

Deja Googled

By Sam Vaknin

Deja Googled

Sam Vaknin
palma@unet.com.mk

Economies in Transition <http://samvak.tripod.com/guide.html>

<http://groups.google.com/>

http://groups.google.com/googlegroups/archive_announce.html

The Internet may have started as the fervent brainchild of DARPA, the US defence agency - but it quickly evolved into a network of computers at the service of a community. Academics around the world used it to communicate, compare results, compute, interact and flame each other. The ethos of the community as content-creator, source of information, fount of emotional sustenance, peer group, and social substitute is well embedded in the very fabric of the Net. Millions of members in free, advertising or subscription financed, mega-sites such as Geocities, AOL, Yahoo and Tripod generate more bits and bytes than the rest of the Internet combined. This traffic emanates from discussion groups, announcement (mailing) lists, newsgroups, and content sites (such as Suite101 and Webseed). Even the occasional visitor can find priceless gems of knowledge and opinion in the mound of trash and frivolity that these parts of the web have become.

The emergence of search engines and directories which cater only to this (sizeable) market segment was to be expected. By far the most comprehensive (and, thus, less discriminating) was Deja. It spidered and took in the exploding newsgroups (Usenet) scene with its tens of thousands of daily messages. When it was taken over by Google, its archives contained more than 500 million messages, cross-indexed every which way and pertaining to every possible (and many impossible) a topic.

Google is by far the most popular search engine yet, having surpassed the more veteran Northern Lights, Fast, and Alta Vista. Its mind defying database (more than 1.3 billion web pages), its caching technology (making it, in effect, one of the biggest libraries on earth) and its site ranking (by popularity and links-over) have rendered it unbeatable. Yet, its efforts to integrate the treasure trove that is Deja and adapt it to the Google search interface have hitherto been spectacularly unsuccessful (though it finally made it two and a half months after the purchase). So much so, that it gave birth to a protest movement.

Bickering and bad tempered flaming (often bordering on the deranged, the racial, or the stalking) are the more repulsive aspects of the Usenet groups. But at the heart of the debate this time is no ordinary sadistic venting. The issue is: who owns content generated by the public at large on computers funded by tax dollars? Can a commercial enterprise own and monopolize the fruits of the collective effort of millions of individuals from all over the world? Or should such intellectual property remain in the public domain, perhaps maintained by public institutions (such as the Library of Congress)? Should open source movements gain access to Deja's source code in order to launch

Deja II? And who owns the copyright to all these messages (theoretically, the authors)? Google, as Deja before it, is offering compilations of this content, the copyright to which it does not and cannot own. The very legal concept of intellectual property is at the crux of this virtual conflict.

Google was, thus, compelled to offer free access to the CONTENT of the Deja archives to alternative (non-Google) archiving systems. But it remains mum on the search programming code and the user interface. Already one such open source group (called Dela News) is coalescing, although it is not clear who will bear the costs of the gigantic storage and processing such a project would require. Dela wants to have a physical copy of the archive deposited in trust with a dot org.

This raises a host of no less fascinating subjects. The Deja Usenet search technology, programming code, and systems are inextricable and almost indistinguishable from the Usenet archive itself. Without these elements - structural as well as dynamic - there will be no archive and no way to extract meaningful information from the chaotic bedlam that is the Usenet environment. In this case, the information lies in the ordering and classification of raw data and not in the content itself. This is why the open source proponents demand that Google share both content and the tools to access it. Google's hasty and improvised unplugging of Deja in February only served to aggravate the die-hard fans of erstwhile Deja.

The Usenet is not only the refuge of pedophiles and neo-Nazis. It includes thousands of academically rigorous and research inclined discussion groups which morph with intellectual trends and fashionable subjects. More than twenty years of wisdom and erudition are buried in servers all over the world. Scholars often visit Usenet in their pursuit of complementary knowledge or expert advice. The Usenet is also the documentation of Western intellectual history in the last three decades. In it invaluable. Google's decision to abandon the internal links between Deja messages means the disintegration of the hyperlinked fabric of this resource - unless Google comes up with an alternative (and expensive) solution.

Google is offering a better, faster, more multi-layered and multi-faceted access to the entire archive. But its brush with the more abrasive side of the open source movement brought to the surface long suppressed issues. This may be the single most important contribution of this otherwise not so opportune transaction.

[Get-Articles.com : 1000's of reprintable business and internet marketing-related articles.](#)

[Submit your article for reprint.](#)