THE GLOBE AND MAIL - 24th of August 2006

e-INSIDER

Pioneers aim to create new language for the Web

SHANE SCHICK

Pierre Lévy asks me, somewhat hopefully, whether I speak French. I do, but not enough to have the kind of conversation I want to have with him, and I say so. He understands. As someone who has spent his life studying collective intelligence, Prof. Lévy has a unique appreciation of how important it is to communicate clearly. That's why he is trying to create a language that will open up a dialogue between computers that has never existed before.

Prof. Lévy, who holds a research professorship at the University of Ottawa, has been toiling away on what he has called the Information Economy Meta Language (IEML: www.ieml.org) for the better part of 15 years. When you understand the scope of what he's trying to do, you can see why it may take even longer.

With IEML, Prof. Lévy hopes to formulate an entirely new alphabet that will express the nuances of all natural languages so that information can be indexed in a much more intuitive way. Instead of searching merely by keyword, for example, we would search by concept. "Today the search for information in computers or on the Web is made by sequences of characters. You can find information on a word like 'dog,' but it won't bring up anything related to '*chien*,' " the French word for dog, Prof. Lévy explains. IEML, on the other hand, would create an "open" search engine that would bring in more in-depth results. "You'll know exactly what algorithms are used to give the answers, what criteria is used to rank pages and even to choose the criteria."

Although the world recently celebrated the Internet's 15th anniversary, Prof. Lévy's efforts with IEML show how much work we have in front of us if we want to create a truly "World Wide" Web. Right now, it's much more of an English-speaking Web, where the rationale behind what appears at the top of the average Google search results page is less than transparent. "They don't give it because it's a commercial secret, but it creates a real limit," Prof. Lévy argues. "We are very, very far from an automatic making of hyperlinks between articles that speak about similar subjects or complementary subjects."
Well, maybe not that far. Prof. Lévy's IEML is one of several projects happening around the world to create a "semantic Web" that will add a more descriptive element to the documents we put on-line. Right now, most Web pages are built with hypertext markup language (HTML), which uses text and some extra information. HTML allows website designers to indicate how information should be presented or structured, but that's about it. Although it can ensure your name and phone number appears in the upper-left-hand corner of a "Contact Us" page, for instance, it doesn't really allow other websites or computer programs to understand that those characters are related to you, specifically.

Semantic Web tools, including Prof. Lévy's IEML, attempt to provide more meaning around information, so that it can be "read" by computers as well as human beings. This is the great vision of Tim Berners-Lee, director of the World Wide Web Consortium (W3C). In a recent interview conducted by the British Computer Society, Mr. Berners-Lee gave some insight as to why someone like Prof. Lévy, whose background is in philosophy, would take on a project such as IEML. "Physics was actually called experimental philosophy at Oxford," Mr. Berners-Lee said. "The Web is now philosophical engineering. Physics and the Web are both about the relationship between the small and the large. . . . At the moment a lot of company knowledge is held on spreadsheets and PowerPoint slides, because companies need to see summaries. But the data has lost its semantics, so it's not usable."

At the World Wide Web Conference held this past May in Edinburgh, Mr. Berners-Lee and others announced a language of their own, Sparql, that would allow software and programs to ask questions about the semantics of on-line information. Sparql is at the "candidate recommendation" phase, which means the group of companies that is trying to make it a standard part of Web development think it is ready to be implemented.

IEML and Sparql may not compete with each other, but there's no certainty that either will be rapidly adopted, either. That's partly because most companies are too busy working on traditional HTML sites that work well enough for their purposes. Semantic Web languages could make the Internet much more useful, but it's a chicken-and-egg situation. Until the majority of websites and Internet software uses the standards, the value isn't readily apparent.

One possible driver is the growing unease Internet users have about the trustworthiness of the information they research on-line. As the Web grows more collaborative through wikis and other data-sharing tools, the questions of reliability and accuracy only grow louder. Prof. Lévy thinks this helps make the case for something like IEML. "When you have language, you have lies. This will never change," he said.

"About the risk of errors, I think the best warranty that we have -- and it's not an absolute guarantee of course -- is the multiplication of sources." A semantic Web based on Prof. Lévy's "open" search engine would provide those extra sources.

The other driver, and one that will emerge over a longer period of time, is the globalization of market economies. Internet technology, in theory, allows even Canada's smallest businesses to compete with rivals in any other country. We speak of a Web "presence" when we talk about sites that can be seen by local customers and far-flung ones. But that presence may be irrelevant if it cannot be clearly understood in the native tongue of those far-flung customers, and especially if the search engines that customers use to find products and services can't derive what that presence really means, either.

Let's hope Prof. Lévy and Mr. Berners-Lee find some measure of success, because it could mean the Internet looks very different when its next major anniversary comes around. We may have a World Wide Web today in terms of reach, but in terms of an instrument for harnessing collective intelligence, it is not yet a global one. You could say that's merely a question of semantics, but that's precisely my point.

Shane Schick is editor of ITBusiness.ca.

sschick@itbusiness.ca