Chapter 16. GEOCODING AND DYNAMIC SEGMENTATION

Geocoding

- Geocoding refers to the process of assigning spatial locations to data that are in tabular format but have fields that describe their locations.
- Address geocoding is the most common type of geocoding, which plots street addresses as point features on a map.
- Address geocoding interpolates the location of a street address by comparing it with data in the reference database.

Geocoding Reference Database

Address geocoding requires two sets of data. The first data set contains individual street addresses in a table, one record per address. The second is a reference database that consists of a street map and attributes for each street segment such as the street name, address ranges, and ZIP codes.

Name,	Address,	Zip
Iron Horse,	407 E Sherman Ave,	83814
Franlin's Hoagies,	501 N 4th St,	83814
McDonald's,	208 W Appleway,	83814
Rockin Robin Cafe,	3650 N Government way,	83815
Olive Garden,	525 W Canfield Ave,	83815
Fernan Range Station,	2502 E Sherman Ave,	83814
FBI,	250 Northwest Blvd,	83814
ID Fish & Game,	2750 W Kathleen Ave,	83814
ID Health & Welfare,	1120 W Ironwood Dr,	83814
ID Transportation Dept,	600 W Prairie Ave,	83815

Figure 16.1
A sample address table records name, address, and ZIP code.

FEDIRP: A direction that precedes a street name.
FENAME: The name of a street.
FETYPE: The street name type such as St, Rd, and Ln.
FRADDL: The beginning address number on the left side of a street segment.
TOADDL: The ending address number on the left side of a street segment.
FRADDR: The beginning address number on the right side of a street segment.
TOADDR: The ending address number on the right side of a street segment.
ZIPL: The zip code for the left side of a street segment.
ZIPR: The zip code for right side of a street segment.

Figure 16.2
The TIGER/Line files include the attributes of FEDIRP, FENAME, FETYPE, FRADDL, TOADDL, FRADDR, TOADDR, ZIPL, and ZIPR, which are important for geocoding.

The Address Matching Process

In general, the address matching process consists of three phases: preprocessing, matching, and plotting.

- The preprocessing phase involves parsing and address standardization.
- In the matching phase, the geocoding engine matches the address against a reference database.
- If an address is judged to be matched, it is plotted as a point feature by interpolating where the address falls within the address range.

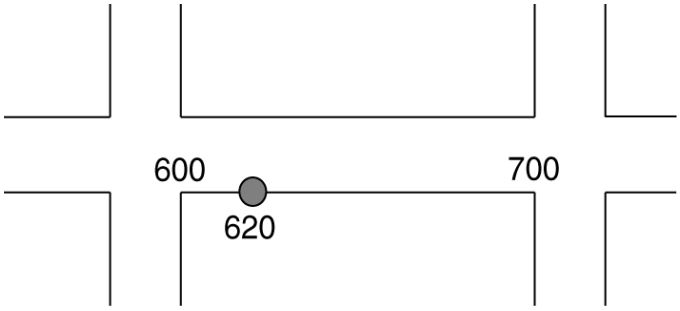


Figure 16.3
Linear interpolation for address geocoding.

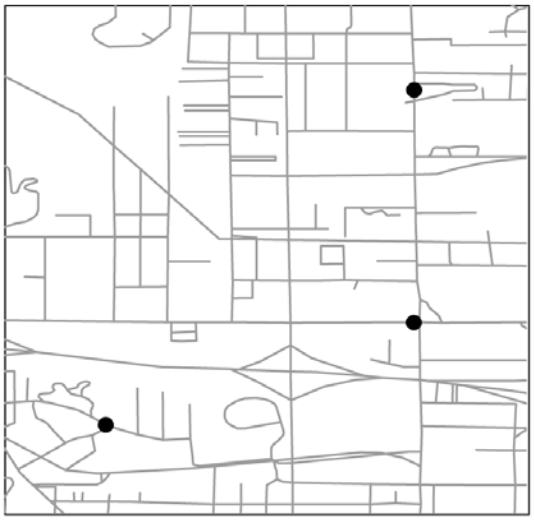


Figure 16.4
Address geocoding plots street addresses as points on a map.

