

FEATURE ARTICLE

"GIS Enabling the Internet"

By Chris Kutler



Chris Kutler's time is spent between maintaining and developing the Archaeology UK website <<http://www.archaeology-uk.com>> and working as a Perl/XML developer at the National Archives, Kew, UK. He has an honours degree in Applied Biological Sciences and a Masters degree in Applied Computing Technology. He has also studied Field Archaeology at Birkbeck College, London.

Imagine you are researching a family history and are looking for web pages relating to the surname 'Martin' and the placename 'Kingston' in 'Devon' (UK). A Google search with this common English surname and relatively common English placename with a wide geographic distribution, will currently generate a list of 58,900 links. However, many of these results will be irrelevant partly because the "Kingston" keyword isn't interpreted in a geographic context - you find you have lots of links to other Kingstons, including people with the name 'Kingston', many, many kilometres away from Devon. Even if you refine the search with something like "martin family kingston devon - hull -surrey", things don't get much better - you still have more than 9,000 sites to sift through and again many of these won't actually relate to Kingston, Devon. Further, in order to build up the family tree, you are likely to want to look for the surname in places near to Kingston, Devon. This would require a knowledge of the names of these places and then repeating the original search with these different placenames. This is going to make the research more laborious because you will need to find out the names of these nearby places and then have to sift through even more irrelevant results lists.

Geographic Information Systems (GIS)

But what if you were able to tell the search engine to perform a query in a geographic context? What if you could tell the search engine to limit your 'Martin' search to this keyword but only report back if the result is relevant to say within 10 or so KM of Kingston, Devon? Surely, this would make things easier?

This is the basis of Geographic Information Systems (GIS), which use positional information such as geographic co-ordinates as search criteria in addition to keywords. We are seeing more of them on the internet, but they have weaknesses because information within them is usually selected by the GIS providers who often make value judgements about what constitutes the best or most relevant information. This could mean that the web pages created by, say, amateur historians about Kingston, Devon, may be under-represented, or not make it into the system at all. This may be because the editor collecting the content for the GIS system hasn't come across these pages in the course of their collation exercise; or the amateur's pages haven't been deemed the best in local community information; yet these may be the most relevant pages to our original search. This situation seems even more frustrating if you consider the huge amount of relevant and well-researched "Community Information" published on the web by individuals and smaller interest groups which is often overlooked when the person doing the research is swamped with irrelevant hits.

Geographically Indexed Community Web Pages

So how can web pages created with a geographic context be made more visible to

Free Virtual Tour of Online Information Online

Take a quick one-page tour of the new 'Online Information Online' virtual exhibition on the VIP site and find out what it's all about:

<<http://www.vivavip.com/oio/>>

internet users? How can these usually unpaid but enthusiastic and highly motivated internet "content generators" mentioned above increase the profile of their work?

First, one has to assume that the major web search engines such as Google will develop their search software to account for geographic information and that the web page indexing programs, called spiders or robots, are able to index web sites in a geographic context. The web page creator then has two main options: They may be able to register the site with the search engine and when they do, also include a "geographic stamp" for the web site such as a co-ordinate; Alternatively, they could embed this stamp within the pages' HTML in a standard place where the search engine spiders/robots have been programmed to look for co-ordinates.

Web Page Registration with GIS-Enabled Search Engine

To illustrate the principle of the first option, a basic working model search engine submission form has been created, and readers are encouraged to submit their own sites or sites they find useful. The search engine's "Go Geo" form can be at <<http://www.gogeo.net>>. Any

GOLD

A nostalgic look back at FreePint, from this time last year to six years ago:

- FreePint No.139 26th June 2003. "Taxation Law Resources Online - Tax Doesn't Have to be Taxing" and "Searching Free Trade Mark Databases on the Web"
<http://www.freepint.com/issues/260603.htm>
- FreePint No.115 27th June 2002. "RSI and the Library and Information Science Professional"
<http://www.freepint.com/issues/270602.htm>
- FreePint No.90 21st June 2001. "Do you Google?" and "Electronic Health Information : A boon and a curse!"
<http://www.freepint.com/issues/210601.htm>
- FreePint No.65, 22nd June 2000. "Internet Intelligence - analysing web-sites for competitive intelligence" and "WAP Technology and Services"
<http://www.freepint.com/issues/220600.htm>
- FreePint No.41, 24th June 1999. "Answering back" and "Finding Links to the Past: archaeological resources on the Web"
<http://www.freepint.com/issues/240699.htm>
- FreePint No.16, 25th June 1998. "Finding information products and services via the Net" and "Bioscience Information on the Internet"
<http://www.freepint.com/issues/250698.htm>

