

Ad hoc encyclopaedia for the information age

April 14 2008

Linking communities and information into a virtual digital library is the 21st century version of the Dictionaire Raisonneé. Better, they can be organised around specific topics, creating vast repositories and networks of experts around a single problem. Best of all, it can be done on demand.

In 1750, Denis Diderot convinced his publisher to support a vast enterprise, the publication of the Encyclopédie gathering all knowledge into one location.

Dozens of writers worked on thousands of articles for more than 15 years to produce the first summary of all human knowledge and, despite the labour and pains of its birth, its entire contents would barely fill one volume of a contemporary encyclopaedia.

Times have changed. And they keep on changing. The pace of discovery in the modern world is such that it is difficult for specialists to stay abreast of their own field let alone be aware of the knowledge in all other fields that may impact on their specialty.

The internet, though useful, makes us aware of our ignorance. It does not reliably fill the gap with relevant and timely information. As an information society, it is becoming increasingly difficult to see the trees for the wood.

"There's a trend in digital libraries now towards combining heterogeneous data from a wide variety of sources. This includes textual,



multimedia objects and, increasingly, sensor and experimental data, or raw data that needs to be processed," explains Donatella Castelli, scientific coordinator of the Diligent project.

Raw data allows virtual digital library (VDL) users to formulate questions that may not have been considered before. But this quantity of data poses huge processing challenges requiring digital libraries to have enormous resources, resources that are not readily available for many institutions.

The virtual digital library

But not, perhaps, for too much longer. Diligent sought to create a test bed to prove the viability of VDL infrastructure on grid-enabled technology. It would behave a little like a wiki, a Hawaiian word that means quick. Like Wikipedia – the world's most famous wiki – a VDL on grids could allow the creation of vast online data repositories from distributed computing sources.

But unlike wikis, Diligent created a system that combines digital libraries with grid computing to provide storage, content retrieval and access services and, most impressively, shared data processing capabilities.

Grids link many computers together to provide a framework for shared processing and storage capabilities. So a grid can take a big, processingintense problem, like weather prediction, and split the problem between a handful, dozens or even thousands of computers. Each only handles a tiny bit or the problem, but combined they provide a huge amount of raw power.

The power of grids is well established, and all that raw data crunching gives physicists and molecular biologists goose bumps. It is the power



behind the SETI@home project, which uses volunteers' computers to analyse cosmic signals in the search for extraterrestrial life.

But grids have never been used for virtual digital libraries, a library that exists only by the combination of data across cyberspace. It is an exciting new use of the technology. But it is not a trivial problem.

"It was very, very difficult," reveals Castelli. "There was a lot of new technology to learn [and] many of the tools we needed were only being defined as we worked on the project."

A better mousetrap

It is like inventing a better mousetrap, but the tools to do the job are only being developed as you hop impatiently from foot-to-foot, waiting for them. Then the tools get changed and you need to go back and reinvent your mousetrap.

But the hard work paid off. Diligent created an infrastructure – a system called g-Cube – and two VDLs to validate how it all works; one among the 'Earth Observation' community, the other in the Cultural Heritage community. It was a resounding success, and now these research communities have VDLs on grids serving their own needs. These are very impressive results and strain the definition of test bed as Diligent literally pushed the available technology to the limit and still came up with a working infrastructure.

They even developed advanced interface tools to set up a VDL. "We have a wizard to set up VDLs and it is very easy to use," notes Castelli.

Nonetheless, work remains to be done. "The system needs to be optimised to improve its quality of service. We need to develop a production infrastructure and deal with issues like real infrastructure



policies. We've started a new project called D4Science, and we'll be working with the Earth Observation and the Fishery and Aquaculture Resource Management Research communities", says Castelli.

Diderot's pride

Diligent has many fine achievements and prompted the interests of a wide range of groups that could usefully share resources. But the real power of the project is the enormous opportunities for fruitful collaboration that their tools will enable in the future.

Scientists, engineers, policy-makers, NGOs and other experts or stakeholders will be able to come together on an ad hoc basis to brainstorm and share relevant data around specific problems, such as disaster relief, fuel efficiency, or even apparently routine tasks like organising a conference.

Diderot, the patron of vast collaborations around a great, hugely ambitious goal, would be proud.

Source: ICT Results

Citation: Ad hoc encyclopaedia for the information age (2008, April 14) retrieved 28 April 2024 from <u>https://phys.org/news/2008-04-ad-hoc-encyclopaedia-age.html</u>

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.