

GIS Workshop

Geocoding – Revisited

Chieko Maene
c-maene@northwestern.edu
Maps and State Documents Librarian
Government and Geographic Information and Data Services
University Library (Evanston)
Northwestern University

Geocoding

- Geocoding:
 - The process of assigning a pair of approximate geographic coordinates to each case (with spatial information, such as: address, zipcode or place name, etc.)
 - Two elements: Source & Reference files
- Address matching
 - Most common geocoding method – matching street addresses with reference street data
 - Address matching methods vary by country

How Does Geocoding Work? (US Streets)

Original Address

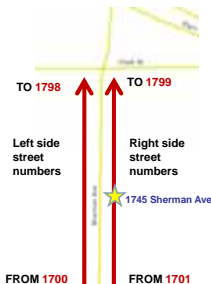
1745 Sherman Ave, Evanston, IL, 60201

Parsed Address

1745 | Sherman | Ave | Evanston | IL | 60201

Street database (reference file) looks for best matching record(s)

- ✓ Find streets (& type): Sherman (Ave)
- ✓ Find street segments: 1745 falls between 1701 – 1799 (right side of the street line segment)
- ✓ Calculate/interpolate a geographic coordinate of the street number, 1745, base on the range of the street line segment (plus a given offset distance)



Geocoding Problems and Solutions

- New or irregular street addresses
 - New subdivisions and rural roads
 - Non-grid street system (Washington DC)
 - International countries (Ex. Japan – street addresses are assigned chronologically)
- Expensive street reference data
 - International street (very expensive to obtain..)

Good reference data and algorithm can handle known problems, such as ArcGIS and ESRI StreetMap data

Solutions

- Instead of street, match by zipcode, or city centroid (see next)
 - Lower accuracy
- Parcel matching
 - Parcel data required (be aware: house numbering mismatch cases)
- Geocode with GPS (latitude & longitude information)
 - Locations were recorded with GPS – highest accuracy

Geocoding Process in ArcGIS

1. Build a geocoding service (a.k.a. geocoder, locator)
 - Define a method, reference data & fields, (quality) thresholds/criteria, and output data
2. Run the service against your table data
3. Manual intervention/corrections
 - Manually match unmatched cases that didn't satisfy the defined service thresholds/criteria
4. Try a lower-quality geocoding service to achieve a higher geocoding rate
 - [Optional: Use this option with caution]
 - Geocode based on second best spatial information
 - i.e. zipcode, instead of street addresses

Geocoding Variations

Reverse Geocoding (New!)

1. "I cannot geocode a location because there is no streets nearby, but I know exactly where it is."
 - We can assign geographic coordinates (or sometimes the closest street address) by clicking on a map .
2. "I know geographic coordinates but I want to know a street address for this location (or a name of the city or zipcode)."
 - By submitting geographic coordinates, some system can assign the closest street address, city and/or zipcode.

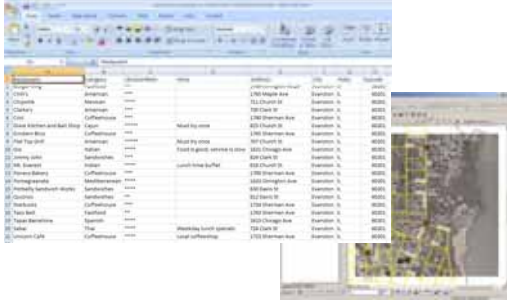
Reverse Geocoding
Pick the closest street address by clicking on a map



For more information:
http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=6407&pid=6402&topicname=Rematching_a_geocoded_feature_class

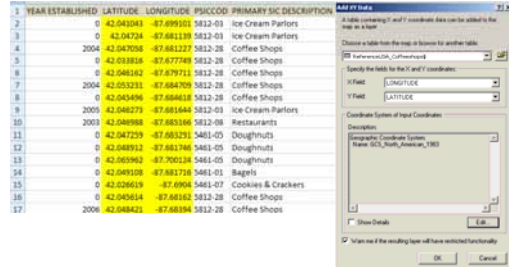
Exercise 7

- Let's geocode campus area restaurants -



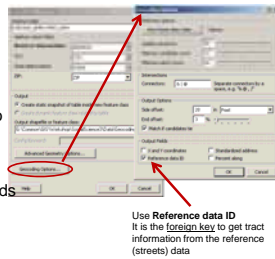
Exercise 7 (cont.)

