

GEOG 204

LECTURE 8
Geocoding

1

Geocoding

- Geocoding
 - The process of assigning spatial locations to descriptive data for those locations
 - from location description to spatial locations
 - Started in the 60's when US Census Bureau was looking for ways of mapping survey data address by address
 - Descriptions of locations
 - points of interest, place names for landmarks, infrastructure stores, addresses, street intersections, postal codes

2

2

Geocoding

- The most common type of geocoding is *address matching*
- Address matching
 - plots street addresses as point features
 - 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 postal code
 - A reference database must have a road network with appropriate attributes for geocoding

3

3

¹ 33 N. Fawn Road
Chatham, QC J8G 2X8

² 9083 Parker St.
Lachute, QC J8H 3C4

³ 23 N. Surrey St.
Youngs Cove, NB E4C 5J3

⁴ 7175 Fawn Ave.
Edmundston, NB E3V 7M1

⁵ 310 SW. Hanover Street
Cobourg, ON K9A 6A7

⁶ 60 Augusta Rd.
Kingston, NB E5S 4M8

Number Column		Street Name Column		ZIP Column (optional)	
Number		Street		(none)	
Street Layer					
OAKVLrds					
Street Name Attribute					
STREET					
Left From Number	Right From Number	From X Attribute	From Y Attribute		
FROMLEFT	FROMRIGHT	(street line order)	(street line order)		
Left To Number	Right To Number	To X Attribute	To Y Attribute		
TOLEFT	TORIGHT	(street line order)	(street line order)		
Left ZIP (optional)	Right ZIP (optional)	Setback (Map Units)			
(none)	(none)	0			

4

4

Field Name	Alias Name
*From Left	FROMLEFT
*To Left	TOLEFT
*From Right	FROMRIGHT
*To Right	TORIGHT
Prefix Direction	PREDIR
Prefix Type	PRETYPE
*Street Name	STREETNAME
Suffix Type	SUFTYPE
Suffix Direction	SUFDIR
Left City or Place	LEFT_MAF
Right City or Place	RIGHT_MAF

Attributes in a reference database for geocoding

5

5

Geocoding

- Address geocoding typically interpolates the location of a street address by comparing it with data in the reference database
 - If the addresses along a block range from 100 to 200, then house number 130 is about one-third of the way along the block
- It is also known as linear referencing
 - Requires a set of addresses associated with a set of linear features

6

6

Geocoding

- In simple terms it requires the starting and ending address for a link in a network (i.e. a road segment in a road network).
- The starting and ending addresses define the address range
- The range is assumed to linearly span the entire link (segment)
- Points are then geographically coded (hence geocoding) by interpolating where they ought to fall on the link
 - In this case interpolating is proportionately approximating the location of an address based on distance from the end points

7

7

Address Matching

- The geocoding process uses a geocoding engine, which is embedded in a GIS
 - In ArcGIS, the geocoding engine is called the Address Locator.
 - In QGIS, the MMQGIS plugin has the Geocode processor
- The geocoding process consists of three phases:
 - preprocessing
 - matching
 - plotting

8

8

Address Matching

- The preprocessing phase parses and standardizes the address
- Parsing breaks down an address into a number of components.
- The result of a parsing process is a record in which there is a value for each of the address components to be matched

9

9

Address Matching

23 N. Surrey St.
Youngs Cove, NB
E4C 5J3

- 23 --- House Number
- S --- Prefix Direction
- Surrey --- Street Name
- St --- Street type
- City --- Youngs Cove
- NB --- Province
- E4C 5J3 --- Postal Code

Some addresses have apartment numbers associated with the house Number

Others have suffixes such as NE following the street name

10

10

Address Matching

- Address Standardization
 - Standardizes variations of an address into consistent form
 - North and N, Avenue and Ave, First, 1st and 1
 - If using speech to text translation, Smith and Smythe may be treated differently

11

11

Address Matching

- Address Matching
 - The engine matches the address against a reference database
 - Mismatches can occur
 - Errors include:
 - Misspelling of street name,
 - incorrect address number, direction suffix, street type
 - Incorrect or missing postal code
 - Unusual abbreviation not recognized by the geocoding engine
 - The reference database can be out of date

