



This article is reprinted from *Search Engine Land.com* by Quova with permission. ©2007 Third Door Media. All rights reserved.

by: Chris "Silver" Smith
August 13, 2007

Geolocation is bandied about quite a bit when discussing aspects of online marketing with location-specific components, but many are blurry as to how it works and how it's being used, so I thought it'd be helpful to outline the basics of it, and to highlight some of the recent developments brought via the expansion of wifi and mobile device use that have improved its precision. Geolocation is coming into broader and broader usage in enhancing the user experience for local search and mobile applications, and it has quietly become a vital component to the policing of fraud—particularly for credit card validation and filtering of PPC advertising clicks.

With all the enthusiasm surrounding the use of geolocation tech, few people really speak to the questions of accuracy with the technology as well— a point that is odd, considering just how integral the technology is to the highly-publicized concerns surrounding the reliability of fraud detection in the paid search marketing industry. By some industry reports, click fraud may be greater than 15% this year, and both the secrecy surrounding the detection technology and the anecdotal problems advertisers have seen in actual conversion rates continue to bedevil the search marketing world. I'll try to clearly outline places where geolocation may still be prone to error, and perhaps others may push for greater industry accountability.

How it works

At its most basic, online geolocation we're referring to is an attempt to identify the actual physical location of internet users. There are a few different ways that this may be accomplished. The best-known method is to take the user's IP address, which is transmitted with every internet request, and to look up the organization and physical address listed as the owner of that IP address. Anyone can do this, by querying the Whois information at ARIN - the American Registry for Internet Numbers. (Note: this is NOT the same as a domain name Whois query! Many IP addresses may not be associated with a domain name at all, so a domain name Whois of an IP address may not get you geolocation info.)

For instance, let's say that I noticed that a visitor to my website came in on IP address 216.64.210.100, according to my server's log files. I can query ARIN for that IP address, and I see that it's an address included within a block of IP addresses owned by The Coca-Cola Company:

[ARIN Home Page](#) [ARIN Site Map](#) [ARIN WHOIS Help](#) [Tutorial on Querying ARIN's WHOIS](#)

Search for:

Search results for: ! NET-216-64-210-0-1

OrgName: The Coca-Cola Company
OrgID: [THECCOC](#)
Address: 1 Coca-Cola Plaza
City: Atlanta
StateProv: GA
PostalCode: 30313
Country: US

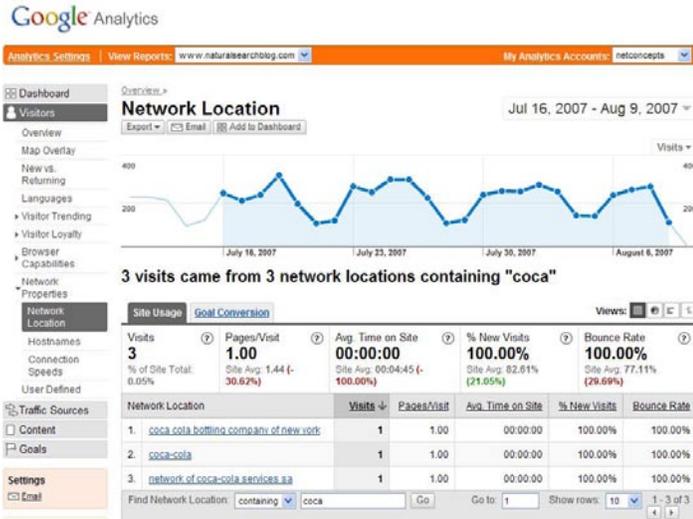
NetRange: [216.64.210.0](#) - [216.64.210.255](#)
CIDR: 216.64.210.0/24
NetName: [SAVV-5262381-0](#)
NetHandle: [NET-216-64-210-0-1](#)
Parent: [NET-216-64-192-0-1](#)
NetType: Reallocated
NameServer: NS1.SAVVIS.NET
NameServer: NS2.SAVVIS.NET
NameServer: NS3.SAVVIS.NET
Comment:
RegDate: 2006-09-06
Updated: 2006-09-06

RTechHandle: KTR34-ARIN
RTechName: Treanor, Kevin
RTechPhone: +1-404-678-8774
RTechEmail: ktreanor@na.ko.com

OrgTechHandle: [KT24-ARIN](#)
OrgTechName: Treanor, Kevin
OrgTechPhone: +1-404-676-8774
OrgTechEmail: ktreanor@na.ko.com

ARIN WHOIS database, last updated 2007-08-08 19:10
Enter ? for additional hints on searching ARIN's WHOIS database.