- Adding X/Y data
 - Sometimes your data comes with handy X/Y (Longitude/Latitude geographic coordinate) information, such as GPS



But, wait –

- You don't know the real geocoding job yet, let me demonstrate it using real world example..
- Example:**
 - Employee residential data of a company A, located in Chicago
 - Total cases (N) = 24,158
- Tip to increase performance:**
 - Pre-sort your input by zone fields (i.e. state, city, zipcode)



Reference Data Sources

- The **traditional** geocoding services require:
 - Reference (geocodable streets) data
- Source options:
 - Census TIGER/Lines**
 - Free and readily available
 - 2000
 - 2008 and later (improved quality)
 - Data must be processed
 - The 2008 data needs to be processed to be used as a geocoding reference data source

Reference Data Sources (cont.)

- Source options:
 - TeleAtlas**
 - Geocode ready
 - Available from ESRI Data & Maps (free for us/NU)
 - Based on 2003 data
 - Also available from ESRI StreetMap **Premium**
 - Better/improved data
 - Updated often
 - Used by Google Maps
 - NavTec**
 - TeleAtlas's rival company
 - Used by Yahoo Maps, otherwise expensive to access

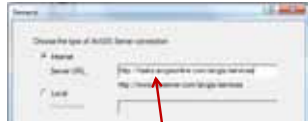
New: Online Services in ArcGIS

- Use of **pre-defined** online geocoding services in ArcGIS
- Pros – Good!**
 - No reference (street) data is necessary
- Cons – Bad!**
 - Not much control over the results – careful follow-up (i.e. quality check) would be necessary
 - These services are typically "composite" geocoders (i.e. they try to use the "best" spatial information but if the first choice doesn't work, they will use the second best information)

ESRI Online Locators

<http://www.esri.com/software/arcgis/arcgisonline/world-geocoding.html>

- Features:
 - Free (or subscription for better services)
 - Data updated frequently
- Type:
 - World Gazetteer
 - US and North American
- Use:
 - Find
 - Batch geocoding (up to 1,000 per session)



To add ESRI Online Locators, find and select GIS Servers > Add ArcGIS Server > Use GIS Services, and then add <http://tasks.arcgisonline.com/arcgis/services> in server URL

Bing Maps (i.e. Virtual Earth) Locator

<http://arcscripts.esri.com/details.asp?dbid=16446>

- Features:
 - Free (or subscription for better services)
 - Data updated frequently
- Type:
 - Address search
 - Business search
- Use:
 - Find
 - Batch geocoding



Geocoding Variations

- Zipcode/City matching
- "My data doesn't have street level address but has a zipcode field"
- We can match the data with zipcode points or polygons
- Two elements
 - Source file (i.e. data with zipcode)
 - Reference file (zipcode point or polygon data: either way, output will be points)
- Tools: ArcToolbox > Geocoding Tools
 1. Create Address Locator
 2. Geocode Addresses

Zipcode data matched with centroid of zipcode polygons

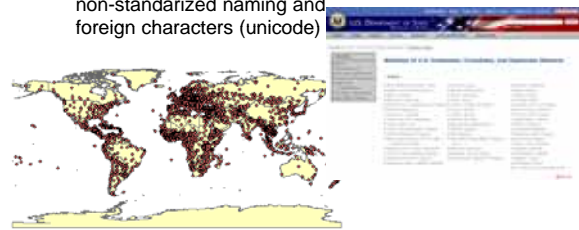


For more information:
<http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Commonly+used+address+locator+styles>

Exercise 8

- Geocoding of U.S. Embassies
 - Note: geocoding international places is not easy because of non-standardized naming and foreign characters (unicode)

Which is correct?
• Florence, Italy
• Firenze, Italy



Reference

- For more information: Geocoding in ArcGIS
 - ArcGIS North America Street Map
<http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName>About+StreetMap+North+America+data>
 - The best street database that the library has. A variety of pre-set geocoding options are included, such as street alias search.
 - Available to the NU affiliated members, contact Chieko Maene, [c:maene@northwestern.edu](mailto:maene@northwestern.edu)
 - Tutorial Video
<http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Geocoding+Tutorial>
 - PDF
http://webhelp.esri.com/arcgisdesktop/9.3/pdf/Geocoding_in_ArcGIS_Tutorial.pdf

Other Geocoding Tools

- ArcGIS is not the only option we have!
 - ESRI Address Coder
 - <http://www.esri.com/software/coder/>
 - Stand-alone geocoding tool
 - Expensive: \$9,000 +
 - Easily add extra data (demographics/marketing data) to your geocoded points



Other Geocoding Tools (cont.)

- Batch Geocoder
 - <http://www.batchgeocode.com>
 - Check output accuracy – I had some problems
- GPS Visualizer
 - <http://www.gpsvisualizer.com/geocoder>
 - Similar to batch geocoder
- Also, visit this near complete [list of geocoding tools!](https://webgis.usc.edu/Services/Geocode/GeocoderList.aspx)
<https://webgis.usc.edu/Services/Geocode/GeocoderList.aspx>