Address Matching Options

- Typically, a geocoding engine has provisions for relaxing the matching conditions but uses a scoring system to quantify the matches at the same time.
- Because of the various matching options, we can expect to run the geocoding process more than once.

Acceptable Matched Rate

- The result of geocoding is expressed as the percentage of addresses matched.
- For crime mapping and analysis, one researcher has stated that a 60% hit rate is unacceptable and another has derived statistically a minimum acceptable hit rate of 85%.
- To bring the rate to 95% or better, as required for competitive location-based services, a current and accurate reference database and additional effort in validating street addresses would be required.

Offset Plotting Options

GIS packages have the side offset and end offset options for plotting geocoded points.

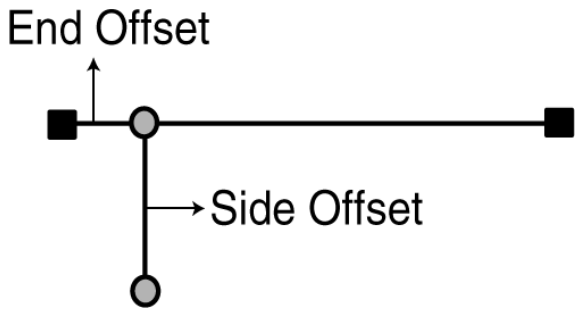


Figure 16.5
The end offset moves a geocoded point away from the end point of a street segment, and the side offset places a geocoded point away from the side of a street segment.

Variations of Geocoding

- Intersection matching matches address data with street intersections on a map.
- ZIP code geocoding refers to the process of matching a ZIP code to its centroid location.
- Reverse geocoding is a process of converting latitude and longitude coordinate data of locations into descriptive addresses.

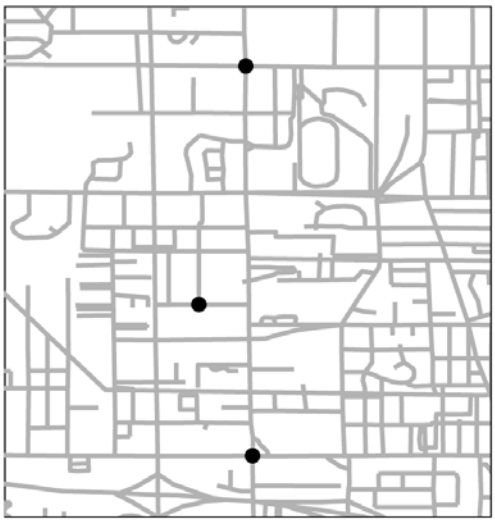


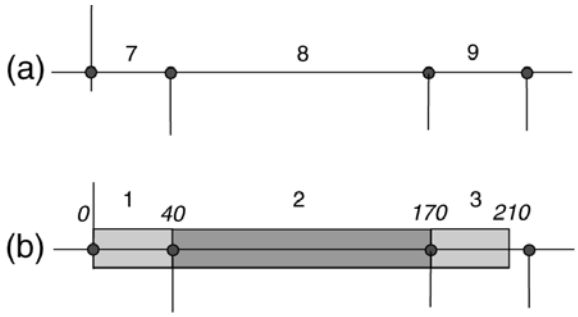
Figure 16.6
An example of Intersection matching.

Applications of Geocoding

1. Location-based services
2. Business applications
3. Wireless emergency services
4. Crime mapping and analysis
5. Public health

Dynamic Segmentation

- Dynamic segmentation refers to the process of computing the location of events along a route.
- A route is a linear feature with a linear measurement system stored with its geometry.
- Events are linearly referenced data that occur along routes.



(c)

Section	Arc	F-Meas	T-Meas	F-Pos	T-Pos
1	7	0	40	0	100
2	8	40	170	0	100
3	9	170	210	0	80

Figure 16.7
An example of a route subclass using the coverage model.

Point	x	y	m
1	x_1	y_1	0
2	x_2	y_2	40
3	x_3	y_3	170
4	x_4	y_4	210

Figure 16.8
An example of a route feature class.

