Grid browser finds the meaning of life

May 20 2009

(PhysOrg.com) -- A web browser that can understand technical terms in life sciences and automatically find additional resources and services has been developed by European researchers. It could lead to a new generation of intelligent search engines.

The life sciences community has built numerous databases - such as for gene sequencing and information about diseases - that are available to researchers as 'grid' services.

"Grid computing is essentially about building virtual organisations that are independent of the physical location where they reside," says Michael Schroeder of Technische Universität Dresden.

The problem is how to link those services to other scientific information found on the web. Schroeder is coordinator of the EU-funded Sealife project which has created a 'semantic grid browser' to make grid services for the life sciences much more accessible.

"We have the web on the one hand and then we have grid computing, with its many services, on the other," he says. A semantic grid browser seamlessly integrates them.

"It tries to understand what it finds on web pages, interprets this content and then links it, on the fly, to services that might be useful to the user."

A matter of semantics

The key to the Sealife browser is a ‘semantic hyperlink’ that shows up on the page to direct users to relevant services. The link is not put there by the website but by the browser itself.

How does it do that?

First, the browser needs to understand the content of the page and identify terms which could be linked to grid services. An example tested in the Sealife project is the naming of genes. Each human gene has an average of 5.5 names, Schroeder points out, but if it can be identified correctly, a link can be made to a wealth of information about that gene.

The browser must also be able to handle ambiguity. “If I see ‘Jaguar’ on a web page, what is it? Is it an animal? Is it a car? Is it the Mac operating system?” Sealife uses specialised algorithms to work out the context from other words on the page and correctly interpret the meaning.

It is still not an exact science, though. The Sealife team entered their algorithm in an international competition with 50 others to identify names of genes. They won, with an 81% success rate, though Schroeder says they have now got that up to 87%.

Background knowledge
The second challenge is the background knowledge that allows the browser to make sense of the identified terms. Such knowledge is formally known as an ‘ontology’, a systematic hierarchy of concepts and their relation to one another. Biology, with its extensive taxonomies, is an ideal field for semantic grid browsing.

“All these efforts of building hierarchical classification systems have been at the core of biology for centuries,” says Schroeder. “Biologists are used to it and there are many efforts to make information exchangeable.”

But outside the life sciences such systematic classification is not so well developed, and the Sealife project has created editors to build ontologies from published literature in any specific field of interest.

“We developed algorithms that grind through this data, identify the key concepts and then the ontology editor offers these concepts to you,” Schroeder explains. “If you agree, it then searches the web to find things that look like definitions. This whole process of building this background knowledge cannot be fully automated but you can ease the pain of doing this quite significantly.”

Different varieties of the Sealife browser build on work by partners in Edinburgh, Manchester, London and Sophia-Antipolis, as well as in Dresden. They have been tested in three scenarios: evidence-based medicine, mining of scientific and patent literature, and in molecular biology. In each case, the focus has been on infectious diseases.

Browser that understands everything?

So successful has the project been that TU Dresden has spun-off a new company, Transinsight, to exploit work done in Sealife. The company has sold semantic browsers to such major customers as BASF and Unilever and runs the GoPubMed search engine, which is linked to the respected PubMed archive of biomedical literature.

But there is no reason why a semantic browser should be confined to specialised academic areas. Could we have a browser that understands everything? Schroeder thinks that is not as far-fetched as it may seem. “The vision is to include every domain,” he says. “For example, if we were able to extract and formalise the knowledge in Wikipedia we would have this general background knowledge that covers all areas.”

Many researchers look forward to a next-generation search engine that

can understand what the user is looking for and return much more relevant results than today's engines can. "This will involve integrating information," says Schroeder, "because very often answers to questions are not provided in one document as a single statement that I can pick up by keywords.

"In the future, we will need background knowledge and this is at the core of Sealife. If we build semantic into search, and make it scaleable, then you will have the next-generation search engine."

More information: www.biotec.tu-dresden.de/sealife/

Provided by ICT Results

Citation: Grid browser finds the meaning of life (2009, May 20) retrieved 2 May 2024 from https://phys.org/news/2009-05-grid-browser-life.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.