

# Postal Codes, Polygons, Perl and Poetry

.. and finding the correct location from some text

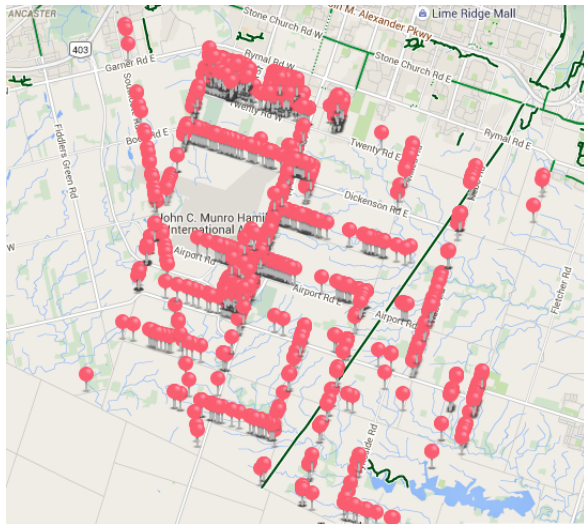
# Postal Addresses in Canada

Street Number, Street Name, City, Province  
Followed by:

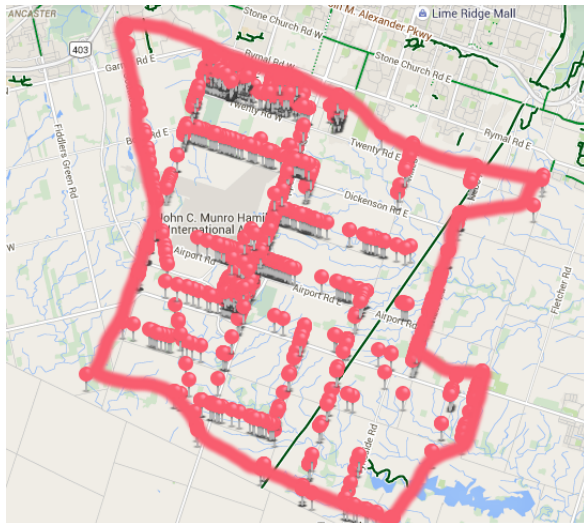
A                      N                      A                      N                      A                      N

K                      2                      C                      1                      N                      5

# A Canadian Postal Code Polygon



# A Canadian Postal Code Polygon Approx



Wait a sec!

Where did all these points come from?

**Forward Geocoding:** | [Reverse Geocoding](#) | [Geocode Street Intersections](#) | [Free Data Downloads](#) | [Legal News](#) | [Development N](#)

### Location (in Canada or USA):

124 Provident, MOUNT HOPE, ON L0R1W0

### GeoCode

| [Geocoding APIs and Tools: XML, CSV, Jsonp and Json APIs](#) | [Batch Geocoding & Geo-Data Visualization](#).

The *Location* field must be a location in North America (USA or Canada) containing a street address, a postal or zip code, a street intersection, a landmark, ip address, a city, a point (latitude,-longitude) or a combination of these location entities.

 [Geocode on a Map](#) / Geocode your ip address: [104.247.231.15](#)

# The Crowd!

Geocoder.ca

Services | Products | Solutions

Terms

Login

Create Account

API

Contact

Beamsville, ON » **124 Provident, Beamsville, ON** » L0R1W0 » [43.314074,-79.627722 Directions](#)

124 Provident, Beamsville, ON L0R1W0 ( [L0R1W0 polygon](#) | [L0R1W0 info](#) ) [Directions](#) [Reverse Geocode](#)

Confidence Score: 0.9

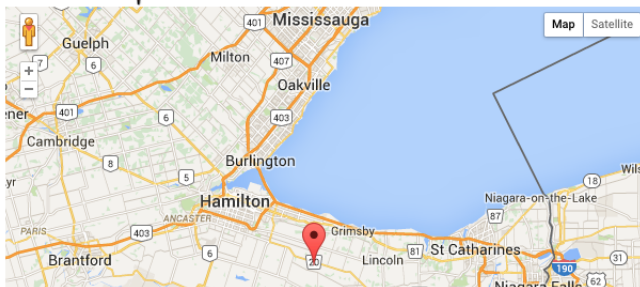
Is the location shown in the Map incorrect?