Creating Routes

Routes can be created interactively or through data conversion.

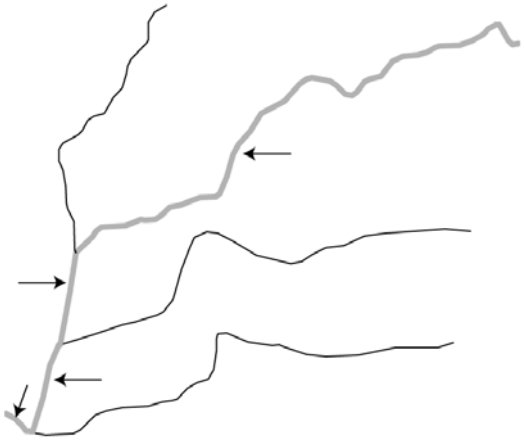


Figure 16.9
The interactive method requires the selection or digitizing of the line segments that make up a route (shown in a thicker line symbol).

Figure 16.10
Interstate highway routes in Idaho.

Types of Routes

1. Simple route
2. Combined route
3. Split route
4. Looping route

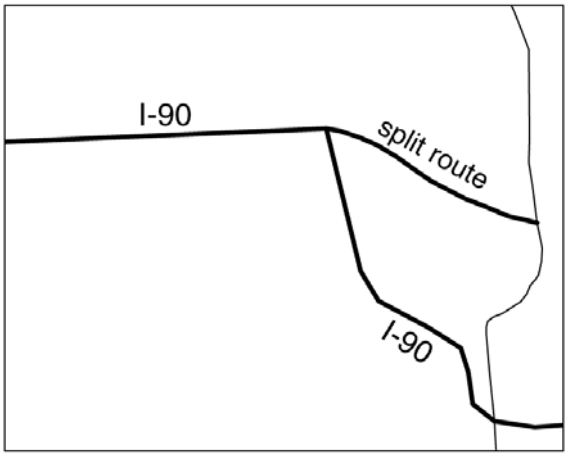


Figure 16.11
An example of a split route.

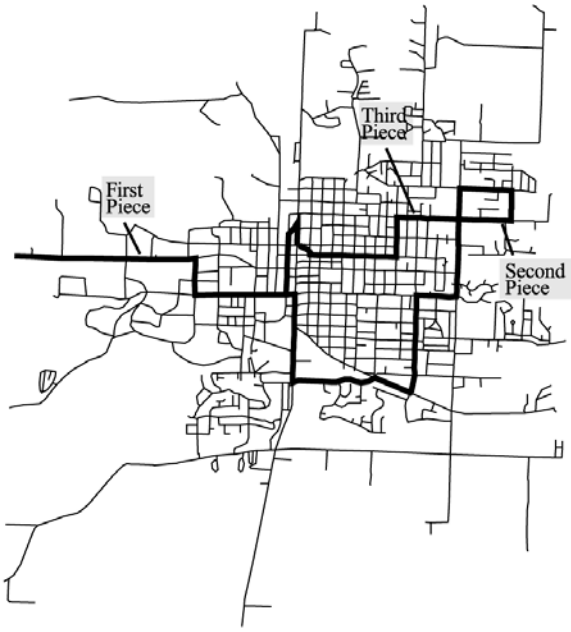


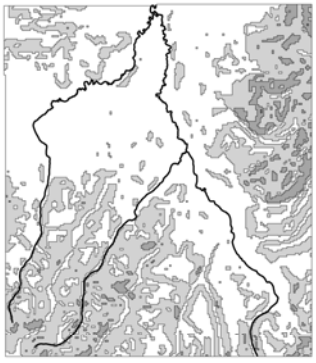
Figure 16.12
A looping route divided into three parts for the purpose of route measuring.

Creating Events

- Events can be point or line events.
- There are two common methods for creating event tables. The first method creates an event table from an existing table that already has data on route ID and linear measures. The second method creates an event table by locating point or polygon features along a route.