I could then perhaps figure that this visitor was an employee of The Coca-Cola Company, perhaps reading an article in the series of pieces I recently did about the Coca-Cola website. Indeed, my Google Analytics report is showing that I got a few visits from people associated with Coca-Cola during that time:



Since I can identify visitors from The Coca-Cola Company, I could deliver up content specific to them - I've heard stories about Google and Yahoo delivering up ads for engineering positions to the employees of Microsoft in Redmond using this method, for instance. More importantly, I can now assume that this user is likely to be physically located in Atlanta, Georgia—so I know their City, State, Zip Code, Designated Metro Area, and Country!

Naturally, it's likely not feasible to automatically perform an ARIN lookup with each visitor to your website before delivering up data, because it would take too long. So, there are a few companies out there who are aggregating and caching the network data and either providing lookup tables or web service lookups to those who wish to deliver location-specific content or who are using the data for reporting or fraud detection purposes.

Some ISPs which provide internet access through hotels may now be providing the physical locations of their networks of access points to the geolocation data aggregators as well, and in many cases these ISPs are hosting the default web page portals of local information to the hotel visitors. Some ISPs may also be quietly providing geolocational data to the aggregators as well, allowing all their customers to be geolocated to varying degrees.

Also, internet service providers who host Wi-Fi hotspots throughout the world are providing data to various of these aggregators, allowing the hotspots' IP addresses to be associated with precise physical addresses.

Mobile phones are able to be geolocated by triangulating their location from area cell phone towers, and there are increasing numbers of wireless devices such as phones, PDAs, and laptops which are getting integrated with GPS satellite pinpointing, paving the way to associate precise coordinates with them. As more mobile devices like the iPhone leverage Wi-Fi access, there will be a variety of geolocational methods which will be able to pinpoint mobile users.

Who provides the geolocation data?

Quova is considered the best-in-class (probably with a price tag to match) of the geolocation data aggregators, and their data is apparently used by Google, Yahoo!, and MSN to geotarget content and ads, and likely for the purposes of analytics and fraud detection as well. They were founded in 2000 and they geolocate users through IP address location data as well as tracing network gateways and router locations. They also likely traceroute users coming through proxies to better determine location to some degree, and they analyze request latency of users passing through proxies to help determine physical distance from the proxy servers' physical locations.

Quova recently partnered with Mexens Technology in order to supplement their IP/network location data with Wi-Fi hotspot locations, device GPS, and wireless tower triangulation.

Quova uses Pricewaterhouse Coopers to audit their geolocation data, and are perhaps the only company allowing independent, third-party validation testing of this sort. Their GeoDirectory Data Sheet states that PwC does this auditing by testing Quova data against "...large, independent third-party data sets of actual web users...". I interpret that to mean that PwC likely obtains IP addresses from some ISPs who tell them the countries and states associated with the IP addresses, and they check to see how accurately the Quova data identifies the locations of those addresses.

Digital Envoy was possibly the first company to work on geolocation, founded in 1999, and their data is apparently based primarily upon IP address data. They may also be performing network routing analysis to some degree, but their documentation doesn't specifically state this as Quova's does, and they do not represent that they have independent auditing. Their product is likely a bit cheaper than Quova's, though, and their clients include AOL, Ask.com, CNET Networks, CNN, DoubleClick, Omniture, and more.

Akamai is primarily a content delivery network service, but their positioning in that space was a natural fit for geolocation service as well, so they added this on as a product called EdgeScope in 2001. Akamai's product is based upon IP address locations along with extensive ability to map network gateways, routers and paths of user requests in order to match up with users' physical locations. Akamai's EdgeScope is probably a bit pricey, but, considering how many large companies are already using their content distribution

services to some degree, there could be some sort of synergy to also contracting them for their geolocation product. Due to their worldwide scope and integration with networks, their IP mapping capability is probably greater in quality than Quova's, but they haven't apparently broadened to include Wi-Fi and mobile location data, nor do they mention independent auditing.