Drag the marker to correct this location

Then click here to send your corrections

**43.314074, -79.627722**

[Geocode this Location on a Map](#)



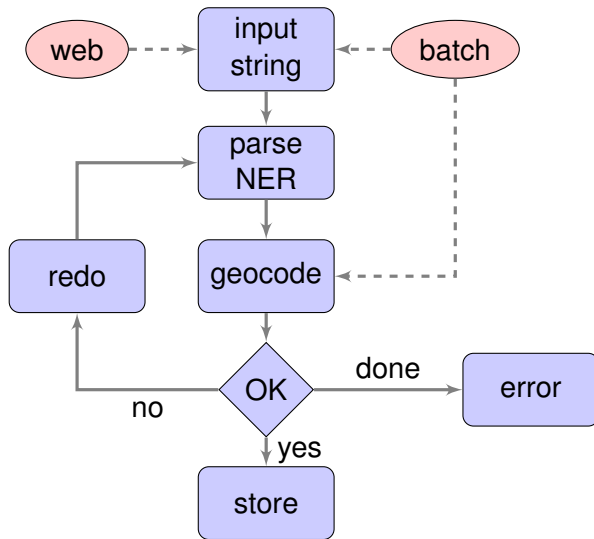
Nearby:

- [Beamsville Restaurants](#)

[View Street Addresses in the L0R1W0 p](#)

[code](#) Note: These street addresses have been su  
geocoder.ca in association with the L0R1W0 post  
Beamsville, ON. They are provided "as is", none h  
verified by geocoder.ca nor does geocoder.ca cla  
a complete list of all existing addresses in the L0  
postal code area.

# The Process. Step 1: Geocoding

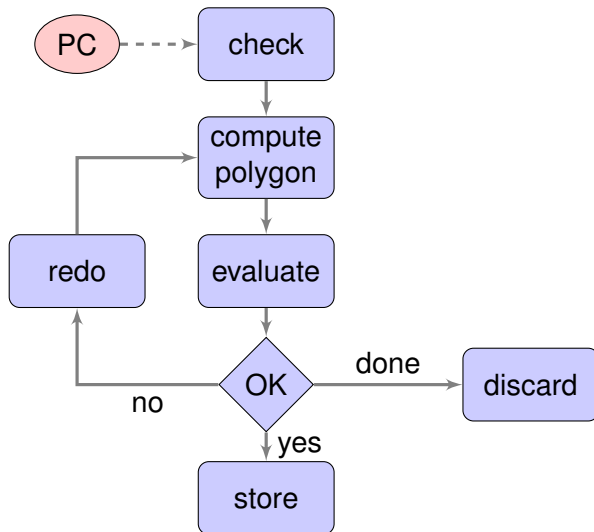




# The Crowd

Main Assumption: The Majority is right!  
Safe bet: A certain minority is wrong!

## The Process. Step 2: Analyzing



# Various assumptions

'Similar' postal codes are located close together. (K2C 1N5, K2C 1N4, etc...). If further than a certain "distance", discard.

K2C1N5

K2C1N

K2C1

K2C

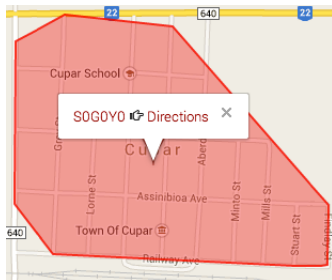
Need to check if a new postal code is processed in the system.

# Limitations

- 1 No way to tell if a certain postal code is no longer valid.
- 2 How accurate can crowdsourcing be?

# A sample postal code polygon

A new postal code can either expand a given polygon, create a new one (or rather a point), or leave the polygon unchanged. A postal code polygon must not intersect with another, so if a new entry attempts to do that, discard.



Either an alpha shape (if enough points) or simple convex polygon

# Numbers on Accuracy (courtesy of Denis Carriere)

COUNT (Geocodes)

=====

Osm: 29141

Esri: 40382

Mapquest: 39413

Bing: 40934

Google: 39583

Geolytica: 40158

Nokia: 40938

(out of about 41,999 verified locations)

<http://datahub.io/organization/addxy>

# Geocoding Test Methodology

- 1 Get data from verified sources (municipality data manually geocoded rooftop:  
<http://open.canada.ca/en/maps/open-data-canada>
- 2 Geocode the data on each provider and compare the result to the verified result

