Avalara Geo for Communications

Getting Started Manual

Avalara for Communications - Contact Information	
Address	Avalara, Inc.
	513 South Mangum Street, Suite 100
	Durham, North Carolina, 27701
Toll Free	800-525-8175
Corporate Site	http://communications.avalara.com/
Comms Platform Site	https://communications.avalara.net
Email	communicationsupport@avalara.com

Document Revision History

The Revision History log lists the date and description of the most recent revisions or versions of the document.

Date	Version	Description
03/24/2016	0004	Avalara branding updates to reflect the transition to the new company and product names have been incorporated into this document. Please see **Appendix A – Avalara Product Names** for specific changes in product references and descriptions.
05/13/2016	0005	Updates throughout document. Most significant updates in **Section 2 Address Geocoding**, **Section 3 AFC Geo Best Practices**, new **Section 4 AFC Geo** Products; a table of endpoints added to **Section 3.7 Endpoints.**
10/31/2017	0006	Updated Avalara contact information (address and support site). Removed **Appendix A – Avalara Product Names**.

Table of Contents

1. What is AFC Geo?

Avalara Geo for Communications (AFC Geo) is a system that provides methods to verify addresses and coordinates and lookup their tax jurisdiction. Tax rates can sometimes vary based on an area of a city, and not just apply universally to the entire city.   AFC Geo is used to identify the taxing jurisdiction that is specific to an exact location.

2. Address Geocoding

Address geocoding is the process of mapping an address to the street network, i.e., returning a location. The input into this process is a set of one or more strings, as well as other input parameters that define options on how to process the address.

Please note that AFC Geo functionality does not support geocoding for an address which includes a Post Office (PO) Box because they are not considered acceptable by tax jurisdictions for tax situsing purposes. An address which starts with a numeric street address must be entered.

Note: AFC Geo only supports geocoding for US, Puerto Rico and other US territory addresses. All addresses in foreign countries (**including Canada and Mexico**) are not supported and will not return any results.

2.1   Parsing

A US address, for example, can be represented in the following ways:

- A first line, city, state, and zip code, e.g.:

16 Tech Circle	Natick	MA	01760

- A first and second line:

16 Tech Circle	Natick, MA 01760

- A single line:

16 Tech Circle, 01760

The parser produces one or more interpretations of the input fields, in the form of associating values with database keys. For example, parsing 16 Tech Circle, 01760 as a single-line address will produce the following key/value combination:

Key	Value
Street Number	16
Street Name	Tech Circle
Secondary Unit	N/A
City Name	N/A
State Abbreviation	N/A
Postal Code	01760

AFC Geo uses a parser, which is capable of producing multiple interpretations of the ambiguous input text. This is different from parsers used in other geocoders, which can only produce one parsing for the given text. For example, parsing 15 GA HWY 21,ATLANTA,GA,12345 will produce the following two interpretations:

Key	Interpretation 1	Interpretation 2
Street Number	15	15
Street Name	GA HWY 21	GA HWY
Secondary Unit	N/A	21
City Name	Atlanta	Atlanta
State Abbreviation	GA	GA
Postal Code	12345	12345

Indeed, the input address could be interpreted either as number 15 on Georgia Highway 21, or as number 15 on Georgia Highway, apartment 21. Both of these interpretations deserve consideration.

2.2   Matching

Next, each of the interpretations is used to look up candidate locations in the street database. This process is similar to a traditional database query. At this stage, some of the keys are treated in a "relaxed" manner, in that a strict match is not required to produce a candidate. For example, streets, whose name

sounds like the input street name, are considered candidates. Therefore, 1 Renee Street, 02134 will match 1 Rena Street, 02134. The algorithm used for matching the sounds is proprietary and is much more powerful than the popular, but relatively primitive, SOUNDEX algorithm.

Note that certain typos may result in street names that do not sound like the correct name. In this case, AFC Geo may not find a match.

If a street has alternate names (aliases) in the database, they will all be considered.

The street number is matched against the address ranges supplied in the vendor's street database. A single street is represented in the street database as a chain of segments called street links. A link is a small segment of the street, usually between two intersections, but sometimes a single block consists of multiple links. Each link is characterized by a range of street numbers on either of its sides. For example, the left side may have the street number range of 2 .. 48, and the right side 1 .. 47. A precise match occurs when the street number supplied for geocoding falls within the address range of a link. For example, if the street number supplied in the above example is 24, then AFC Geo assumes that the location is approximately in the middle of the link on the left side, and computes the geographic coordinates accordingly.