IP2Location.com was founded in 2001, with headquarters in Penang, Malaysia, and their data is likely based entirely upon IP address data. They have a number of data products, and are probably the cheapest of the IP data providers, particularly if you only need a restricted set of data lookup tables, though one suspects that they're possibly also of the lowest quality.

Mexens Technology uses GPS, Wi-Fi, and wireless towers (partnered with Quova, as previously mentioned) to provide some level of geolocation ability. They also have an interesting API.

Skyhook Wireless provides geolocation service based off of Wifi users' locations. Skyhook was recently named one of FierceBroadband Wireless's "Fierce 15" of 2007:

"As LBS applications in the mobile world heat up, Skyhook Wireless will play a key role. The company has pioneered the development of the first-ever metro-area positioning system that leverages WiFi rather than GPS satellites or cellular towers to deliver precise location data," said Lynnette Luna, Editor of FierceBroadbandWireless.

The Skyhook Wireless Wi-Fi Positioning System (WPS) includes a database with the known physical location of more than 19 million WLAN access points across a growing coverage area that reaches over 70% of the U.S., Canadian, and Australian populations, making it the world's most comprehensive WiFi database. With the market for local search exploding, the applications for WiFi positioning software are endless. Some of these include Internet search, proximity advertising, search and recovery, E911, fleet management, buddy finders, and more.

Skyhook's also set up a locational beta service for users called Loki, along with an interesting free open API for developers.

Downside of Skyhook: precision best in Wifi-dense locations such as centers of major cities—otherwise uses Skyhook is using IP location addresses as a fall back.

I've just touched on some of the companies that are most interesting to me who are providing geolocation products and services. There are likely quite a number of companies which are also doing this in-house to some degree. For instance, I wouldn't be surprised if Google wasn't geolocating through querying and caching of ARIN data on top of data they're receiving from other providers listed above. Considering how vital geolocation data is to the policing of click-fraud, Google could be building out their own complete geolocation data aggregation infrastructure. Further, it's also been suggested that Google is likely using domain's registration data through

Google's status as a registrar to assist in associating websites with geographic locations for Google Maps—not precisely the geolocation of users I'm covering here, but a closely related method that could be useful to local SEO.

Many mobile service providers are also using the geolocational information associated with their devices in order to deliver up location-specific information on their own, without the assistance of the geolocation data aggregators.

How geolocation is used in the local space and in general internet marketing:

- **Targeting Ads to user's locality** - ads could be targeted by varying levels of locality including ZIP Code, City, Metro Area (DMA), Region, State, Company, Country, and Time Zone. For example, I just performed a search in Google for "personal injury lawyers", and you can see that they displayed a number of ads for lawyers who've targeted ads to the Dallas, Texas metro area where I'm writing this article:

The screenshot shows a Google search for "Personal Injury Lawyers" with approximately 4,430,000 results. The search results are divided into sponsored links and organic search results. Several results are circled in red, highlighting local targeting. The circled results include:

- Sponsored Links:** "Dedicated Injury Lawyers" (Qualified personal injury lawyers, Over 60 years experience, Local office, Dallas, TX), "Personal Injury Lawyer" (Dallas-Fort Worth Firm, Recognized lawyer, medical & more, www.toddsmithlaw.com, Dallas-Ft. Worth, TX), "Serious Injury / Death" (Texas law firm represents seriously injured people and their families, www.dearmlone.com/injurydeath.html, Dallas-Ft. Worth, TX), "Personal Injury Lawyers" (Nationwide directory of personal injury lawyers. Free case review, www.PersonalInjuryLawyer.com), "Personal Injury Lawyers" (Page One, Settlement You Deserve, Dallas, Blakeley/Ramey.com, Dallas, TX), "You Have Legal Rights" (Serious injuries require serious lawyers. Contact us today for help, www.mattthewlawfirm.com, Texas), and "Rad Law Firm" (Dallas, TX, Personal Injury Lawyers).
- Organic Search Results:** "Personal Injury Lawyers" (National Network and Directory of Personal Injury Lawyers. Provides a forum, bulletin for personal injury lawyers, www.napil.com - 30k - Cached - Similar pages - Note this), "PERSONAL INJURY LAWYER - ATTORNEY NATIONWIDE!" (Personal Injury Lawyers experienced in wrongful death, auto accidents, trucking accidents, motorcycle injuries and many other types of injury litigation, www.Ipersonalinjurylawyers.com/ - 41k - Cached - Similar pages - Note this), "Find a personal injury lawyer for information on law, lawsuits ..." (Personal Injury FYI is a resource for finding a personal injury attorney and learning about personal injury law, including the elements of a lawsuit, ... www.personalinjuryfyi.com/ - 22k - Cached - Similar pages - Note this), "Personal Injury Lawyer. Claim your Compensation. Find an Attorney" (Small directory of personal injury lawyers searchable by state, www.personalinjurylawyer.com/ - 35k - Cached - Similar pages - Note this), "Personal Injury, Car Accident, Medical Malpractice - Lawyer ..." (San Jose, CA - Highly Experienced Injury Lawyers Practicing in the Bay Area since 1975, Free Case Analysis: (408) 246-5500 ... injury.findlaw.com/ - 56k - Cached - Similar pages - Note this), and "Personal Injury Lawyer, Lawyers, Attorney, Attorneys, Law ..." (Stages of a Case - Information on the stages of a personal injury case, including contacting a lawyer, fact-finding, settlement, trial, and collecting ... injury.findlaw.com/personal-injury/ - 35k - Cached - Similar pages - Note this).

- **Targeting locally apropos content to users**, including language delivery, currency such as pounds/euros/dollars/yen/etc—providing native users' currency on e-com pages and order forms, location-specific text/images, customization of web search results which may have a local component, automating Store Locator pages for retailers, etc.
- **Content Restriction:** there are frequently some contractual/legal limits on what products and services can be sold where. Uses include restricting online gambling from US users; enforcement of trade embargoes so that certain items won't be sold to countries disallowed by federal laws; some items can only be sold in particular areas of the world and some promotional contests are only allowed by certain states or provincial rules.

- **Financial Fraud Detection:** denying sales to possibly compromised credit cards or bank accounts - for instance, if the IP address of the online user is in suspect foreign country, but account owner address is in the US.
- **Identity Fraud Detection:** geolocation provides additional signal for logins for protecting user identities.
- **Advertising Fraud Detection:** filtering out invalid or fraudulent clicks - products/services only available in one country, but Pay-Per-Click advertising clicks are coming from another.
- **Potential Detection of DoS Attacks:** many requests coming in from a wide variety of natural-looking IP addresses, but geolocation of requestors shows requests actually coming all from one primary location.
- **Internet Analytics Applications:** analyzing and showing from where visitors viewed a website, and quantifying how many come from particular locations.
- **Site Server Locations for SEO:** there's some supposition that websites hosted in the country who's audience they're targeting might actually get better rankings within search engines targeting that country's users. (See Ian McAnerin's article on Geolocation for SEO.)

The issue of error rates

From the very beginning, geolocation providers have been asked about how much error is involved in their ability to pinpoint web users, and from the very beginning geodata consumers have noticed some amount of errors happening. There are a lot of anecdotal tales of ads and content being incorrectly displayed for users when their geolocation has been incorrectly assessed.

The classic example of IP locating error is caused where a large internet service provider may provide web access across the world, but the block of their users' IP addresses are all associated with the ISP's corporate headquarters or network office in one location. With simplistic IP address mapping, all those users could be geolocated by aggregators to that single corporate office location, even though they might in actuality be spread out in many areas. The most famous example of this is the AOL proxy server issue wherein geolocation aggregators were originally unable to pinpoint AOL users and incorrectly associating them all with their Virginia address.

Quova used to claim to have beat the AOL proxy barrier to identify where their requests originate, but specific terminology touting this ability has been considerably toned down these days in Quova's collateral materials, and their GeoDirectory data sheet merely mentions that they have included a flag for AOL. One assumes that their confidence factors rating for geolocation and general proxy detection/locating ability might be used to give some level of AOL user identification ability, but the flag must be provided so that the geodata consumers could opt to not geolocate AOL users if

they presumed the data to be too error-prone.

