</i>	<i>Priorität</i>
PLZ + Ortschaft + Str. + Hnr.	1
Ortschaft + Str. + Hnr.	2
PLZ + Str. + Hnr.	3

Unabhängig davon, ob die Georeferenzierung mit unveränderter oder mit einer veränderten Hausnummer erfolgte, wurde die in Tabelle 4 aufgezeigte Reihenfolge eingehalten. Prioritär werden Einträge gesucht, die auf allen vier Feldern übereinstimmen. Für die verbleibenden Einträge wird zuerst die PLZ-Information fallengelassen.

Danach folgt eine zweite, halbautomatische Stufe. Dabei werden die verbleibenden Ziele in der Wegedatei manuell mit der Punktedatei verglichen. Es werden dort ähnliche Einträge gesucht, die den Einträgen in der Wegedatei entsprechen könnten. Beispielsweise wird der Eintrag „Oberdorf“ durch „Oberdorfstr.“ ersetzt, falls in der entsprechenden Gemeinde keine Strasse mit der Bezeichnung „Oberdorf“ existiert, hingegen eine „Oberdorfstr.“ vorhanden ist. Die korrigierten Einträge werden dann automatisch geprüft. Weil die Zuweisung nicht so sicher ist wie beim vollautomatischen *matching* erhalten sie jedoch ein anderes *Rating* (siehe folgender Abschnitt).

Der dritte Schritt ist manuell. Die verbleibenden Ziele werden manuell über die Software „Telinfo“ von der Firma „Swisscom Directories“ Anschriften zugewiesen. Diese Software stellt ein Telefonbuch der Schweiz mit integrierter Schweizerkarte zur Verfügung. Die meisten Einträge der Telefonbuchs sind georeferenziert und können per Knopfdruck in der Karte des Programms angezeigt werden. Zwar wird nur die Lokalität und nicht deren Koordinaten angezeigt, das Programm zeigt jedoch die Koordinaten für jeden Punkt an dem sich der Mauszeiger befindet an. Die durch Eingabe von Ortschaft, Strassenname und Hausnummer ermittelten Koordinaten werden manuell in die Wegedatei übernommen. Die Genauigkeit dieses Verfahrens wurde an einigen Beispielen überprüft. Die ermittelten Koordinaten lagen in keinem Fall weiter als 30 m vom effektiven Punkt entfernt. Beispiele für dieses Vorgehen, räumlich dargestellt auf der Telinfo-Karte, finden sich im Anhang.

Einträge, die manuell nicht geokodiert werden konnten, wurden in zwei Klassen unterteilt:

- **Plausible Einträge:** In diese Klasse fallen, z.B. auch alle Einträge bei denen nur die Ortschaft aufgeführt ist. Bei diesen Einträgen ist keine weitere Überprüfung der Plausibilität möglich.
- **Fragwürdige Einträge:** Beispiele hierfür sind „Weinfelden - Hafenstr. 50d“ (Weinfelden verfügt über keine Hafenstrasse), „Winterthur - Rigi-Kulm“ (die Rigi liegt in der Innerschweiz), „Frauenfelden - Frauenfelderstr.“ (nur in seltenen Fällen benennt eine Gemeinde Strassen auf ihrem Gebiet nach sich selber, in Frauenfeld existiert keine Frauenfelderstrasse) Genauere Erläuterungen zur Bestimmung fragwürdiger Einträge finden sich in Kapitel 5.2.

Allen plausiblen Einträgen wurden abschliessend die Koordinaten der Postleitzahl-Punkte zugewiesen. Den fragwürdigen Einträgen wurden keine Koordinaten zugewiesen.

Einträge innerhalb von Deutschland wurden extern mit Hilfe des mit Hausnummern attribuierten Navtech-Strassennetzes geokodiert. Die Geokodierung wurde dabei mit einem *Rating* versehen, welches in das *Rating* gemäss Tabelle 5 übertragen wurde.

3.3.2 Bewertung der Genauigkeit (*Rating*)

Jedem georeferenzierten Eintrag wird ein *Rating* zugewiesen, welches Auskunft über die Qualität der Zuordnung gibt (siehe Tabelle 5). Beispiele zum *Rating* auf Karten räumlich dargestellt finden sich im Anhang.