# Numbers on Accuracy (courtesy of Denis Carriere)

AVERAGE ERROR

=====

Osm: 347m

Esri: 39m

Mapquest: 76m

Bing: 80m

Google: 87m

Geolytica: 187m

Nokia: 77m

<http://datahub.io/organization/addxy>



# Numbers on Accuracy (courtesy of Denis Carriere)

CRITICAL ERROR (+1km)

=====

[5.04%] Osm: 1469

[0.06%] Esri: 28

[0.26%] Mapquest: 106

[0.22%] Bing: 93

[0.70%] Google: 278

[2.08%] Geolytica: 836

[0.22%] Nokia: 92

<http://datahub.io/organization/addxy>

# Numbers on Accuracy

MARGINAL ERROR (+250m)

=====

[6.26%] Osm: 1826

[1.27%] Esri: 515

[2.99%] Mapquest: 1180

[1.13%] Bing: 466

[3.36%] Google: 1330

[7.76%] Geolytica: 3117

[1.12%] Nokia: 461

<http://datahub.io/organization/addxy>

# Is the crowdsourced postal code db accurate enough?

Good Question.

# About Me and Perl

- 1 I do not like to change the world, only what I do not like.
- 2 I like Perl.
- 3 I am not a very good perl programmer
- 4 But I still can use it to solve complex problems
- 0 That is why I like Perl.

# Why go into all this trouble?

- 1 I do not like getting lost.
- 2 I visited Barcelona last month.
- 3 I stayed at two Airbnb's
- 4 Both times I was sent to the wrong location by Google Maps
- 0 Is there more than one way to do it?

# Carrer de Colomines, 2 Entresuelo 200AA, Barcelona, Catalunya 08003, Spain

The screenshot shows a mobile application interface. At the top, a search bar contains the address "Carrer de Colomines, 2 Entresuelo 2ª, Barcelona, C" with a search icon to its right. Below the search bar, a dropdown menu displays "Search nearby: hotels · restaurants". The main content area features a card for a location named "Ank". This card includes the address "Bruc 19, Entresuelo 2ª 08010 Barcelona Spain", a phone number "+34 717 12 75 79", and buttons for "Directions" and "Save". A "Send to device" link is also present. Below this card, an advertisement for "Book a room" is displayed. It shows a check-in date of "Thu, Aug 20" and a check-out date of "Fri, Aug 21". Two booking options are listed: "Booking.com" and "BudgetPlaces.com", both offering a rate of "\$104 / night". Each option has a "Book" button. At the bottom of the ad, there is a link to "View 1 more booking option at \$103". The background of the interface is a map showing the location of "Ank" on a street grid in Barcelona, with labels for "Carrer de les Corts Catalanes", "Carrer del B", and "Carrer d'Au".

Carrer de Colomines, 2 Entresuelo 2ª, Barcelona, C

Search nearby: hotels · restaurants

**Ank**  
Bruc 19, Entresuelo 2ª  
08010 Barcelona  
Spain  
+34 717 12 75 79  
Send to device

Directions Save

**Ad Book a room**

Check-in  Check-out

Booking.com \$104 / night **Book**

BudgetPlaces.com \$104 / night **Book**

View 1 more booking option at \$103

# http://Geocode.xyz - A geocoder for Spain

GeoCode.xyz

API


BARCELONA, ES » 2 COLOMINES, BARCELONA, ES » [41.3855238000,2.1789661000](#) [Directions](#)

2 COLOMINES Calle, BARCELONA, ES ( [BARCELONA,ES polygon](#) ) [Directions](#) [Reverse Geocode](#)

Confidence Score: 0.

Is the location shown in the Map incorrect? 

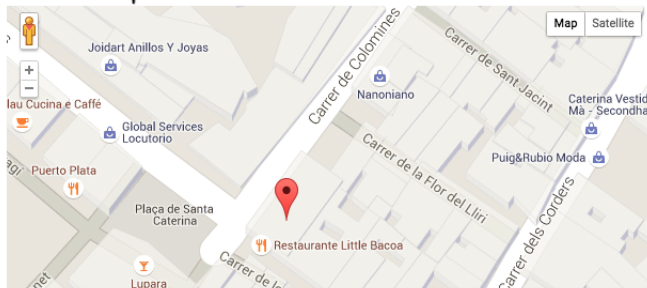
Then click here to send your corrections

Drag the marker  to correct this location

41.3855238000, 2.1789661000



[Geocode this Location on a Map](#)



# Data sources

An aerial photograph of a dense city skyline, likely New York City, serves as the background. Overlaid on this is a semi-transparent map of Spain, which is filled with a dense pattern of red dots representing address locations. The map includes labels for 'Vigo', 'Bilbao', 'ANDORRA', 'SPAIN', 'Valencia / València', 'Córdoba', 'PORTUGAL', 'Balearic Sea', and 'Mediterranean Sea'. In the top right corner of the slide, there is a small circular logo with a green and white checkmark. Centered over the map is the 'OpenAddresses' logo, which consists of a stylized white outline of a city skyline with the text 'OpenAddresses' in white. Below the logo, the text 'The free and open global address collection' is written in a large, bold, white font.