Related FreePint links:

- 'Geography and Mapping' articles in the FreePint Portal <<http://www.freepint.com/go/p189>>
- Post a message to the author, Chris Kutler, or suggest further resources at the FreePint Bar <<http://www.freepint.com/bar>>
- Read this article online, with activated hyperlinks <<http://www.freepint.com/issues/010704.htm#feature>>
- Access the entire archive of FreePint content <http://www.freepint.com/portal/content/>

website address added to the "Archaeology" subject form will also be included in the hit list output from a search of Archaeology UK's ARCHI (Archaeological Sites Index) database at <<http://www.digital-documents.co.uk/archi/archi.htm>>.

Here's how the model works. When you go to the site, you simply enter the address (URL) of the website you want to add to the search engine, together with a placename or a co-ordinate in the British Ordnance Survey Grid Reference format (OS). The placename or co-ordinate is the georeference for the webpage and stamps the web page with a geographic identity. Note that the system can convert a placename into a co-ordinate, so you don't always have to know a place's grid reference in order to submit a web page.

When a web user searches the internet they would simply enter the keywords plus a geographic reference such as a placename or OS co-ordinate together with a distance value, say 10 km. This distance value tells the search engine to only return results from web sites which have a geographic stamp within that distance from the placename of interest (e.g. Kingston, Devon). This is going to have the obvious benefit of excluding all Kingston places outside of Devon and all sites relating to Devon which have the word Kingston in them but not within 10km of the place Kingston.

However, in addition, websites within the 10km radius which don't contain the Kingston placename will also be included as long as the web site has been submitted to the search engine in a geographically enabled way. This overcomes one of the problems mentioned at the start of this article - the need to know the names of nearby places in order to find possibly relevant web pages.

GIS-Enabled Web Pages

A problem with the need to register web pages with the various search engines is that it takes time and there might be hundreds of different search engines to which you want to submit your site. It might be easier to type the geographic locational information (i.e. co-ordinate) directly into web pages' HTML. This sounds difficult, but the technicalities may be much easier than one might think. Again, we have to assume that the major search engines' spiders have been programmed to look for a page's georeference when they index the web page. They already look for keyword information in a page's <meta> tags anyway, so this might make an appropriate starting point, e.g. <meta name="Keywords" Content="martin,family,history, kingston,devon, SX 63 47">.

There are problems with the above in that there isn't anything to tell the spider/robot that the co-ordinate (SX 63 47) is associated with the Ordnance Survey British National Grid Referencing System, but the format of the "keyword", i.e. two upper case characters followed by two sets of 2 to 6 digits which may or may not be separated by a space, makes it likely. How can we find out if a search engine spider looks for these coordinates?

Business information product reviews from VIP

For in-depth product reviews and news analysis for the business information industry, subscribe to VIP and VIP Eye.

Monitor the VIP Wire free for press releases and announcements.

<<http://www.vivaVIP.com/>>

XHTML encoding of Dublin Core metadata

When talking about web pages, the Dublin Core Metadata tags provide a simple and standard framework which allow web page creators to insert general information about the information on the page directly into the HTML of the page. For example, there are tags for 'publisher', 'language', 'subject', and for our purposes 'coverage', which can be used to specify the geographic location to which the information on the web pages relates. If standards like this are used, it is going to be much simpler to program the robots mentioned earlier to find their way to the geographic stamp in a web page. Basically, this just means adding a couple of extra tags to the HTML from the Dublin Core set of tags. Adding the particular Dublin Core tag, called "coverage", is very straightforward and the process shouldn't be allowed to frighten or intimidate anyone. For an example of how one might encode a web page in this way, click the "geocodes" link at <<http://www.gogeo.net>>.

Conclusions

With the number of internet sites ever expanding, adding geographical relevance to a search provides an obvious benefit to many researchers. It provides speed and accuracy to what is often a long-winded and frustrating experience for those interested in genealogy, local history, archaeology and potentially many other fields.

Links

Understanding the British National Grid Referencing System <<http://digbig.com/4bhps>>

Dublin Core

<http://dublincore.org/>
<http://digbig.com/4bhpt>

GIS-Enabled Search Engines

<http://local.google.com/lochp>
http://local.google.com/help/faq_local.html

GIS-Enabled Web Sites

<http://www.gogeo.ac.uk/>
<http://www.digital-documents.co.uk>
<http://www.ahds.ac.uk/archaeology/index.htm>

GIS Discussions
<http://www.vivavip.com/go/w165>