Tabelle 5 Bewertung der Genauigkeit (*Rating*)

Klasse	Genauigkeit	Beschreibung
A1	Gebäude	Automatisch referenziert; Ort, Str, Hnr übereinstimmend
A2	Gebäude	Gebäude - Automatisch referenziert; Ort, Str übereinstimmend, Hnr +-2
A3	Gebäude	Gebäude - Automatisch referenziert; Ort, Str übereinstimmend, Hnr +-4
AS	Strasse	Automatisch referenziert; Ort, Str übereinstimmend
B	Bahnhof	Bahnhof – Automatisch referenziert über Bahnhof-File, Genauigkeit vergleichbar A1
C1	Gebäude	wie A1, jedoch referenziert nach vorgängiger manueller Korrektur
C2	Gebäude	wie A2, jedoch referenziert nach vorgängiger manueller Korrektur
CS	Strasse	wie AS, jedoch referenziert nach vorgängiger manueller Korrektur
M1	100 m Radius	Manuell referenziert über „Telinfo“, Unsicherheit kleiner 100 m
M2	500 m Radius	Manuell referenziert über „Telinfo“, Unsicherheit kleiner 500 m
M3	1000 m Radius	Manuell referenziert über „Telinfo“, Unsicherheit kleiner 1000 m
O	Gemeinde	Gemeindemittelpunkt – Automatisch referenziert; Ort übereinstimmend
?		Fragwürdiger Eintrag, wahrscheinlich fehlerhaft (Ortschaften verwechselt, ...)
...-ptv		Eintrag in Deutschland welcher von der Firma PTV geokodiert wurde

4 Ergebnisse

4.1 Übersicht

Nach Abschluss der Arbeiten konnten die in Tabelle 6 und 8 dargestellten *matching*-Quoten erzielt werden („Unsicherheit kleiner 100 m“ bedeutet, dass der effektive Punkt maximal 100 m vom gewählten Punkt entfernt ist):

Tabelle 6 Schlussbilanz der Geokodierung

Klasse	Wege im			Alle	
	Thurgau	Schweiz	Ausland	Anzahl	Anteil
Wege	31180	5325	316	36824	100 %
Unsicherheit kleiner 100 m (AS, CS ausgeschlossen)	80.0%	62%	54%	28382	77.1%
Unsicherheit kleiner 500 m (AS, CS eingeschlossen)	96.8%	85%	54%	34900	94.8%
Georeferenziert	99.3%	98%	94%	36504	99.1%
Nicht georeferenziert	0.7%	2%	6%	321	0.9%

Tabelle 7 Gegenüberstellung der Unsicherheit bei Start- und Zielpunkt aller Wege (Total 36'824)

Unsicherh. Ziel \ Unsicherh. Start	100 m	500 m	1000 m	Ortschaft	Fragwürdig
100 m	22207	4921	42	984	236
500 m	4942	1181	10	326	52
1000 m	40	9	4	6	
Ortschaft	958	352	3	215	15
Fragwürdig	237	51		15	18

Tabelle 8 Gegenüberstellung von Unsicherheit bei Start- und Zielpunkt aller Wege in % des Total

Unsicherh. Ziel \ Unsicherh. Start	100 m	500 m	1000 m	Ortschaft	Fragwürdig
100 m	60.3	13.4	0.1	2.7	0.6
500 m	13.4	3.2	0.0	0.9	0.1
1000 m	0.1	0.0	0.0	0.0	
Ortschaft	2.6	1.0	0.0	0.6	0.0
Fragwürdig	0.6	0.1		0.0	0.0

Tabelle 9 und Tabelle 10 im Anhang zeigen die Qualität von Start und Endpunkt aufgeschlüsselt nach allen möglichen Rating-Kombinationen.

4.2 Überprüfung der Resultate

Zur Prüfung der Plausibilität der Geokodierung wurde eine Qualitätskontrolle vorgenommen. Die Menge der automatisch zugeordneten Einträge ist zu gross, als dass jeder Eintrag überprüft werden könnte. In einem Nachgang wurden alle Einträge der Wegedatei mit den Postleitzahl-Punkten ortsgenau geokodiert und die Luftdistanz zwischen effektiver Zuordnung und diesem Kontrollmatching berechnet. Das Mass der Luftdistanz erlaubte eine Einschätzung der Plausibilität des effektiven Matchings. Dieses Verfahren ermöglicht es einen Teil der groben Fehler aufzufinden. Dennoch können auch fehlerhafte Zuordnungen mit grossen Abweichungen nicht vollständig ausgeschlossen werden.

In diesem Kontrollschritt konnten auf insgesamt 36824 Einträge etwa zehn fehlerhafte Geokodierungen in den Wegedateien eruiert werden. Alle Einträge mit einer Abweichung von grösser als 2 km wurden nachkontrolliert. Abweichungen von bis zu 120 km kommen vor. Sie sind bedingt durch falsche Postleitzahlen bei Gemeindenamen, die in der Schweiz mehrfach vorkommen (Siehe Kapitel 5.2). In diesen Fällen ergibt die ortsgenau geocodierte falsche Resultate.

5 Mögliche Fehlerquellen

Das Hantieren mit grossen Datenmengen birgt immer die Gefahr der Scheingenaugkeit der Resultate. Deshalb soll im Anschluss an die Ergebnisdarstellungen nochmals auf die Qualität der Ausgangsdaten und des Verfahrens eingegangen werden.