2.3   City + State vs. Postal Code

In addition to the mandatory street address, i.e., the house number and street name, a valid address must contain information about the locality. In the US, this information is represented by the city name, state abbreviation, and/or postal code. In many cases, an address can be unambiguously geocoded with only city and state, or only the zip code. In some cases, the client application does not supply a zip code, or the city name, etc. Here is how AFC Geo approaches these situations:

- If the postal code is supplied, it is used for matching.
- If both the city and state are supplied, then city and state are also and independently used for matching.
- The two sets of results are merged.

An important implication of this approach is that if *the right city and state, but the wrong postal code are supplied, or vice versa*, the match will still succeed. There are implications for scoring, in that a 100% score will not be achieved, but the correct location will be returned (unless it falls under the scoring threshold).

One way to look at this process is as producing multiple interpretations, just like during the parsing stage. This, of course, has nothing to do with parsing, and everything to do with the virtual redundancy of city/state versus postal code.

For instance, the address 16 Tech Cir, Natick, MA, 12345 (wrong zip code) will produce the following two interpretations:

Key	Interpretation 1	Interpretation 2
Street Number	16	16
Street Name	Tech Cir	Tech Cir
City Name	Natick	N/A
State Abbreviation	MA	N/A
Postal Code	N/A	12345
Notes	Will match a location in postal code 01760	Will not match any locations

When the results are merged, the one correct location will be returned. If the two interpretations have each produced some results and all matches have been requested, then both sets of results will be scored and returned.

2.4   Fallback Modes

Some of the keys are normally treated in a strict manner. However, if the query fails to produce a match, AFC Geo enters the so-called fallback mode. The fallback mode can be thought of as relaxing certain criteria. There are two modes that can be independently turned on or off by the client application.

Each of the fallback modes results in a slight decrease of the performance of the query, which is why the application developer is given control over this behavior.

The most useful fallback mode is called *Street Number Snapping*. Either an erroneously entered street number, or outdated street data can result in a situation when none of the address ranges contains the supplied street number. When the Street Number Snapping mode is on, AFC Geo finds the link with the address range closest to the supplied number, and returns that link. The location is "snapped" to the end of the link with the closest street number. For example, geocoding 200 Tech Cir, 01760 *with Street Number Snapping on* will fall back to 98 Tech Cir, 01760, because 98 is the largest even street number present in the data set.

When the snapping is off, no match will be generated. There are links in the data set that do not have an address range assigned. This happens usually when the new street is digitized, but no houses have been built. Later, even when the street is populated, there may be a lag before the address ranges appear in the street data set. Links without address ranges are approached according to the following rules:

- Links without address ranges are not considered unless Street Number Snapping is on.
- If there is even a single street link with an address range, it is preferred to unassigned links.
- Certain streets, especially in new developments, may not have *any* address ranges assigned. Obviously, in this situation it is impossible to return *the closest address range*. In this

situation, AFC Geo will return an arbitrary segment on the street, to at least get you in the neighborhood.

The other supported fallback is called *Postal Centroid*. It involves reverting to the centroid of a postal code area, when:

- Geocoding failed to produce any match at all;
- The postal code was supplied in the input address;
- The street data set contains at least one street in this postal code.

The centroid is a very coarse approximation of the real location.

2.5   Scoring

Finally, all candidate locations are scored according to how well database values match specified input values. The score of 1.0, or 100%, results from an exact match between the inputs and the actual values. The score of 0.0 is a theoretical score that would be produced if nothing matched at all, although such locations never make it to the scoring phase.

An important point here is that the scoring is applied only to the keys that are specified by the calling application. The user is never "punished" by score for *not supplying* a certain key, such as the zip code. For instance, geocoding 16 Tech Cir, Natick, MA succeeds with a 100% score (even though the zip code is not specified). However, if the user does supply a value, then the scoring engine considers it. For instance, geocoding 16 Tech Cir, Natick, MA, 01700 (the zip code is incorrect here) succeeds with a 76% score.

