PAGC: Recent Advances and Future Developments

Dan Putler and Walter Sinclair, Anemoi Analytics
Stephen Woodbridge, iMaptools
Agenda

- About PAGC and a little about its history
- An overview of what PAGC provides
- A glimpse under the covers of how PAGC works
- The current roadmap for future development
- Examples of how PAGC is currently being used
- Q&A
The PAGC Elevator Pitch 1

• What is PAGC?

The Postal Address Geocoder (or PAGC) is an open source library and web service framework for geocoding locations based on either postal addresses, street intersections, parcels, or landmarks. The system offers a number of features that differentiate it from other open source geocoding software solutions.
The PAGC Elevator Pitch 2

• Why should you care?

If you need a tool that takes a textual description of a location, such as an address, landmark, or parcel, PAGC can look that description up in a reference database in order to produce a location (a latitude and longitude), then this is a tool that will provide that service. As a library this can be linked directly into other source code packages to create a seamless integration. As a webservice, you have the flexibility to make requests from a variety of platforms via a simple HTTP request.
How to Get PAGC

Project Site:  
http://www.pagcgeo.org

SourceForge Page (the path to SVN):  
http://sourceforge.net/projects/pagc/
PAGC – Background

- Development of PAGC began in 2000
- **Objective**: Create a *command line* geocoding program with functionality equivalent to the ArcView 3 geocoder, but make it cross platform
- PAGC becomes an open source project in 2002
- Starting in 2006 PAGC started to transition from a standalone program to a library based suite
Key Features 1

- PAGC is written in ANSI C, currently runs on all major operating systems, and can be easily ported to nearly all operating systems.
- PAGC is pre-configured to use data from several different providers (TIGER/Line, Statistics Canada's Road Network File), and through the creation of a simple configuration file, it can easily work with data from a large number of other providers.
Key Features 2

- PAGC makes use of an advanced three-method, probabilistic matching algorithm to provide the greatest possible match rate for street addresses
  - Exact match
  - Soundex
  - Pointer-less trie combined with an edit distance measure
- PAGC can be used with multiple input data sources
  - Shapefiles
  - SQLite/SpatiaLite database tables
  - PostgreSQL/PostGIS database tables
Key Features 3

- PAGC works with multiple data stores
  - BerkeleyDB
  - SQLite
  - PostgreSQL (under development)
- PAGC has the ability to provide an ID number in addition to geographic coordinates, which can greatly simplify database operations for location-based applications
Key Features 4

- PAGC allows the user to edit road segment address ranges on the fly, enabling the user to overcome suppressed or altered address ranges frequently encountered in government originated road network layers.

- PAGC as a web service allows you to pass an address parsed into fields or as a single text field that will be parsed into fields internally, giving you a lot of control over your requests.
What PAGC Does Under the Covers

**Build:** Create a standardized back-end database of street segments, property parcels, and/or landmarks that can quickly be searched.

**Match:** Take user input, standardizes it in a way that is compatible with the back-end database, search the database, and return the search results to the front-end software.
PAGC's Software Components 1

- **libpagc**: The underlying, API-based workhorse
- **libds**: A data store abstraction layer that allows PAGC to work with both raw and standardized street segments, property parcels, and landmarks in several different underlying formats
  - Shapefile/DBF/BerkeleyDB
  - SpatiaLite/SQLite
  - PostGIS/PostgreSQL (in development)
PAGC's Software Components 2

- Front-end geocoding software to obtain user input and interaction
  - The web geocoder service
  - Command line tool for bulk geocoding, user interaction, and “on the fly” editing of the underlying road segment database

- Front-end build tool for configuring the standardization and loading data into the back-end database
Other PAGC Components

- Standardization files
  - rules.txt
  - lexicon.csv
  - featwords.csv
  - gazeteer.csv
The Build Phase