Die Restriktionen der Geokodierung ergeben sich vor allem aus:

- der Qualität der Punktdaten und
- der Qualität der Umfrage-Daten.

5.1 Qualität der Punktdaten

Problematisch sind die fehlenden Ortsnamen bei den Gebäudepunkten in den ländlichen Gebieten des Kantons Thurgau. Der Thurgau weist allgemein eine sehr dezentrale und zergliederte Siedlungsstruktur auf. Zahlreiche Gemeinden (Bsp. Hinterneunforn) bestehen nur aus Weilern, die ohne ein klar ausgeprägtes Zentrum über eine grosse Fläche verstreut liegen.

Aufgrund der geringen Grösse vieler Weiler, kann alleine über den Weilernamen eine sehr genaue Geokodierung erfolgen. Weil aber die Ortsnamen im Punktdatenfile des Kantons Thurgau bzw. im Gebäuderegister des Schweizer Bundesamtes für Statistik fehlen, war für Einträge ohne Strassenbezeichnung nur eine automatische Geokodierung über die Postleitzahl möglich. Eine solche ist in Anbetracht der grossen Gemeindeflächen mit einer entsprechenden Ungenauigkeit verbunden.

Es kann zu Fehlern kommen, wenn in verschiedenen Weilern dieselbe Strasse existiert. Ein Beispiel einer durch Zufall entdeckten falschen Georeferenzierung ist folgender Eintrag:

„8556 - Engwang - Dorfstr. 2“.

Die automatische Georeferenzierung hat dem Eintrag die Koordinaten für „8556 – Illhart - Dorfstr. 2“ zugewiesen. Korrekt wären die Koordinaten der Dorfstrasse in Engwang. Derartige Fehler können nicht ausgeschlossen werden solange zum Georeferenzieren ein Datensatz ohne Ortsnamen verwendet wird.

5.2 Qualität der Umfrage-Daten

Die Rohdaten scheinen eine allgemein gute Qualität aufzuweisen. Es wurden im Verlauf des der Arbeiten jedoch einige typische Fehlerquellen entdeckt, die nachfolgend exemplarisch aufgelistet werden. Die meisten dieser Fehler wurden während des manuellen Arbeitsschrittes bemerkt und korrigiert. Es ist aber bei bestimmten Konstellationen möglich (vor allem bei weit verbreiteten Strassennamen wie „Dorfstrasse“ oder „Hauptstrasse“), dass Einträge automatisch falsch zugeordnet wurden. Derartige Fehler sind grundsätzlich schwer aufzuspüren.

Falsche Postleitzahlen

Aus einigen Einträgen wurde ersichtlich, dass bei einem Teil der Umfragedaten nachträglich fehlende Postleitzahlen zugewiesen wurden. Ein Teil dieser Zuweisungen ist falsch. Es betrifft dies vor allem Ortschaften, deren Namen in der Schweiz nicht einmalig ist. Beispielsweise erhielt „Nussbaumen TG“ die PLZ von „Nussbaumen AG“. Es kann angenommen werden, dass diese Fehler nicht stark ins Gewicht fallen. Derartige Einträge sind kaum automatisch zugeordnet worden, weil sich nur selten eine zusätzliche Übereinstimmung beim Strassennamen finden lässt. Die meisten dieser Einträge wurden spätestens beim manuellen Arbeitsschritt entdeckt. Ein Beispiel:

5646 - Abtwil - Gaiserwaldstr.

5646 ist die PLZ von Abtwil, AG, die Gaiserwaldstrasse befindet sich jedoch in 9030 Abtwil, SG. Der Eintrag wurde auf „9030 – Abtwil SG – Gaiserwaldstrasse“ geokodiert (Rating: CS). Gemäss „Telinfo“ gibt es in der Schweiz nur eine Gaiserwaldstrasse, es kann also von einer sehr sicheren Zuweisung ausgegangen werden.

Vermischte Zielangaben

Eine zweite mögliche Fehlerquelle in den Umfrage-Daten sind falsche Angaben seitens der befragten Personen. Bei vielen Einträgen gewinnt man den Eindruck, zwei Ziele seien von den befragten Personen zu einem Ziel vermischt worden:

8180 - Nussbaumen – Hüttwilerstr. 5

8180 ist die PLZ von Bülach, welches einige Dutzend Kilometer von 8537 Nussbaumen, TG entfernt ist. Korrekt wäre: „8537 - Nussbaumen - Hüttwilerstr. 5“. Der Eintrag wurde als

fragwürdig (Rating: ?) eingestuft. Es ist möglich, dass die Person in Wirklichkeit von Bülach nach Nussbaumen, TG gefahren ist, beim Ausfüllen des Fragebogens dann aber die beiden Wegpunkte „8180 Bülach“ und „8537 Nussbaumen, TG“ vermischt hat. Ein weiteres Beispiel:

