Geocoding - Advanced Techniques

Agatha Wong
Brad Niemand
Sergey Ivanenko
- Workflows on improving matching results
- Overview of the geocoding engine and matching process
- Fine tuning locators
- Road ahead for geocoding
What This Workshop Does Not Cover

- Basic geocoding process
  - Covered in the *Geocoding: An Introduction* Technical Workshop
- Programming with ArcObjects, Web APIs
- Address data model and tools

* Please meet with the geocoding team at the Spatial Analysis Island to discuss about these topics or issues you may have.
Improving matching results

One of the key questions when working with geocoding: How to match more addresses?

- Two workflows:
  - Create/manage composite locators
  - Create locators from manually matched addresses
Demo: Composite Locators

- A workflow using Geoprocessing model
Demo: Locators from Manual Matches

• Create point address locators from manual matches
Create locators from manual matches

**Assumption:** manual-pick addresses contain the true location and address information
Create locators from manual matches

Setup:

- Composite locator consisting of
  - Point locator
    - Built using address point feature class
    - Use US Address - Single House style
    - Set Minimum Match Score to a high value (95)
  - Street centerline locator
Create locators from manual matches

**Workflow:**

- After each rematch session
  - Select records from the geocoded feature class that are matched using the **Pick Address from Map** tool
  - Append records to the point feature class
  - Rebuild the **point** address locator
Create locators from manual matches

Sample workspace with the Geoprocessing model is available on the UC Proceedings media
and Esri Geocoding Resource Center
Overview of the geocoding engine at ArcGIS 10

- Geocoding process
- Grammar
- Data structure
- Aliases
- Scoring

Understanding the basic components of the geocoding engine can help matching addresses more effectively
Geocoding process

Unstructured input

100 fifth avenue ny

100 5th Ave, New York NY 10011

4600 Pennsylvania Ave NW, Wash. DC

380 New York St, Redlands CA 92373

1355 N Hrbr Drive, San Diego CA 92109

898 West Mission Bay Drive San Diego California 92109
Geocoding process

Unstructured input

100 fifth avenue ny

Structured reference data

House: 400, 100
PreDir: N
PreType: 5th
Name: New York
SufType: Ave
SufDir: St
City: Redlands, New York
Geocoding engine: grammar

- All supported forms of addresses explicitly defined in a grammar.
Grammar example:

```
Address: House StreetName City // 380 New York St Redlands
    | StreetName "&" StreetName City // Main St & 2nd Ave Springfield
    | SpatialOperator Address ; // 100 ft SW from 5 Main St San Diego

StreetName: PreDir PreType Name SufType SufDir ;

House: number // 123
    | number "-" number // 17-100
    | number letter ; // 100A

PreDir: "N" | "E" | "W" | "S" | "NW" | "SW" | "NE" | "SE" | ;

PreType: "Ave" | "Hwy" | ;

SufType: "Ave" | "St" | "Rd" | ;
...```

Handles ambiguous addresses
No "standardization"
Geocoding engine: data structure

- Locator stores a snapshot of reference data
- List of unique values for each reference data field
- Many-to-many relationship between fields

Locator stores a snapshot of reference data
- List of unique values for each reference data field
- Many-to-many relationship between fields

72 N Alabama St, Amarillo
Geocoding engine: data structure

Name

City
Geocoding engine: data structure

Address: House StreetName City ...
StreetName: PreDir PreType Name SufType SufDir ;

72 N Alabama St Amarillo

House PreDir PreType Name SufType SufDir City
11 72 100 302 400 520 1000 ...
N E S W
Ave 1st 2nd 5th
Albama Alpine Bacon Brockton Cimarron E New York ...
Zzyzx
Ave Ct Ln Pkwy Rd St N NE SE S W

Geocoding process is driven by both grammar and the reference data
Reference Data is an extension of the grammar