While the AOL proxy issue is the most famous, many other ISPs likely have some similar barriers to pinpointing their users. Using one of the previously-mentioned geolocation services, I just now checked my IP address and was mapped to Keller, Texas, even though I'm writing this 20 miles away. Large corporations likely have this going on as well. For instance, in the Coca-Cola IP address example I gave above, I'd bet that the company is large enough that they probably have offices throughout the states and world, and their employees addresses might be prone to being incorrectly mapped to their headquarters locations.

Since IP address mapping using ARIN registrar data could be so prone to error at the more granular levels, a number of the geolocation providers rush to quote accuracy estimates based on the broader, country and regional levels:

Quova: "...In audited tests using large, independent third-party data sets of actual web users, Quova's country level accuracy was measured at 99.9%. US state level accuracy was measured at approximately 95%."

IP2Location: "...over 95 percent matching accuracy at the country level..."

Another factor occurs when users specifically choose to route their requests through a proxy in order to anonymize their internet usage, either for privacy reasons, or for the sake of hiding criminal activities. A number of sites out there provide free or paid anonymizing services, allowing users to submit their internet requests which then get filtered through another layer of services before the requests reach content providers' servers.

Obviously, geolocation accuracy could be more accurate through network route mapping and enhancing IP registration data with data from the large ISPs, along with Wi-Fi and mobile device location data.

Users browsing the internet through mobile phones and other wireless devices now pose an additional proxying problem, since most of the wireless carriers will display only a central IP address for all of their users, and any attempts at network routing will be stymied by the fact that wireless network traffic isn't being monitored. For the companies who are providing content through these wireless carriers' mobile portals, they may be supplied geolocation info by the carriers, but this may not help most webmasters who don't have such partnerships. As more mobile device users demand open access to the entire internet, the mobile carrier's proxies may become an increasing source of error in geolocation data.

Freshness of data weighs in as well since IP address blocks change over time, so if an IP location source doesn't update their database, it can result in incorrect targeting, just as with this incident related by Barry Schwartz where a Texas school district kept getting content from Google Canada.

The biggest problem in assessing the error rates of geolocation data is the simple fact that there's no way to really test well for accuracy. The one and only company which publicly states that it uses external auditing (Quova), provided by Pricewaterhouse Coopers, is apparently testing by comparing their geodata with large datasets where they know the physical locations of the users associated with the IP addresses. But, how broad is that comparison data? Is the testing comparison working the same as when users are dynamically being geotargeted through the data in real-time? Does data from just a few major ISPs (assuming that's what's being used) really represent the majority of internet users? Does it take into account the huge amount of corporate employees browsing during their workdays? (I'd guess not, since most large corporations probably shouldn't be sharing the locational information associated with their employee's IP addresses.) What's the estimate for accuracy at the city-level and postal-code level?

At best, this is only an estimation and not direct test results for accuracy, so we don't know what the error rate really is.

To be fair, it's simply not possible for any of us to know the actual error rates involved, since it's impossible to assess whether all internet users are being accurately geolocated through any of these services. We can only sample some amount of users, and decide whether that sample set should be considered representative of all usage or not.

On one hand, this inability to assess error rates more precisely is highly concerning, particularly for the paid search industry, since it makes the entire policing structure of click fraud appear to be built upon a house of cards.

On the other hand, the filtering of suspect clicks is primarily based upon identifying the country where the click is originating. Countries with higher apparent rates of fraudulent clicks tend to be flagged as less-trustable, and those clicks are discounted from billing. Based on the logic that most ISPs are fairly country-specific, and that most large companies might use completely different IP address blocks for their employees in different countries, I'm willing to believe the industry's published accuracy rates of 99.9% to 95% at country-level geolocation. But, when you're speaking in terms of processing billions upon billions of clicks, and millions of dollars, 5% to 0.1% can still amount to a whole lot of money...

Even considering the higher accuracy of country/regional geolocation, there's still cause for concern for advertisers who are buying ads and targeting at the more granular levels—are their ads being shown to the right demographic groups, and are their clicks coming from the qualified buyers they're seeking? The more granular levels of geolocation are apparently still considered to be much more error-prone, and the industry remains quiet about it.