8570 – Weinfelden – Hafenstr. 50d

„8570 – Weinfelden“ liegt nicht am Bodensee und verfügt demzufolge auch über keine Hafenstrasse. Hingegen existiert im Gebäudekataster ein Eintrag „8280 Kreuzlingen Hafenstrasse 50d“. Der Eintrag wurde dennoch als fragwürdig bewertet (Rating: ?). Es ist möglich, dass zwei Ziele miteinander vermischt wurden. Die ursprüngliche Fehlerursache des folgenden Beispiels lässt sich nicht mehr genau erruieren:

3981 - Brig - Furkastr. 5

Die PLZ von Brig ist 3900. 3981 ist die PLZ von „Geschinen“ (ca. 30 km talaufwärts von Brig). In beiden Gemeinden existiert eine Furkastrasse. Der Eintrag wurde als fragwürdig (Rating ?) bewertet. Eine Georeferenzierung wäre allenfalls unter Einbezug der Wegekette möglich.

Benachbarte Gemeinden

Sehr häufig sind fehlerhafte Einträge aufgrund der Verwechslung benachbarter Gemeinden. (Die Einträge dieser Klasse lassen sich in vielen Fällen nicht einwandfrei von Einträgen unterscheiden bei denen wahrscheinlich Ziele vermischt wurden.) Nachfolgendes Beispiel wurde mit dem *Rating* M1 taxiert:

8500 - Frauenfeld - Hinterhorben.

Hinterhorben existiert in der Schweiz gemäss „Telinfo“ nur einmal in „8524 – Buch bei Frauenfeld“. Weil „Buch bei Frauenfeld“ direkt an „Frauenfeld“ grenzt, wurde angenommen, dass der Eintrag sich auf „Buch bei Frauenfeld“ bezieht.

Ungenauere Ortsangaben

Eine grosse Gruppe stellen die Einträge mit ungenauer Ortsangabe dar. Viele Befragte gaben anstelle einer Strassenbezeichnung einen Bereich wie „Altstadt“ oder „Waldstück Herzen-

holz“ an. In solchen Fällen wurde versucht die räumliche Ausdehnung dieses Bereiches abzuschätzen und ein entsprechendes *Rating* (M1, M2 oder M3) zuzuweisen. Ein Beispiel ist der Eintrag:

8260 – Stein am Rhein – Altstadt

Abbildung 1 zeigt die Situation und das gewählte *Rating*. Die Ausdehnung der Altstadt von Stein am Rhein ist nur unwesentlich grösser als ein Kreis mit 100 m Radius. Daher wurde der Eintrag auf einen Punkt im Zentrum der Altstadt geokodiert und mit dem *Rating* M1 bewertet. Siehe auch die weiteren Beispiele im Anhang (Abbildung 2 bis Abbildung 4)

5.3 Subjektivität des manuellen Arbeitsschritts

Manuelle Geokodierung ist subjektiv. Zuerst muss die Glaubwürdigkeit des Eintrages abgeschätzt werden und in einem zweiten Schritt muss visuell bestimmt werden, welche Genauigkeit vorliegt. In einigen Fällen ist bekannt, welches Objekt gemeint ist, subjektiv ist aber die Abschätzung der Ausdehnung des Objekts. Ein Beispiel hierfür ist der oben vorgestellte Eintrag „Stein am Rhein – Altstadt“. In anderen Fällen ist bereits die Wahl des Objekts problematisch. Lautet die Strassenbezeichnung „Oberdorf“, es existieren aber nur „Oberdorfstrasse“ und „Im Oberdorf“, dann wurde ein Mittelpunkt der beiden Strassen gewählt. Auf diesen Punkt wurde der Eintrag geokodiert. Für die Wahl des *Rating*s wurde die entstandene Unsicherheit in Metern abgeschätzt.

6 Empfehlungen

Bei der Bearbeitung des Projektes wurden Erkenntnisse gewonnen, die an dieser Stelle als Empfehlungen für weitere Geokodierungs-Arbeiten aufgeführt werden sollen. Für eine Besprechung von Erfahrungen, die im Zusammenhang mit einer deutlich aufwendigeren Geocodierung gewonnen wurden, sei auf Jermann (2003) verwiesen.

Das Adress-Geokodierung stellte sich unabhängig von der Qualität der Punktdaten als mühsamer Arbeitsschritt heraus. Die Harmonisierung der Wegebezeichnungen, Umlaute und Sonderzeichen ist umständlich. Ebenfalls viel Zeit wurde beansprucht um Rechtschreibung der Daten zu korrigieren (Oberstrasse statt Oberstrasse, Reinstr. anstelle Rheinstr. usw.)