The names, such as city and street names, are matched based on the algorithm known as Edit Distance. Edit Distance is essentially the minimum number of single-character changes, removals or additions required to convert one string into another. The more typos one makes, the larger the edit distance between the typed and the perfect values. The edit distance of 0 produces the perfect score, and so on. For instance, geocoding 1 Great Plaine ave, 02492 (the correct name is Great Plain) succeeds with an 84% score, and 1 Great Plane ave, 02492 results in an even worse, 79% score. Note, by the way, that these reasonable typos result in a relatively small decrease of the score.

The combined score for the location is computed as a weighted average of the individual scores by key.

3. AFC Geo Best Practices

This section contains a list of best practices for using AFC Geo recommended by Avalara, Inc. Each organization should carefully evaluate its own business requirements when determining how to use AFC Geo with its applications.

3.1 Minimum Score or Scoring Threshold

The minimum score is the threshold above which an address is considered a matching address. The minimum score should not be defaulted to zero, because that could allow matches that are not at all close to the desired input address. Avalara, Inc. recommends not using addresses with a score of less than 70%. This would allow for a reasonable number of typos when entering the address. This scoring requirement may be increased or decreased to allow for a higher or lower level accuracy, which may vary depending on the quality of the input data.

3.2 Street Number Snapping

Avalara, Inc. recommends that clients do not use street number snapping (Options value 1) as it almost always results in returning an incorrect street number. In many applications it may be acceptable to enable street number snapping to get the tax jurisdiction (FIPS Code or PCODE) and use the original street number. This is accurate unless there is a long section of the street with no address range assigned.

3.3 Fallback to Postal Centroid

The postal code centroid fallback (Options value 2) is also not recommended by Avalara, Inc. because the address returned will have no relationship to the input address other than they will both have the same zip code. In addition to an invalid address, there are also several cases where the tax jurisdiction returned will also be incorrect.

3.4 Special Tax Districts

It is highly recommended to set the option to return special tax districts (Options value 16). Special tax districts are applicable for taxes associated with fire, police, ambulance, library, roads, economic development, townships, etc. Without this option turned on, jurisdictions for only city or county will be returned.

3.5   Numeric Street Number

In most cases, an address will not geocode unless the street address starts with a numeric street number. For example, "100 Main Street" is acceptable, but "PO Box 100" will not produce a valid result. A letter as a suffix is often allowed, but a letter prefix is typically not allowed. For example, "100B Main" could be OK, but "B100 Main" is not. However, there are a few states that allow letters to be included within the street address, but those addresses may not be recognized as having a valid street number. Therefore, whenever possible, it is best to always have a numeric street number.

3.6   CASS Validation

CASS Validation (Coding Accuracy Support System) is a process by which a USPS database is used to ensure that the geocoded address is as accurate as possible. Therefore, it is a best practice to use this option to increase the accuracy of the results. Please note; however, that if the address is in Florida, CASS Validation will always be performed, no matter what the value of the input flag.

3.7   Endpoints

AFC Geo SaaS Pro has two sets of endpoints. One set returns a NULL when there is a problem within the geocoding process. The other set of endpoints (ending in /2.0) return exceptions. It is a best practice to use the endpoints that return exceptions and to handle them within your code. Please reference the table below for examples of endpoints.

Binding	URL
Basic HTTP	https://ezgeoasp.billsoft.com/LocatorService.svc/2.0
	https://ezgeoasp.billsoft.com/LocatorService.svc
Custom	https://ezgeoasp.billsoft.com/LocatorService.svc/SSL/2.0
	https://ezgeoasp.billsoft.com/LocatorService.svc/SSL
WS HTTP	http://ezgeoasp.billsoft.com/LocatorService.svc/2.0
	http://ezgeoasp.billsoft.com/LocatorService.svc

4. AFC Geo Products

There are three different product offerings that are built around AFC Geo functionality. They all use the same core geocoding functions, but provide different types of access. The different products are as follows.

- AFC Geo SaaS Pro
- AFC Geo SaaS Standard
- AFC Geo Direct

AFC Geo SaaS Pro provides a SOAP message interface, where client's applications can geocode individual addresses in real time.

AFC Geo SaaS Standard provides an FTP batch mode, where clients can transfer a comma separated value (CSV) file to a server and have an entire set geocoded to produce a results file.

AFC Geo Direct provides a web browser interface that allows a user to manually geocode a single address.

AFC Geo License provides on sight functionality to directly run the geocoding process on the client's computer.

The most common and preferred product offering is the AFC Geo SaaS Pro version. For additional details on each of these products, refer to the AFC Geo Product User Manual.