

December 26, 2004

TECHNO FILES

At I.B.M., That Google Thing Is So Yesterday

By JAMES FALLOWS

SUDDENLY, the computer world is interesting again. The last three months of 2004 brought more innovation, faster, than users have seen in years. The recent flow of products and services differs from those of previous hotly competitive eras in two ways. The most attractive offerings are free, and they are concentrated in the newly sexy field of "search."

Google, current heavyweight among systems for searching the Internet, has not let up from its pattern of introducing features and products every few weeks. Apart from its celebrated plan to index the contents of several university libraries, Google has recently released "beta" (trial) versions of Google Scholar, which returns abstracts of academic papers and shows how often they are cited by other scholars, and Google Suggest, a weirdly intriguing feature that tries to guess the object of your search after you have typed only a letter or two. Give it "po" and it will show shortcuts to poetry, Pokémon, post office, and other popular searches. (If you stop after "p" it will suggest "Paris Hilton.") In practice, this is more useful than it sounds.

Microsoft, heavyweight of the rest of computerdom, has scrambled to catch up with search innovations from Google and others. On Dec. 10, a company official made a shocking disclosure. For years Microsoft had emphasized the importance of "WinFS," a fundamentally new file system that would make it much easier for users to search and manage information on their own computers. Last summer, the company said that WinFS would not be ready in time for inclusion with its next version of Windows, called Longhorn. The latest news was that WinFS would not be ready even for the release after that, which pushed its likely delivery at least five years into the future. This seemed to put Microsoft entirely out of the running in desktop search. But

within three days, it had released a beta version of its new desktop search utility, which it had previously said would not be available for months.

Meanwhile, a flurry of mergers, announcements and deals from smaller players produced a dazzling variety of new search possibilities. Early this month [Yahoo](#) said it would use the excellent indexing program X1 as the basis for its own desktop search system, which it would distribute free to its users. The search company Autonomy, which has specialized in indexing corporate data, also got into the new competition, as did [Ask Jeeves](#), [EarthLink](#), and smaller companies like dTSearch, Copernic, Accoona and many others.

I have most of these systems running all at once on my computer, and if they don't melt it down or blow it up I will report later on how each works. But today's subject is the virtually unpublicized search strategy of another industry heavyweight: [I.B.M.](#)

Last week I visited the Thomas J. Watson Research Center in Hawthorne, 20 miles north of New York, to hear six I.B.M. researchers describe their company's concept of "the future of search." Concepts and demos are different from products being shipped and sold, so it is unfair to compare what I.B.M. is promising with what others are doing now. Still, the promise seems great.

Two weeks before our meeting, I.B.M. released OmniFind, the first program to take advantage of its new strategy for solving search problems. This approach, which it calls unstructured information management architecture, or UIMA, will, according to I.B.M., lead to a third generation in the ability to retrieve computerized data. The first generation, according to this scheme, is simple keyword match - finding all documents that contain a certain name or address. This is all most desktop search systems can do - or need to do, because you're mainly looking for an e-mail message or memorandum you already know is there. The next generation is the Web-based search now best performed by Google, which uses keywords and many other indicators to match a query to a list of sites.

I.B.M. says that its tools will make possible a further search approach, that of "discovery systems" that will extract the underlying meaning from stored material no matter how it is structured (databases, e-mail files, audio

recordings, pictures or video files) or even what language it is in. The specific means for doing so involve steps that will raise suspicions among many computer veterans. These include "natural language processing," computerized translation of foreign languages and other efforts that have broken the hearts of artificial-intelligence researchers through the years. But the combination of ever-faster computers and ever-evolving programming allowed the systems I saw to succeed at tasks that have beaten their predecessors.

One example is question answering. Google-type search engines are fabulous at retrieving random data, but mediocre at handling subtler queries. Using Google or Ask Jeeves, you can eventually find out how many of the world's Web pages are in each of the major languages, but it's slow and frustrating compared with finding out, say, Mozart's birthplace. Jennifer Chu-Carroll of I.B.M. demonstrated a system called Piquant, which analyzed the semantic structure of a passage and therefore exposed "knowledge" that wasn't explicitly there. After scanning a news article about Canadian politics, the system responded correctly to the question, "Who is Canada's prime minister?" even though those exact words didn't appear in the article.

The Semantic Analysis Workbench, demonstrated by Eric Brown and Dave Ferrucci, showed another way of exposing latent meaning. The I.B.M. officials said the best use for this technology would be customer-support call centers: As representatives took notes on the problems people were having with their cars or computers or prescription drugs, automatic interpretation of the results would reveal useful patterns. Arthur Ciccolo, an I.B.M. strategist for its unstructured-information project, said that call centers would be the first place for new search systems to be applied. Genomic-research projects, where unexpected correlations can be crucial, might be the second. But the demonstration suggested another likely market, since every bit of sample text was a transcript of intercepted phone calls, apparently among people suspected of terrorism. ("He made two calls from Frankfurt on these dates ...") Whether these were real, I still don't know.

Salim Roukos demonstrated a system I would like to have tomorrow: an assortment of news headlines, roughly comparable to Google News, but from non-English language sources. The system automatically - and comprehensibly - translated the headlines and leads of each article. If you wanted to read more, you pressed a button and in 15 or 20 seconds had a

good-enough translation.

MR. CICCULO, the search strategist, said that in a way his team was trying to match - and reverse - what Google has achieved. "As Google use became widespread, people began asking why it was so much easier to find material on the external Web than it was on their own computers or in their company's Web sites," he said. "Google sets a very high standard for that Web. We would like to set the next standard, so that people will find it so easy to do things at work that they'll wonder why they can't do them on the Internet." How soon might this happen? He said, with a chuckle, "Well, if I could freeze what everyone else is doing, it could be in two years." The great part is, the competition won't be frozen. At least this part of the future looks bright.

James Fallows is a national correspondent for The Atlantic Monthly. E-mail: tfiles@nytimes.com.