Zeitersparnisse können sich ergeben, wenn mit einem Programm gearbeitet würde, das Buchstabenähnlichkeiten erkennen kann¹. Access ist dazu nicht fähig und kann nur exakt übereinstimmende Einträge verknüpfen. Andererseits birgt die automatische Erkennung von Ähnlichkeiten auch Gefahren. In sehr vielen Dörfern und Städten gibt es Strassen deren Namen sich sehr ähnlich sind (Beispielsweise „Im Unterdorf“ und „Unterdorfweg“).

7 Literatur

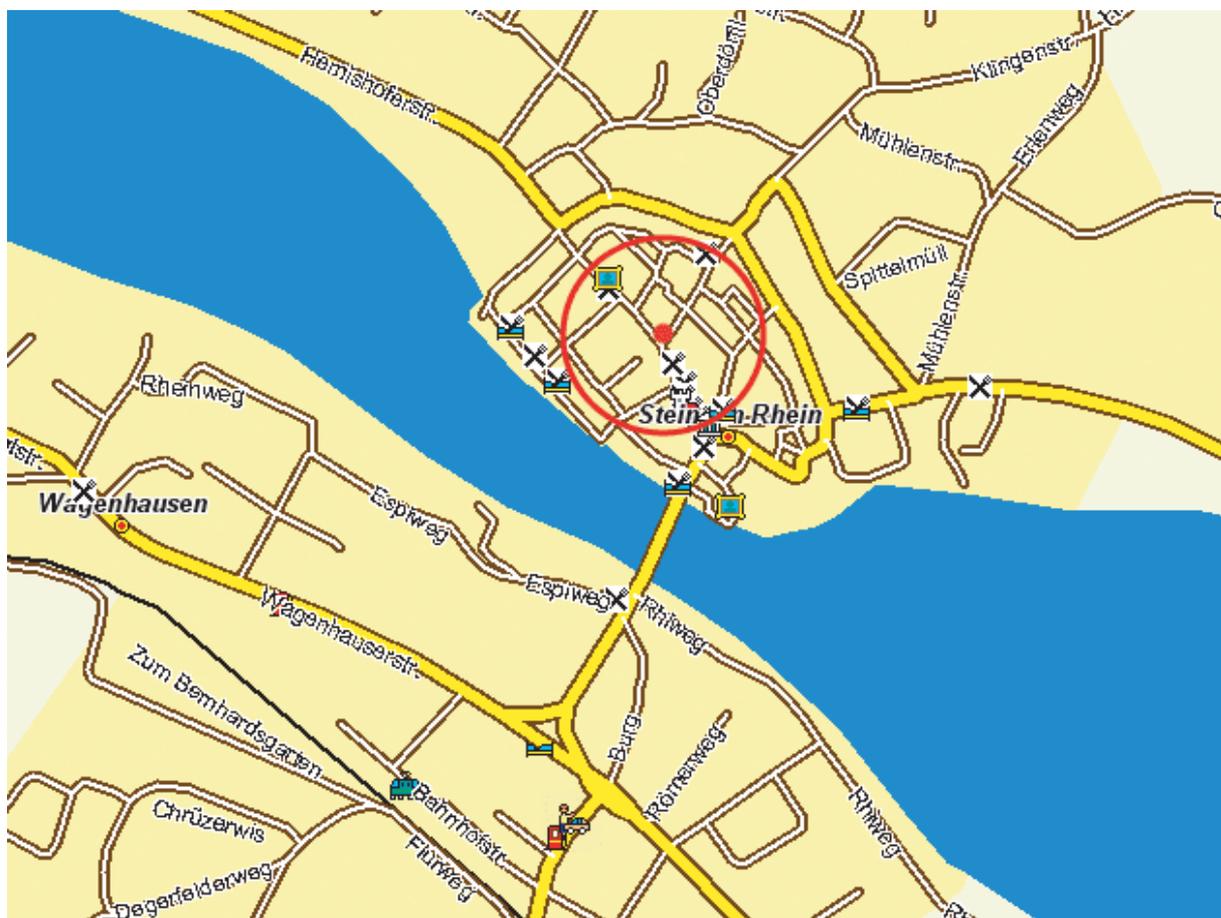
Jermann J. (2002) *GIS-gestützte Modellierung von Anmarschwegen auf Haltestellen des öffentlichen Verkehrs*, Diplomarbeit UNIGIS MAS, Institut für Geographie und angewandte Geoinformatik, Universität Salzburg, Salzburg.

Jermann J. (2003) *Geokodierung Mikrozensus 2000*, Arbeitsbericht Verkehrs- und Raumplanung, **177**, IVT, ETH Zürich, Zürich.

¹ Als Beispiel kann hier die als Fachschale des Statistikprogramms SAS konzipierte Software DataFlux genannt werden (www.dataflux.com).