Raw Data Source
- Shapefile
- SpatiaLite table
  - *PostGIS table*

Schema File
- DBF
- SQLite table

Front-End Build Tool
libds

libpargc

libds

Back-End Data Store
- BerkeleyDB
- SQLite
  - *PostgreSQL*

Note: *Italics* indicate that use of this format is under development

Postal Address Geo-Coder
The Schema (Configuration) File

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTRIB, C, 13</td>
<td>COMPARE, C, 26</td>
<td>NAME1, C, 10</td>
<td>NAME2, C,7</td>
<td>NAME3, C,9</td>
<td>NAME4, C,7</td>
</tr>
<tr>
<td>HOUSE</td>
<td>NUMBER_INTERVAL_LEFT_RIGHT</td>
<td>LFROMADDR</td>
<td>LTOADDR</td>
<td>RFROMADDR</td>
<td>RTOADDR</td>
</tr>
<tr>
<td>PREDIR</td>
<td>CHAR_SINGLE</td>
<td>PREDIRABRV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRETYP</td>
<td>CHAR_SINGLE</td>
<td>PRETYPABRV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STREET</td>
<td>CHAR_SINGLE</td>
<td>NAME</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUFDIR</td>
<td>CHAR_SINGLE</td>
<td>SUFDIRABRV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUFTYP</td>
<td>CHAR_SINGLE</td>
<td>SUFTYPABRV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUALIF</td>
<td>CHAR_SINGLE</td>
<td>QUALIFIER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CITY</td>
<td>ALT_CHAR_LEFT_RIGHT</td>
<td>PPLACE</td>
<td>PPLACER</td>
<td>CPLACEL</td>
<td>CPLACER</td>
</tr>
<tr>
<td>PROV</td>
<td>CHAR_LEFT_RIGHT</td>
<td>STATEL</td>
<td>STATER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POSTAL</td>
<td>POSTAL_LEFT_RIGHT</td>
<td>ZIPL</td>
<td>ZIPR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOURCEID</td>
<td>NO_COMPARISON</td>
<td>TLID</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GEOMETRY</td>
<td>SPATIALITE_WKB</td>
<td>the_geom</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Building PAGC Standardized Data

• Analyze your data and decide on a schema
• Create a schema (configuration) file
• Join the data if needed
  ○ tiger_street_join utility
  ○ SELECT INTO statements for SpatiaLite tables
• Run the pagc_build_schema utility to standardize and load the records
The Match Phase

Front-End Geocoding Software
- Web geocoder service
- Command line tool
- Third-party tools

Back-End Data Store
- BerkeleyDB
- SQLite
- PostgreSQL

Note: *Italicics* indicate that use of this format is under development
The PAGC Roadmap 1

• Determine ways to improve performance with large data sets
• Revitalize the command line tools for bulk geocoding
• Simplify the build and setup process
• Better support for languages other than English
  ○ UTF8 support if/as needed
  ○ Support for country specific gazetteers, abbreviations, etc
The PAGC Roadmap 2

- Improve user and administrator documentation
- Provide language binding for libpagc
  - Python
  - Java
- Library changes to support embedding PAGC into a databases systems such as PostgreSQL
Metropolitan Mosquito Control District

Map of an area with a marker at 600 Country Trail East, Jordan, 55352.

Address: 600 Country Trail East
City: Jordan
Zip Code: 55352

Or search by intersection. Both street names are required for an intersection search.
Street 1:
Street 2:

Or search by Point of Interest/Landmark name, such as a park, school, or lake.
Place Name:

Postal Address Geo-Coder
Anemoi Analytics Prototype
MetroGIS Geocoder for ArcGIS

Geocoder Tool for ArcMap

The MetroGIS Geocoder Tool for ArcMap was created so users could easily use the MetroGIS Geocoder Web Service within ESRi's ArcMap 9.3 Software. The tool allows users to enter an address, geocode it using the MetroGIS Geocoder Web Service, display the results and zoom ArcMap to each selected result. This tool is currently only compatible with ArcMap 9.3.
Thanks to Our Sponsors

- MetroGIS
- Social Sciences and Humanities Council of Canada

---

• MetroGIS and the Metropolitan Mosquito Control District

• Social Sciences and Humanities Council of Canada
Questions?

The PAGC Project
http://www.pagcgeo.org

Anemoi Analytics
http://www.anemoianalytics.com

iMaptools
http://www.imaptools.com