

Crowd sourcing the shape of neighbourhoods

Tom Taylor

It's exciting to see that the AGI Geocommunity adopting the value of place as its theme this year. Because if we want location based services to be natural and comfortable to use then they need to speak to people in terms they can relate to rather than coordinates and radius searches.

Right now it's possible for me to pull out my iPhone and to search for restaurants nearby and be presented with a map, full of red dots, each with distance accurate to the nearest meter. And whilst this has a value, it doesn't feel very human to me. I want these services to converse in human terms about location and space; to understand that cities are vast and different; that home is important and that my perception of the city is not just about co-centric rings. I want them to understand place.

In order to turn 'place' into something usable by developers and designers of location based services we need to find ways of codifying it and making it as usable in software as coordinates are to us already.

Of course, truly codifying place is almost impossible, as no database can capture the complexity and nuances of detail that make up how we perceive somewhere. But we need to try. At its simplest, we need to attempt to understand the shape of places, where the boundaries between areas lie, and where they're fuzzy and where they're clear. And from there we can begin to project upon that.

But one person's opinion won't do. Where I think an area lies is probably different to where you think it lies. To gather a useful set of data we need the collective opinions of the residents and visitors to an area, aggregate those and produce usable data.

The traditional response would be to survey, perhaps with questionnaires and door knocking. To do this across the country at neighbourhood level would cost millions of pounds and years of effort in the capture and analysis of data. To recoup some of the cost, the data produced would probably be expensive and inflexible in terms of licensing.

And yet it's already been done. For free. Almost accidentally. It's updated every few months. The results in densely populated areas such as London, New York and Paris are fantastic, and it's usable and improving in less dense areas.

It's been done by a photo sharing website called Flickr [<http://flickr.com>]. They're owned by Yahoo and they hold over three billion photos from tens of millions of individuals worldwide.

Flickr helps you share and organise your photos, with tags, descriptions and comments. A couple of years ago Flickr added geotagging: the act of adding metadata to a photo with location information. Users can just click 'add to map' next to their photo and position the marker where the photo was taken.

Some mobile phones such as the iPhone or high end Nokia phones have GPS receivers built into them, allowing the location to be captured at the time the photo was taken, allowing it to be geotagged without any effort from the user.

Over 100 million photos have been geotagged on Flickr as of February 2009.

So, that's one half of the equation.

The other half of the equation is Yahoo's purchase of the Where on Earth database, a hierarchy of places. For example, Homerton, where I live, is a neighbourhood in Hackney, which is a borough in London, which is a city in England, which is a country in the United Kingdom, which is a sovereign state in Europe, which is a continent on Earth.

Each place in the WOE database has a bounding box. A rectangle encompassing it.

Sometime last year Flickr introduced the ability to guess the name of the place where your photo was taken, using an educated guess from the bounding boxes of places nearby. This is called reverse geocoding, and it's quite hard, as Aaron and Dan have explained in a previous post.

Because bounding boxes overlap, it's only a guess, but it's often a good one. However, people can override this guess. If Flickr suggests that it's taken in Homerton, I can say "no it wasn't", and change it to another nearby place.

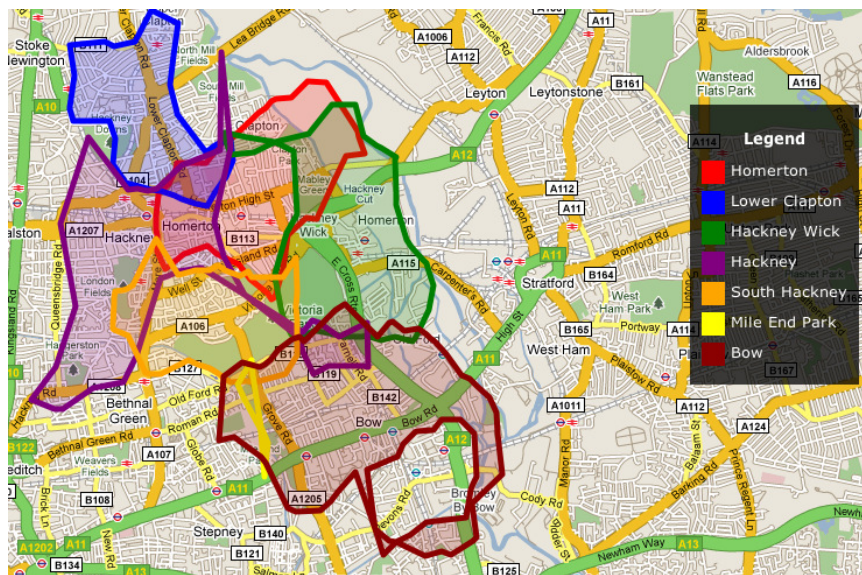
That rectangular block gets chipped away and chipped away, until it starts to approximate what people collectively perceive that shape of that neighbourhood as.