12

12

Address Matching

- Error Examples

1 Joy Ridge Stt. Eskazoni, NS B1W 4P6	Eskazoni/Eskasoni
59 Miles Driv. Windsor, LB A2B 2B7	Driv/Dr
8003 E. Drive Lantz, NS B2S 1B8	E. Drive/East Drive
243 Anderson St. Saint-Luc, QC J2W 0L7	J2W 0L7/J2W 0L7
Courtland Dr. Huntsville, ON P1H 3Y2	Miss stree number

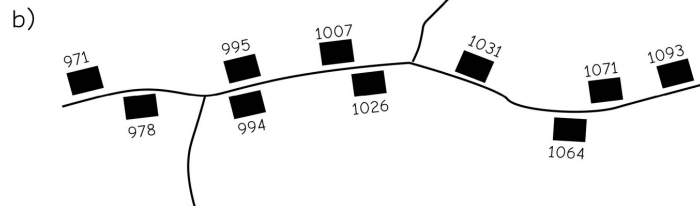
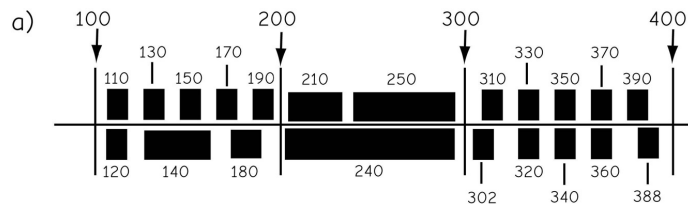
13

13

Address Matching

- **Error Examples**

- The address 250 is not halfway between 200 and 300, and address 240 takes up an entire block. This ordinal/interval mismatch may be particularly bad in rural areas, where irregular development may result in nonlinear address arrangements.
- Part b) illustrates how nonlinear addresses can cause substantial confusion. Address 1007 almost opposite address 1026, and numerous inconsistent intervals; for example, the 22 address units between 1071 and 1093 are separated by a shorter distance than the 12 address units between 995 and 1007.



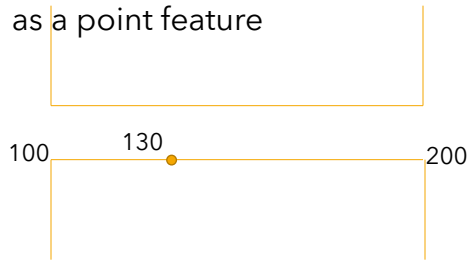
14

14

Address Matching

- Plotting

- Essentially linear interpolation is used to approximate where an address falls on a line segment (road segment)
- If an address is judged to be matched the final step is to plot it as a point feature



Note: Odd street numbers are on one side of the street.
Even numbers on the other

15

15

Address Matching Options

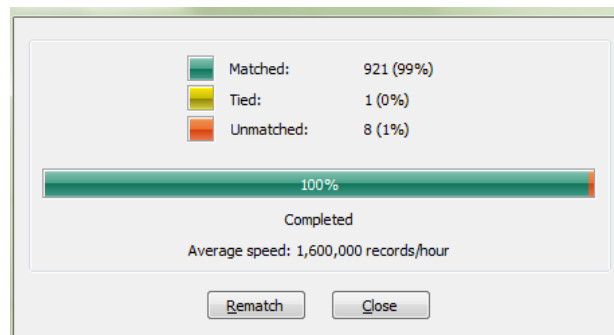
- Some Geocoding engines will have options for relaxing the matching conditions
 - ArcGIS will provide minimum matching score and candidate score
 - Candidate score -> if an address has a likely candidates in the reference database
 - Matching Score - > if an address actually matches

16

16

Quality of Geocoding

- The quality of the geocoding results is sometimes expressed as the percentage of addresses matched.



ArcGIS results

17

17

Variations of Geocoding Process

- Intersection matching matches address data with street intersections (or corners) on a map. E.g. Victoria St & 7th Ave
- ZIP/Postal Code code geocoding refers to the process of matching a ZIP/Postal code to its centroid location.
- Reverse geocoding is a process of converting latitude and longitude coordinate data of locations into descriptive addresses.
- Photo geocoding uses location information of photographs to map the locations
- Parcel-level geocoding, given a parcel reference databases, a parcel number is matched to the centroid of the parcel

18

18

Applications of Geocoding

- Location-based services
 - Google, Yahoo, Apple
- Business applications
 - Matching customer addresses and/or postal code to census data
- Wireless emergency services
 - Phone GPS receiver used to identify location. Wireless providers are required to provide location accuracy
- Crime mapping and analysis
 - Geocoding addresses
- Public health
 - Geocoding and mapping for neighborhood

19