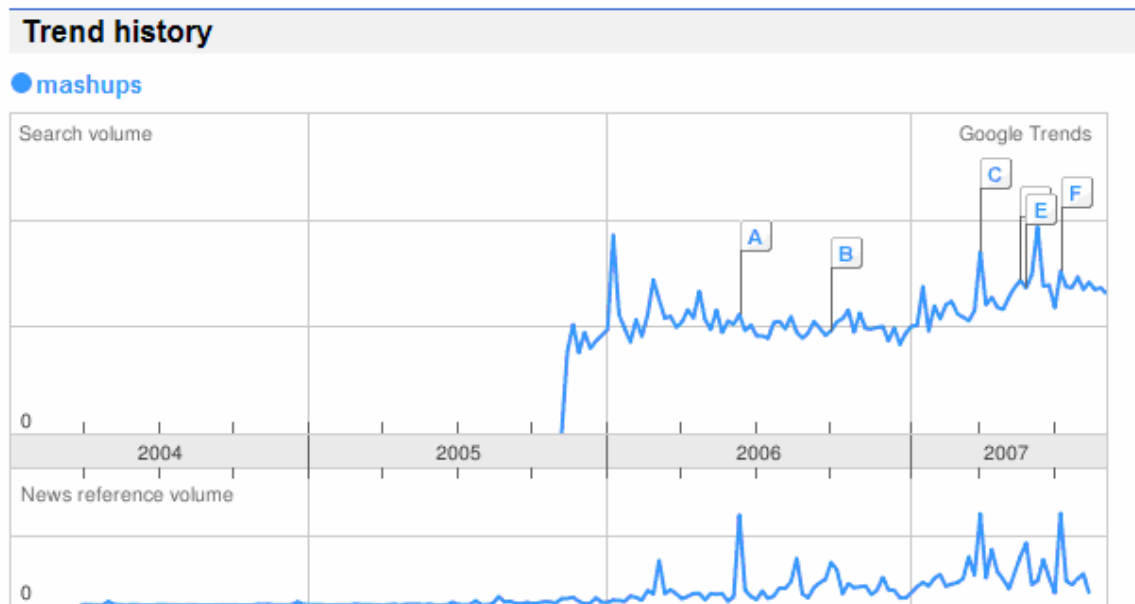


Mechanisms for validation of volunteer data in open web map services

A web ‘mashup’ is a software system that combines content from one or more sources usually from a database (local or web) to web maps usually exposed by Open Web Map Services (OWMS) like Google Maps and Yahoo Maps. The Internet and cheap computers provide fertile soil for the rise of citizen journalism by anyone for anyone. There may or may not be rational reason for adding a web mashup. This major trend has been dubbed “Generation C”, where the C stands for the avalanche of new content (usually photographs or text) by Internet users. Anyone with a bit of creativity and some spare time can create and publish content online using OWMS, and millions do so everyday. In the current online era, number of online searches carried out for an entity can be roughly linked to its popularity and significance in the online world. Google Trends analyzes a portion of Google web searches to compute how many searches have been done for the terms mashups, relative to the total number of searches done on Google over time¹. A graph with the results (search-volume graph) plotted on a linear scale is shown below.



<http://www.google.com/trends?q=mashups&ctab=0&geo=all&geor=all&date=all&sort=0>

Graph shows a steady state in the year 2006 but a sudden surge in the year 2007. This surge is mainly due to the increasing availability of mashup tools in online market. This trend is expected to see an exponential rise in coming days.

With cheap GPS units, computers and the Internet, they can form teams of civilian surveyors that construct maps such as OpenMumbaimap, OpenStreetMap projects. They

¹ <http://www.google.com/intl/en/trends/about.html>

offer spatial data that is free of any legal restrictions with their use. While this may provide opportunity to deceive misinform and offend, licenses and user agreements with terms of use are implemented to quash these potential problems. Still there is scope for legal problems. Acceptable Use Policies in mashups need to be examined more closely in the future, however for the most part, map mashups have revolutionized on-line mapping. Though a few citizens have become amateur surveyors, many more are actually creating map “mashups” or uploading information such as photographs from GPS phones that can be added to collaborative online mapping sites, hence the rise of the amateur cartographer. Paul Rademacher, a software engineer, is credited with sparking the mashup explosion, when he combined real estate listings from craigslist with Google maps to make HousingMaps.com. Paul was frustrated with looking up real estate listing in the newspaper, then logging onto Google Maps to look up the property addresses. Most importantly, though, the pushpin applications are gathering data. It has been noticed again and again that the Web 2.0 idea that a user’s data is valuable is one that’s hard for traditional data companies to understand. They are heavily invested in driving streets, and spending many millions of dollars to generate their data based on real ground truths. They’re very uncomfortable with the idea that the intangible expressive associations of place (“I got engaged here,” “best steakhouse in New Delhi,”) are also valuable. The established players have been watching mash-up developers and are opening up their own data APIs. The social data layer isn’t the only thing coming from users. Both the Open Street Map project and the Mumbai Map project take data contributions (usually in the form of GPS traces and manually typed notes) from users and make them freely available. As the tools become better these types of projects will increase in number and scope.

Mashup on maps provided by OPWMS is a recent phenomenon. Number of OPWMS is increasing with time because there is tremendous response to the existing OPWMS. Google expects developers will use technologies such as a MySQL database, a Linux OS, and programming languages such as Ruby, Python, or PHP². Mash up on maps is yet to get its momentum and even if it continues at the given pace, it may face the following problems:

1. Density of mash up will increase so much that user viewing the application will get confused with the extensive no. of mashups and mashups will loose its relevance.
2. Presently there is no check or validation for adding the mashups. This raises a question on the authenticity of the mashups. Users may loose the faith on the mashups.
3. Mashup may reveal the private information of a third person. E.g. pictures of celebrities in privacy. This may amount to intrusion in privacy.
4. Data servers providing the mashups to application may not be in a position to absorb the massive load of new mashup once this craze spread to internet savvy people in highly populated countries like India and china.

There is no fool proof mechanism to tackle with the above problems but following steps can be taken to put a check on the above mentioned problems. OPWMS are provided by big companies like Google and Yahoo etc but they don’t have the expertise in the

² http://www.infoworld.com/article/07/05/31/google-day_1.html

validation of spatial data. They procure GIS data from vendors like Navteq, TeleAtlas etc. These vendors must be entrusted with the responsibility for screening and validation of data behind mashups. These vendors procure data themselves and also from third party vendors and they have existing mechanisms for screening and validation of spatial and attribute data. The data from mashups can also be treated as third party data. They only need to modify the existing mechanisms for validation and screening of mashup data. There must be some change at the other end too i.e. at the user end who is adding the mashup. Open Geospatial Consortium must come up with standards for mashup features on the lines of Well Known Text (WKT) and Well Known Binary (WKB). User will follow a standard while entering and editing the mashup and it will be easy for the data providers to screen and validate this data.

If last decade was of Google search engine, next decade will be of web map mashups.