FID	Route-ID	Measure	RDLR
1	90	161.33	R
2	90	161.82	L
3	90	198.32	R
4	90	198.32	L

Figure 16.13
An example of converting point features to point events.



Route-ID	F-Meas	T-Meas	Slope-Code
1	0	7,638	1
1	7,638	7,798	2
1	7,798	7,823	1
1	7,823	7,832	2
1	7,832	8,487	1
1	8,487	8,561	1
1	8,561	8,586	2
1	8,586	8,639	1
1	8,639	8,643	2

Figure 16.14
An example of creating a linear event table by overlaying a route layer and a polygon layer.

Applications of Dynamic Segmentation

1. Data management
2. Data display
3. Data query
4. Data analysis

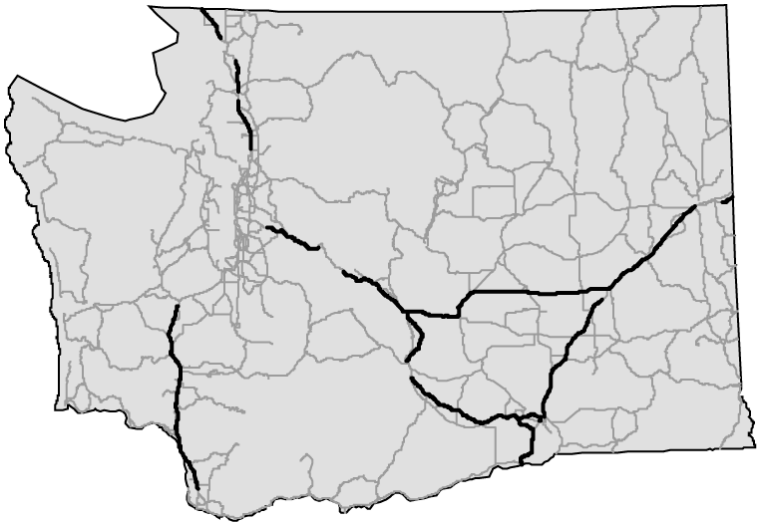


Figure 16.15
The thicker, solid line symbol represents those portions of the Washington State's highway network that have the legal speed limit of 70 miles per hour.

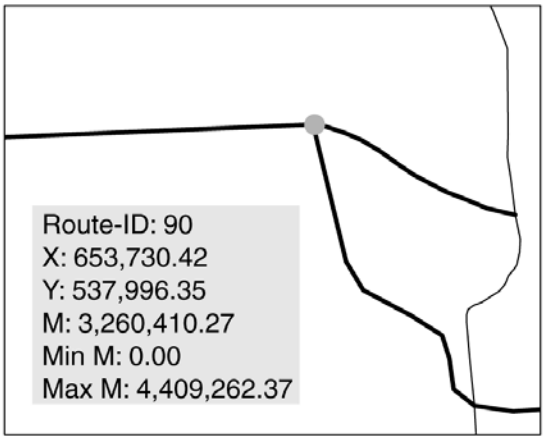


Figure 16.16
Data query at a point, shown here by the small circle, shows the route-ID, the *x*- and *y*-coordinates, and the measure (*m*) value at the point location. Additionally, the beginning and ending measure values of the route are also listed.

Google Maps
http://www.google.com/
U.S. Census Bureau MAF/TIGER system
http://www.census.gov/geo/www/tiger/index.html
Tele Atlas
http://www.teleatlas.com/
NAVTEQ
http://www.navteq.com/
Trillium Software
http://www.trilliumsoftware.com/
GeoDirectory, Ireland
http://www.geodirectory.ie/
Sanborn
http://www.sanborn.com/
MapQuest
http://www.mapquest.com
CrimeStat
http://www.icpsr.umich.edu/NACJD/crimestat.htm
Washington State Department of Transportation GIS Data
http://www.wsdot.wa.gov/mapsdata/geodatacatalog/default.htm
ESRI's Geodatabase Website
http://support.esri.com/datamodels

ESRI's Download Website
http://www.esri.com/data/download/census2000_tigerline/index.html
National Hydrography Dataset: NHDinGEO
http://nhd.usgs.gov/geodatabase_review.html