```
SufType: "Ave" | "St" | "Rd" | "Ct" | "Ln" | "Pkwy" | ;
```

<table>
<thead>
<tr>
<th>PreDir</th>
<th>PreType</th>
<th>Name</th>
<th>SufType</th>
<th>SufDir</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td></td>
<td>1st</td>
<td>Ave</td>
<td>N</td>
<td>Amarillo</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td></td>
<td>NE</td>
<td>Redlands</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5th</td>
<td></td>
<td>E</td>
<td>San Diego</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alabama</td>
<td></td>
<td></td>
<td>New York</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alpine</td>
<td></td>
<td></td>
<td>Zion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bacon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Brockton</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cimarron</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>E</td>
<td></td>
<td>S</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>New York</td>
<td></td>
<td>W</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zzyzx</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reference data help dealing with some local addressing oddities without the need for locator customization.
Geocoding engine: aliases

- Common abbreviated forms (aliases)
  - “Mt”, “Mtn”, “Mount”, “Mountain” mean the same in the context of a City name
  - Same for “View” vs. “Vw”
  - There are many ways to spell “Mountain View”
    - Mountain View
    - Mt View
    - Mtn View
    - Mount View
    - Mountain Vw
    - Mt Vw
    - Mtn Vw
    - Mount. Vw
Geocoding engine : scoring

- Grammar has information about relative importance of the components

Address: House \( \text{StreetName} \) \( \text{City} \)

<table>
<thead>
<tr>
<th>House</th>
<th>StreetName</th>
<th>City</th>
</tr>
</thead>
</table>

StreetName: PreDir \( \text{PreType} \) \( \text{Name} \) \( \text{SufType} \) \( \text{SufDir} \)

There is no penalty if some of the components (i.e. City, ZIP) are missing from the input address.
Geocoding engine: scoring (continued)

Address: 100 fifth avenue ny

\[\frac{2 \times 1.0 + 6 \times 0.95 + 3 \times 1.0}{2 + 6 + 3} = 0.97\]

House: 100

StreetName: fifth avenue

\[\frac{5 \times 0.0 + 6 \times 1.0 + 70 \times 1.0 + 7 \times 1.0 + 5 \times 1.0}{5 + 6 + 70 + 7 + 5} = 0.95\]

PreDir: 0.0

PreType: 1.0

Name: fifth (alias applied)

SufType: avenue (alias applied)

SufDir: 1.0

City: ny (alias applied)

House

PreDir

PreType

Name

SufType

SufDir

City

100

N

5th

Ave

New York
Geocoding engine : scoring (continued)

Address: 100 fifth avenue

\[
\frac{(2 \times 1. + 6 \times 0.95)}{(2 + 6)} = 0.96
\]

House: 100
StreetName: fifth avenue

\[
\frac{(5 \times 0. + 6 \times 1. + 70 \times 1. + 7 \times 1. + 5 \times 1.)}{(5 + 6 + 70 + 7 + 5)} = 0.95
\]

PreDir: 0.
PreType: 1.
Name: fifth
SufType: avenue
SufDir: 1.

(city applied)

House 100
PreDir N
PreType 5th
Name Ave
SufDir New York
ArcGIS 10 Geocoding engine summary

- Handling of ambiguous addresses
  - Example: 10 West Ridge Dr
- No penalty for missing zone information
- Can parse any input field, not just Street Name
- Returns matching address as it is written in the reference data
- Single line input
- Highly configurable
  - Single XML file
Fine tuning locators
Displaying Locator Contents on Internet Browser
Fine tuning locators (with XML Editor)

- Aliases

- Character mapping (spelling variations)
Demo: Find locations

- Spatial offset
- Dealing with challenging addresses
Customizing ArcGIS 10 Locators

- An Esri Geocoding Technical Paper
- http://resources.arcgis.com/gallery/file/geocoding
Road ahead for Geocoding at ArcGIS 10.1x

- Better geocoding performance
  - Batch matching
  - REST API
Road ahead for Geocoding at ArcGIS 10.1x

- Locator Package / Sharing via ArcGIS.com
Road ahead for Geocoding at ArcGIS 10.1x

- Geocode Services in the Cloud
Road ahead for Geocoding at ArcGIS 10.1x

- New Locator Properties user interface
Road ahead for Geocoding at ArcGIS 10.1x

• Enhanced searching
  - Search by extent
    - Within the current map extent
    - Return candidates within a bounding box

• Return candidates based on rank (simple rank, or population)
  Example:
  New York, NY
  New York, TX
Road ahead for Geocoding at ArcGIS 10.1x

- Global geocoding initiatives – international locator styles at Esri Resource Centers, and geocode services on ArcGIS.com.
Resources and References

- ESRI Resource Centers
  http://resources.arcgis.com

What is Geocoding?

Geocoding is the process of transforming a description of a location—such as an address or a place name—to a location on a map. You can geocode by entering one location description at a time or by providing many at once in a table. The resulting locations are output as geographic features with attributes, which can be used for mapping and spatial analysis.

See What’s new in geocoding in ArcGIS 10.
Additional Geocoding Sessions

• **Geocoding – An Introduction**
  - *Wednesday 1:30 PM, (Room 14B) – Offering II*

• **ESRI Showcase Software Island Demo Theater**
  - **Road Ahead – Geocoding** *(Tuesday 5:00 PM)*
  - **From a table of addresses to locations on the map** *(Thursday 11:30 AM)*

• **Visit the Spatial Analysis Island in the Exhibit Hall**
Questions?

&

Thank you for attending

Please fill out and submit your evaluation form

www.esri.com/sessionevals