

Walter Sinclair
3715 West Tenth Avenue
Vancouver, BC
V6R 2G5
Telephone: 604-228-8223
E-mail: whitedwarf@deadwrite.com

Proposal for Support of Landmarks in the PAGC Geocoding Library and Geocoding Web Service

Submitted to: The PAGC Developer's Group
Date: March 15, 2009

OVERVIEW

It has been suggested that support for landmarks (or points of interest) would be a useful addition to the PAGC library and geocoding web service. An implementation of this support is proposed here.

PROPOSAL

It is proposed that the PAGC library and webservice software be expanded to incorporate support for landmarks. Landmarks (or points of interest) are sites identified by name, rather than by a number and street address. The geocoder, so expanded, would accept the name and type, a restriction to a particular area, and return scored candidates, each with latitude and longitude – and the site address if available.

To do this the library software will be expanded to identify, match and score on new fields -- fields not currently utilized in address geocoding. The geocoding web service will be expanded to handle a landmark request, returning data from these (and other) fields in a manner consistent with the way it now handles intersections and site addresses.

Landmark Request

The geocoding webservice would accept a LandmarkSite request consisting of

- a) LandmarkName (used in matching/scoring)
- b) FeatureType (used in matching/scoring)
- c) CountyName (used in matching/scoring)
- d) CityName (used in matching/scoring)
- e) Zip/Postal Code (used in matching/scoring)
- f) State/Province Name (used in matching/scoring)
- g) MethodName, Version, CountryCode, MaximumResponses, ResponseFormat (as with current requests)

This request is passed to the PAGC library, which standardizes, matches, scores, and returns to the geocoding service a list of scored candidates. The geocoder returns that list, suitably formatted, to the requester.

Landmark Response

Each candidate returned, in addition to fields representing the dataSource field and id, will have geographic position and score, standardized or official name values corresponding to the

requested fields, as well as the Address Number and Street verbatim, if available. The Address data returned is not used in matching or scoring. The presentation and packaging of the response will be consistent with that now employed for site and intersection responses.

Landmark Data and Processing

The PAGC library and builder (`pagc_build_schema`) will need to be modified to support the changes required handle this new, non-address schema type. New configuration flags are needed to identify the fields, named here to correspond with the draft Street Address Data Standard: `LandmarkName` (SAD-2nd 1.7.4) , `CountyName` (SADS-2nd 1.7.5.4), `FeatureType` (SADS-2nd 1.8.3.2). The `LandmarkName` will be stored in two forms, the official name and the standardized name, but only the official name returned. The `FeatureType` will be for this version stored and returned as just a standard code. The `CountyName` will be stored and returned as official name only.

A dataset for Landmarks will contain, at a minimum, the `LandmarkName`, and may contain other address attributes. However only those indicated will be used for scoring and matching.

It will be necessary for the library (accessed through `pagc_build_schema`) to create an internal record with fields for each landmark site, and indices for approximate, soundex and regular searches.

The standardizer for the landmark name may require an additional standardization file, similar to `lexicon.csv`, or may employ the current lexicons. This will be determined as the standardizer evolves. Many changes to the standardizer in particular are expected, due to the difference in nature between a site or intersection address and a landmark name.

New library routines will need to be written to perform the different kind of standardization required for the landmark name, to handle the building of the landmark name records and indices, to handle the searching, matching, scoring and formatting for the response.

Responder

The responder will need to be expanded to handle the new elements of the request and the response. It will also need to handle multiple reference datasets. This it will do by conducting an ordered search on the set of datasets. In other words, search dataset 1 and if score is not high enough, search dataset 2 etc. This is basically what we are currently doing with precise and interpolated site addresses, but here it will be with the same geocoding (precise) in each case.

Other Changes

New Documentation will be necessary for these new features. The library interface and configurations will change, expanding to handle landmark requests.

International Note

Some field names proposed, `CountyName` for instance, may appear to be unduly US-centric. It should be noted that the field names used and returned are specified – and can be different -- for each reference schema. The concept of a county – as a district name somewhere between city and province – can be applied to many environments outside of the United States.

It should also be noted here that some of the functionality that would be introduced here would also be useful in environments where name rather than number is the more significant identifier in a site address.

Note on Data Editing

One feature requested for inclusion in PAGC's handling of landmarks is the ability to do online edits of the data. This feature is not included in this proposal. To facilitate software that supports this feature, however, the current build-once philosophy behind the building of standardized reference coverages will need to be abandoned. This will require the addition of modification functions, functions that allow addition of records or sets of records after the initial build.

Workload

This work will require major changes to the handling of standardization, schemata, and scoring and matching. It is estimated that this work will require in the neighbourhood of 60 hours of programming spread over a 3-4 month period. The work itself will require funding to complete, but an outline of budget requirements is not part of this proposal.