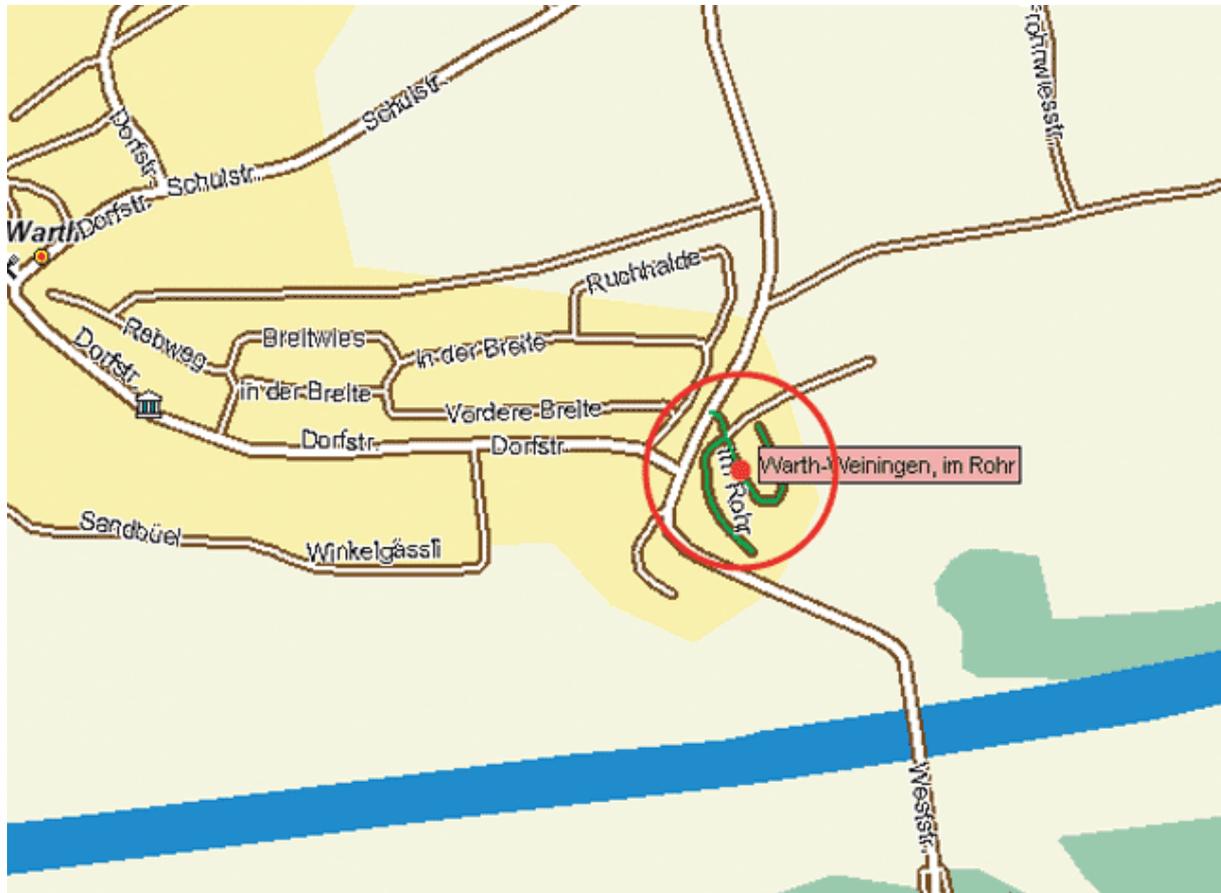
Appendix A Beispiele zur Genauigkeit der manuellen Geokodierung

Abbildung 1 Beispiel für *Rating* M1, Kreisradius 100m: „Stein am Rhein – Altstadt“

