

Web 2.0 / 3.0 Semantics Needed; Teaching Human Geography to the Internet

Gary Gale, Yahoo! Geo Technologies

Scope

The computers that power the Internet use algorithms to recognize geographic features and places, and prefer concrete disambiguated place references. The humans that use the Internet prefer abstract, ill-defined and colloquial references to place. This talk provides an overview of the issues surrounding describing place, provides commentary on the benefits of improving the semantic recognition of place and gives details of the ongoing collaboration between the Geo Technologies and Research arms of Yahoo! in this area.

Position

“The nature and importance of geography are not well understood. The differences between physical and human geography and the shared core of concepts can be confusing ... understanding of geography outside the discipline often extends no further than the colloquial”. *Geography: A Very Short Introduction, Matthews and Herbert, 2008.*

Location has taken off; sales of location enabled devices are predicted to exceed 800M and GPS enabled mobile handsets are predicted to exceed 450M by 2011.

Location also has proven benefits; Yahoo! has experienced a twofold increase in time spent on network, a five-fold increase in page views and two-fold growth in local queries year on year since enabling location recognition technologies across the network.

As data increasingly becomes either open, free or licensed in a way that it is able to be used to create derivative works (see the City of Vancouver Data license model) it becomes a commodity. This commoditization, whilst of benefit to the user also increases the amount of content that is available to the user; the information *hosepipe* of the Internet becomes increasingly more difficult for the user to control.

Domain, product and market leadership are thus available to organizations and individuals who are able to tame the information hosepipe and deliver relevant content to the user, such that the act of requesting information and receiving it becomes as second nature as a street corner conversation.

The fast pace of technological development and an increasing impatience on the part of the user means that systems which are able to understand personal, ill defined, ambiguous and colloquial geographic terms will be positioned to help deliver the much heralded, but also much delayed, semantic web.

©Copyright 2010 belongs to the Author identified in this document.

The views expressed in this study are the personal opinions of the writer(s) and do not necessarily reflect the policies and view of the AGI or any of its individual or corporate members or the authors' employer.

The AGI Foresight Study - The UK Geospatial Industry in 2015

An Expert Paper



Just as "The Queen's English" and "Received Pronunciation" are not used in day-to-day conversations, so formal and fully disambiguated place referenced are being used less and less when searching for content on the internet.

Similarly attempts to geoparse and/or geotag extant content needs to be able to recognize colloquial and informal geographic references in order to be able to deliver relevancy to users subsequent searches.

Recognising that the semantics of place are bizarre, capricious and utterly human, Yahoo! has moved from a non-scalable human editorial model to a scalable machine learned tagging and disambiguation system. Doing so yields a 21% increase in precision and a 16% increase in recall when recognising place in queries and in structured and unstructured content.

Such location recognition technologies give insight that Joe Orlando, Jack London and Panama Hat are not references to a place, that London, Londra and Londres are the same place, and which one of the 23 Londons is being referred to.

Anticipated Changes

Whilst current location recognition systems have shown quantifiable benefits, the technologies are still naïve and in their infancy.

It is anticipated that convergence of user location and location recognition technologies will allow location itself to become location aware. Technologies such as the W3C Geolocation API, onboard mifi, cell tower triangulation, GPS and public wifi MAC address triangulation will allow a local focus to be applied to understanding place references, allowing terms such as the Tri Countries, downtown and the High Street to be parsed, understood and applied give the relevance that today's users crave but which today's systems are currently unable to deliver upon.

Summary

1. Location has taken off; 800M location enabled devices and 450M GPS enabled mobile handsets are predicted by 2012.
2. Computers prefer concrete disambiguated place references.
3. Geography is a human construct; bizarre, eccentric, capricious and irregular. We prefer abstract, ill defined and colloquial place references.
4. Understanding colloquial and personal geographic references improves relevance and delivers quantifiable benefits.

©Copyright 2010 belongs to the Author identified in this document.

The views expressed in this study are the personal opinions of the writer(s) and do not necessarily reflect the policies and view of the AGI or any of its individual or corporate members or the authors' employer.

The AGI Foresight Study - The UK Geospatial Industry in 2015
An Expert Paper



5. We can make the Internet work better by making it understand how we talk in the real world

©Copyright 2010 belongs to the Author identified in this document.

The views expressed in this study are the personal opinions of the writer(s) and do not necessarily reflect the policies and view of the AGI or any of its individual or corporate members or the authors' employer.