OpenAddresses

**The free and open global address collection**



# Spanish Data License

```
{  
  "coverage": {  
    "ISO 3166": {"alpha2": "ES", "country": "Spain"},  
    "country": "es"  
  },  
  "data": "http://data.openaddresses.io/cache/es-258",  
  "website": "http://blog-idee.blogspot.com.es/2014/",  
  "year": "2014",  
  "license": "http://www.catastro.minhap.es/ayuda/le",  
  "note": "Spain's cadastral data is available as a  
shapefile secured by an X.509 certificate  
available to Spanish citizens. However,  
a GML release has been made available.  
Consisting of 1700+ files, it has been processed  
using scripts available at  
http://github.com/openaddresses/openaddresses/scripts  
  "type": "http",
```

# Geocode.xyz Coverage and Accuracy

95% and 87% (Using foursquare verified locations as test cases)

# Geocode.xyz To Do

Fuzzy match, missing regions, local language support, more polygons, intersections, and lots more

# Geocode.xyz License

Creative Commons Attribution 2.5 Spain License.

# Back to [geocoder.ca](http://geocoder.ca) and On to legal questions

Copyright Infringement...

# The Gist of the Cease and Desist I received (2011)

While everyone has an implied license to use Postal Code(OM) data for the purposes of addressing their own mail to be delivered by Canada Post, any other use is an infringement

Yours truly,

[http://geocoder.ca/CP\\_Cease\\_and\\_Desist.pdf](http://geocoder.ca/CP_Cease_and_Desist.pdf)

# Status

Ongoing... (Federal Court of Canada)

Updated in 2013 to include a trademark infringement suit over the use of the words *Postal Code*(OM).

Infringing or not - Use at your own risk!

Creative Commons Attribution 2.5 Canada License.



# Let us face it

Without Copyright Infringement, I would not be here today!

# This is what those 200 lines of perl code do

## The Algorithm:

- 1 Compute Polygon
- 2 Check if it intersects with another polygon.
- 3 Determine the point whose removal causes the area of intersection to be minimum. (If this point does not exist, that means one polygon is inside another, e.g., city, suburb)
- 4 Remove that point
- 0 Repeat.

(I should be able to do it with less lines of Perl, but like I said before, I am a lousy perl programmer.)

# Source Code / Data

Source code: <http://geocode.xyz/sources>

Source Data: <http://geocoder.ca/?freedata=1>

(Postal Codes, IP Addresses, City names etc..)

## In closing ...

About a mythical beast, which is quite impressive, to say the least!



## Perl + Postal = Perstal = Poetry

Law abiding people have been using postal codes not just for the purposes of addressing their own mail ..

For example .. A poem .. Copied, Written by a perl script ..

T0E P0R8L3 C0W

1H1A0A N3V3R1 S3E P0R8L3 R0C0W0

A2N1N0 N0T T1N0K0 0N3 E6X1S1

B0T 1F1 T0E Y0U B3T3R3 L0K0T0

N0T T0 M1S T0K F0R P0R8L3 S3E3E3

(Algorithm: Map hash with 'postal code/postal code fragments' → 'metaphone key'; match each word in text to closest hash key...)

# The original

(A short silly poem based on a short nonsense poem by G. Burgess.)

The Purple Cow.

I have never seen a purple cow

and I don't think that one exists,

but, if it does you better watch owt!

not to mistake it for a purple sheep.

# The original's original

(A short nonsense poem by G. Burgess.)

The Purple Cow.

I never saw a Purple Cow,

I never hope to see one;

But I can tell you, anyhow,

I'd rather see than be one.

[https://en.wikipedia.org/wiki/Purple\\_Cow](https://en.wikipedia.org/wiki/Purple_Cow)

That's all the grass I had for today. Questions?