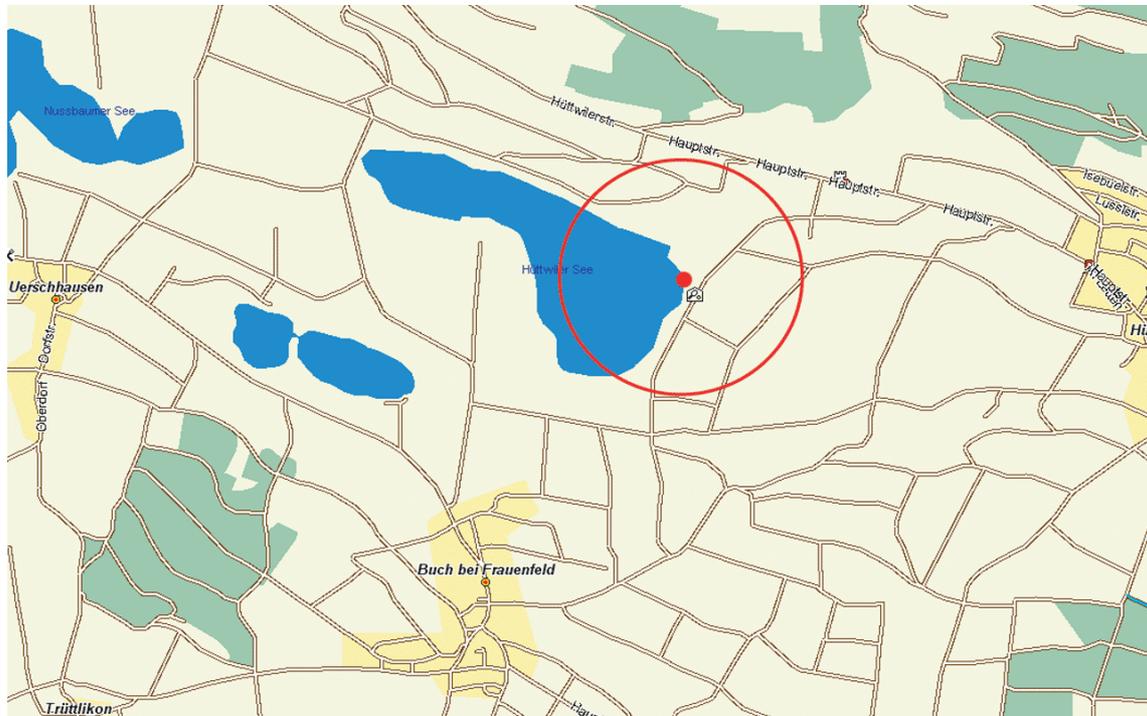


Quelle Kartengrundlage: Telinfo

Abbildung 2 Beispiel für *Rating* M1, Kreisradius 100m: Warth-Weinigen - Im Rohr



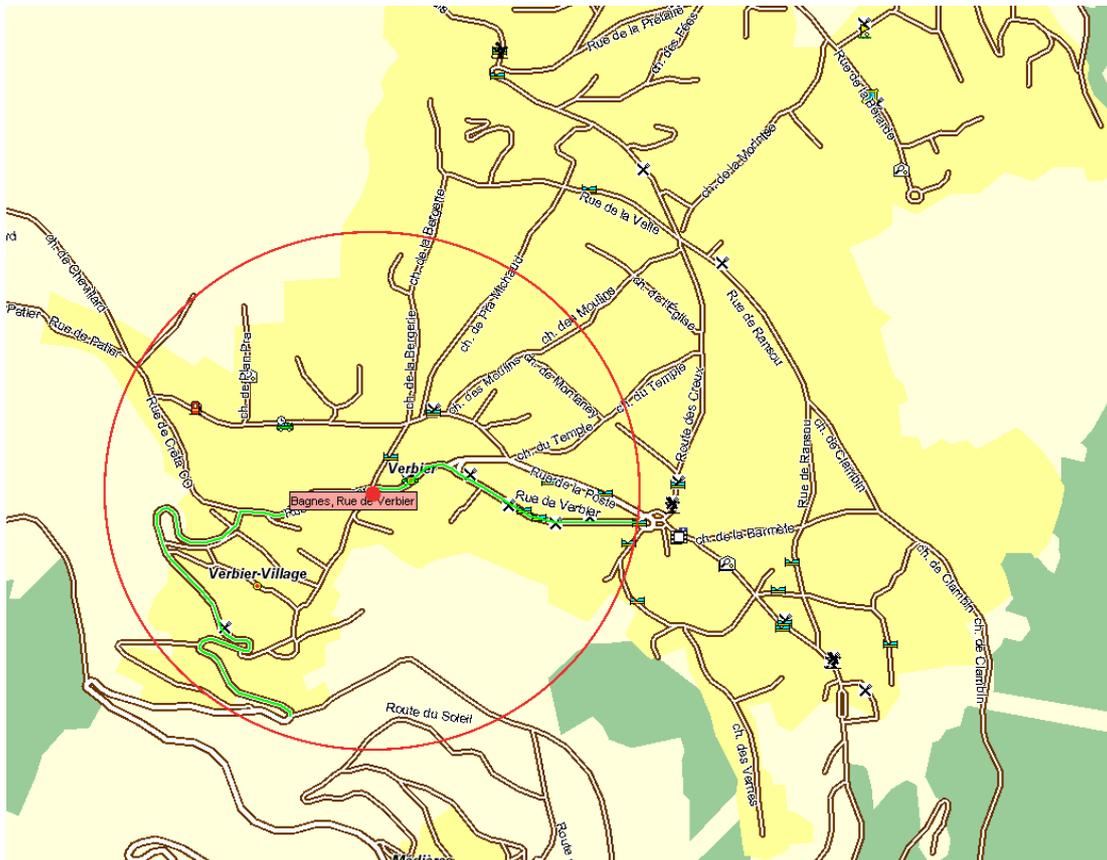
Quelle Kartengrundlage: Telfino

Abbildung 3 Beispiel für *Rating* M2, Kreisradius 500m: Hüttwiler See

Das Wegziel „Hüttwiler See“ wurde häufiger genannt. In vielen Fällen kann angenommen werden, dass die Befragten sich im Strandbad Hüttwilersee aufgehalten haben. Die genaue Position des Strandbades konnte über „Telinfo“ ermittelt werden und der Eintrag wurde auf diesen Punkt geokodiert. Trotzdem wurde der Eintrag mit einer Unsicherheit von 500 m bewertet. Es kann nicht ausgeschlossen werden, dass die Personen sich an einem anderen Punkt des Sees aufgehalten haben

