

Knowledge on the Web

state of the nation and outlook for the future

Daniel **Rivers-Moore** <daniel.rivers-moore@rivcom.com>

Abstract

This paper takes stock of where we are at in the evolution of the Web, seen as a tool for knowledge sharing and navigation. It argues that any knowledge system necessarily and intimately involves people, and so a computer-based knowledge system has to be seen as including the users as well as the machines. The paper as given at the Conference will therefore include the author's perspective on current developments in the community process that is developing the current generation of fledgling knowledge systems and their related specifications.

The Web has been through several revolutions. HTML gave us pages on the Web; improved security and server-side processes gave us shopping on the Web; XML gives us data on the Web; schemas and ontologies will give us interoperable data on the Web. But to bring us knowledge on the Web requires bridging the gulf between human and computer information processing paradigms.

Knowledge is never complete, never perfect, never fully defined. Yet computers require strict formalisms and deterministic algorithms in order to function. What is needed to bridge this gap and make computers aid us effectively in our quest to access and augment humanity's rich store of knowledge, is some way of creating synergy between these two domains.

Tim Berners-Lee, in his recent presentation on the Semantic Web (at XML 2000 in Washington DC, December 2000) described the Semantic Web as being about what is machine-processable. The 'semantic test', he says, is 'If you give a machine a piece of data, does it do the right thing with it?'. But this begs the question of what the 'right' thing is, and this can only be meaningful in relation to some human purpose or intent. By limiting the perspective to the machine alone, we effectively exclude the domain of knowledge.

However, if we ask the same question of a *system*, where the system includes both

people and machines, things get more interesting. The test then becomes: 'If you give a system a piece of data, does it do the right thing with it'. The system that is the Web consists of the computers, the network connections *and the users*. The question of whether the system does the right thing is now answerable within the system domain, because the system includes people, whose wishes and intentions are part of the determination of what is 'the right thing'. The question of machine intelligence becomes one, not of whether a machine alone can be intelligent, or conscious, or have knowledge. It becomes whether machines are able to contribute to improving the intelligence or increasing the knowledge of the entire system, including its human participants.

One of the most interesting aspects of the XML phenomenon is that it has raised the profile of international efforts to reach consensus on ontologies, taxonomies, schemas, and various other mechanisms for giving formal structure to aspects of human understanding. The fact that XML has come into being as a universal syntax for the exchange of data between computer systems has led not only to a new generation of *software*, built on the basis of open interchange formats and, in many cases, open source code, but also to a new generation of human effort to agree upon clear definitions and representations of aspects of shared human knowledge.

The work of a standards committee agreeing on a DTD, a schema, an ontology or a taxonomy, is 'part of' the system that is the Web. The results of this work feed directly into the machines and are processed and worked on by machines, then fed back to humans, and the cycle of sharing and augmenting knowledge continues. A number of recent initiatives recognise explicitly the need to be aware both of the human and the machine players in the knowledge process. Doug Lenat's Cyc initiative involves large numbers of people working interactively with computers to refine their reasoning algorithms by inputting information about common-sense inferences and hypotheses that people use when they themselves perform reasoning. The NewsML specification includes formalisms for stating which person and/or system has allocated a given piece of metadata and how confident they are in its accuracy. Topic Maps include formalisms for identifying resources that represent 'non-addressable subjects', which are concepts meaningful to humans but unable to be understood directly by machines, though the machines can manipulate them and draw inferences from them and the relationships between them within the topic maps formalism. It is these kinds of initiative, involving formalisms that explicitly bring

together human and machine intelligence and processing in ways that recognise their differences and exploit the strengths of each in optimal ways, which will bring us the next generation of knowledge systems.

The work is already well under way, and is evolving rapidly. It will be exciting to see it bear fruit over the next few years. Given the speed with which this effort is unfolding, any snapshot of the current state presented at the time the printed proceedings of the conference goes to press, will be severely out of date by the time the Conference takes place. The live version of this paper will include a brief summary of recent developments, and will present the author's perspective on the current state of play and likely future directions.

Biography

Daniel **Rivers-Moore**

Director of New Technologies

RivCom

Swindon

United Kingdom

Email: daniel.rivers-moore@rivcom.com

Daniel Rivers-Moore - Daniel Rivers-Moore is Director of New Technologies at RivCom, a consultancy and services company specializing in helping businesses adopt XML technologies to meet their information management and distribution needs. He has a decade of experience helping organizations gain business benefit from the adoption of standards-based publishing solutions and been actively involved in the development of the XML family of standards for many years, having been a member of the original XML Special Interest Group, joint project leader of the STEP/SGML harmonisation initiative under ISO, software development lead in the recently completed European XML/EDI Pilot Project. Daniel is a Founder Member of TopicMaps.org and a member of the XML Topic Maps Authoring Group. He is the editor of the NewsML standard and chair of the provisional steering group of KnoW (Knowledge on the Web), a collaborative initiative aimed at fostering the development of Web technologies for the sharing and interchange of knowledge.