The shapes are generated with a tool called Clustr, that takes a collection of points and uses that to generate a polyline that best fits the majority of them. Flickr publish these shapes as XML, available through their API or as a download.

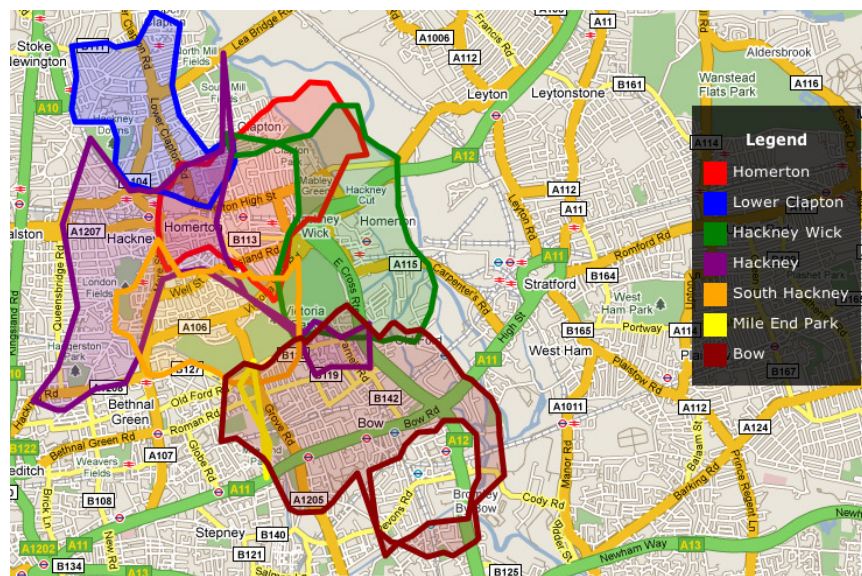
I was interested in visualising this data, so I built a tool called Boundaries to do just that. You can use it at <http://boundaries.tomtaylor.co.uk>.

It enabled you to type in a name of a neighbourhood or suburb and visualise the shapefiles that Flickr generates, along with those of its neighbours.

Here's one for Homerton, London.



And one for Shoreditch.



Every few months Flickr regenerate them, so as more photos get added, and the old ones become less relevant, the shape evolves. Homerton might shift north, encroaching on Clapton, due to zoning changes or a school renaming, or maybe just Clapton, as a name, falling out of fashion.

We might see cities expand, such as Los Angeles in the 80s and 90s, or die, such as Detroit, right now. As a historical tool it's going to be fascinating.

The Flickr shapefiles are licensed under the public domain, and the WOE database is available under Creative Commons Attribution License, meaning that anyone can take this data and build on it, for any purpose they like, without commercial restrictions.

So, what kind of things can you build on top of this?

One example is Dopplr, a service to help people travel smarter. One component of it lets you see and share what's good in a city to eat at, stay in, or explore. When you mark a place in Dopplr as 'liked', it shares that with people you trust, such as friends or colleagues.

Dopplr can build up a profile of what kind of restaurants you like, and thanks to the Flickr shapefiles, which neighbourhoods or suburbs they're in.

They can start to perform calculations, matching you against people who like similar places, and using that suggest areas you'll like in other cities that you visit.

For example, people who like to go out in Hackney, London, might like the Mission, San Francisco, the Mitte, Berlin and De Pijp, Amsterdam. And people who go out in Kensington, London might like similar places in those respective cities. Almost like twin neighbourhoods, based on the opinions of millions of Flickr users.

I hope this has given you a little bit of an insight into the value of the Flickr shapefiles and the Where on Earth database. I'm exciting about using them in my next project and I'm looking forward to seeing what other people can do with them.