Quelle Kartengrundlage: Telinfo

Abbildung 4 Beispiel Rating M2, Kreisradius 500m: Verbier – Route de Verbier



Im Falle der Route de Verbier, wurde ein Punkt in der Mitte der Strasse gewählt und die Unsicherheit über die Länge der Strasse bestimmt.

Quelle Kartengrundlage: Telinfo

Tabelle 9 Rating von Start- und Endpunkt. Anzahl Einträge pro Kombination. Total 36'824

Ziel Rating Start Rating	?	A1	A1-ptv	A2	A3	AS	B	C1	C2	CS	M1	M1-ptv	M2	M3	M3-ptv	O	O-ptv
?	18	106		36	75	48	4	2	1	2	12		1			15	
A1	104	12194	54	2111	263	2885	323	376	82	342	647	11	357	27	3	723	4
A1-ptv		65	43	7	3	11	1				11	6			1	5	1
A2	32	2118	5	749	130	631	54	86	7	51	174		73	3		128	
A3	80	245	5	116	69	68	19	32	24	7	9		33	2		39	
AS	40	2819	15	707	77	676	69	73	65	57	176		151	5	2	249	
B	4	323	2	44	18	65	26	8		13	23		28			17	
C1	6	384	2	74	30	63	9	24	1	9	22		6	1		21	
C2		78		10	15	67	1			4	3		5			2	
CS	4	367		47	5	46	7	6	2	8	9		11	1		30	
M1	9	673	11	152	10	176	22	27	3	8	158		16	5		38	1
M1-ptv		8	6			2				1	1					2	3
M2	8	363		58	22	163	27	4		9	24		60	2		46	1
M3		27		4	1	6		1		1	3		1	2		5	
M3-ptv		1	1		1	1						1			2	1	
O	15	730	7	123	29	275	11	11		32	36	2	44	3		213	
O-ptv		2	3								1	3	1				2

Tabelle 10 Rating von Start- und Endpunkt. Anteil jeder Kombination in % am Total

Ziel Rating Start Rating	?	A1	A1-ptv	A2	A3	AS	B	C1	C2	CS	M1	M1-ptv	M2	M3	M3-ptv	O	O-ptv
?	0.0	0.3		0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0		0.0			0.0	
A1	0.3	33.1	0.1	5.7	0.7	7.8	0.9	1.0	0.2	0.9	1.8	0.0	1.0	0.1	0.0	2.0	0.0
A1-ptv		0.2	0.1	0.0	0.0	0.0	0.0				0.0	0.0			0.0	0.0	0.0
A2	0.1	5.8	0.0	2.0	0.4	1.7	0.1	0.2	0.0	0.1	0.5		0.2	0.0		0.3	
A3	0.2	0.7	0.0	0.3	0.2	0.2	0.1	0.1	0.1	0.0	0.0		0.1	0.0		0.1	
AS	0.1	7.7	0.0	1.9	0.2	1.8	0.2	0.2	0.2	0.2	0.5		0.4	0.0	0.0	0.7	
B	0.0	0.9	0.0	0.1	0.0	0.2	0.1	0.0		0.0	0.1		0.1			0.0	
C1	0.0	1.0	0.0	0.2	0.1	0.2	0.0	0.1	0.0	0.0	0.1		0.0	0.0		0.1	
C2		0.2		0.0	0.0	0.2	0.0			0.0	0.0		0.0			0.0	
CS	0.0	1.0		0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0		0.0	0.0		0.1	
M1	0.0	1.8	0.0	0.4	0.0	0.5	0.1	0.1	0.0	0.0	0.4		0.0	0.0		0.1	0.0
M1-ptv		0.0	0.0			0.0				0.0	0.0					0.0	0.0
M2	0.0	1.0		0.2	0.1	0.4	0.1	0.0		0.0	0.1		0.2	0.0		0.1	0.0
M3		0.1		0.0	0.0	0.0		0.0		0.0	0.0		0.0	0.0		0.0	
M3-ptv		0.0	0.0		0.0	0.0						0.0			0.0	0.0	
O	0.0	2.0	0.0	0.3	0.1	0.7	0.0	0.0		0.1	0.1	0.0	0.1	0.0		0.6	
O-ptv		0.0	0.0								0.0	0.0	0.0				0.0