Other downsides to use of geolocation:

Geolocation is probably a very bad method for targeting languages! Better to use content negotiation through browsers, using the language-accept headers to choose which languages to display to users (this is what the W3C recommends). While using geolocation to choose which language to deliver up to an user, search engine spiders may all come in from a central location or from one of their regional data centers, so using geolocation for language targeting would not be best practice and could result in less-optimal natural search marketing.

Even delivering up local-oriented content by geolocation of users can be dicey, if one doesn't properly handle search engine spiders. Last year, I informed representatives from Amazon.com on how their geolocation for the purpose of delivering up their yellow pages links was ruinous to their SEO of that section, since Googlebot was apparently being delivered up all Washington, D.C. content, keeping the rest of their national content unavailable for indexing. Geolocation can be great for targeting content to users, but design a default for unidentifiable users and search engine bots.

Geolocation can creep out users who don't understand how it works and can raise user privacy concerns. Most users still don't realize their physical locations are being mapped while they're browsing, so many still don't quite know enough about the technology to be concerned. The industry hasn't really addressed this as well as it could. Quova's FAQ is rather dismissive of privacy concerns, saying only "Since accuracy is limited to zip code level, Quova does not pinpoint individual user locations...", though this seems a bit inaccurate since they are also apparently incorporating GPS, W-Fi, and wireless tower triangulation through Mexens Technology - meaning the pinpointing of users could be a whole lot more accurate than mere ZIP code level.

Geolocation can reveal some information you wanted to keep confidential, which is why it should be on the radar screens of privacy advocates. Don't want your competitors knowing you're examining some of their pages every day? If you're viewing from a unique city where average users are unlikely to be viewing your competitor's site pages, you might want to try dialing up through an ISP outside of your town or going through a distant proxy or two before viewing their pages, just to try to obscure your geolocation info. Or, call up a friend in another state to send you screen-grabs of the site.

For travel-based industries, filtering out PPC clicks from suspect foreign countries could result in undercounting of valid consumer traffic. That's cool if you're a travel business

advertising in PPC networks, since it may get you more free ads and higher apparent conversion rates. But, it's not so cool for the ad network companies and publishers displaying those ads - they're likely getting a little less revenue than they should since some of the "good" traffic is inevitably going to be thrown away with the "bad".

Summary

Geolocation is here to stay in the online local space. Its use in fraud detection and regulatory compliance is only deepening, and geolocation reporting in web analytics has become a standard. Geolocation data is a necessity for the geotargeting of ads, and that would appear to be an increasingly popular choice amongst marketers as online advertising continues to gain traction among local businesses.

Geolocation use in targeting relevant content to users is still in something of an experimental stage, and few sites seem to be really making simultaneously extensive and effective use of it.

It should not really be used in content mediation for delivering different languages, since this likely will not allow the various translations of the site pages to be properly indexed in the search engines for various countries/tongues.

Geolocation may have a factor in effective SEO—anecdotal evidence and logical reasoning would indicate that it could make sense that a site hosted within a particular country might be more relevant to that country's citizens than in other countries. I would guess that this factor wouldn't apply as much for higher-PR sites or publicly-traded companies, but there's not a lot of research evidence out there.

The biggest issue with geolocation is the lack of transparency in how the aggregators are gathering the data, and how high the error rates may be with all the levels of granularity. The geolocation providers all desire to keep their methods proprietary, but this competitive need for confidentiality makes it difficult for companies to try to estimate relative levels of accuracy amongst the providers. Many companies may be using cheaper providers than they should for the purposes of advertising click-fraud detection, leaving themselves open to liability of fraud claims, and causing innocent advertisers to be paying higher amounts than they should. Considering how geolocation has become such a major component of the policing of click-fraud, it's surprising that there hasn't been a wider demand for transparency and standardized methods for testing accuracy. The leaders in the industry should pursue a greater degree of openness and a greater variety of auditing methods to check accuracy.



Quova Corporate Headquarters

707 California St.
Mountain View, CA 94041
USA
Tel: 877.737.8682 (US only)
sales@quova.com

Quova B.V. European Headquarters

Jan Luijkenstraat 94
1071 CV Amsterdam
The Netherlands
Tel: +31.20.888.45.30
sales-eu@quova.com