The *Discussion Forum* provides a medium for airing your views on any issues related to the pharmaceutical industry and obtaining feedback and discussion on these views from others in the field. You can discuss issues that get you hot under the collar, practical problems at the bench, recently published literature, or just something bizarre or humorous that you wish to share. Publication of letters in this section is subject to editorial discretion and company-promotional letters will be rejected immediately. Furthermore, the views provided are those of the authors and are not intended to represent the views of the companies they work for. Moreover, these views do not reflect those of Elsevier, *Drug Discovery Today* or its editorial team. Please submit all letters to Joanna Owens, Acting News & Features Editor, *Drug Discovery Today*, e-mail: Joanna.Owens@elsevier.com

The semantic web and biology ▼

The vision of a semantic web was first put forward in the late 1990s by Tim Berners-Lee, the father of the World Wide Web (http://www.w3.org/2001/sw/). Since then, the idea of a semantic web, loosely defined as a web with machine understandable content, has grown to become a significant driving force in the development of standards for describing web content with the use of XML (eXstensible Mark-up Language)-based languages. Several concepts related to this vision are likely to be useful for bioinformatics, in particular for the organization and discovery of knowledge from life science documents. Not only has the XML format largely been adopted by publishers and citation databases (e.g. MEDLINE) for the encoding of documents, but XML-specified ontologies are now also used for defining and describing data in biological databases. The efficient use of an ontology requires wide acceptance of the terminology of concepts and the relationships between concepts described therein. Broad acceptance generally requires either a central defining authority or a large concerted effort in deriving a consensus over the global view.

A central question is whether the different scientific communities that are involved have the will to speak the same language. Even in science, which is generally much less controversial than politics, the necessary level of agreement will take time to achieve. The extent of present knowledge that has been recorded within biology, as refereed literature and its digitally available summary MEDLINE, is organized as free-text. Several efforts, including the Gene Ontology Consortium (http://www.geneontology.org/) and our own effort PubGene (http://www.pubgene.org), are in development to move science toward the semantic web vision. However, such efforts are severely hindered by the lack of digitally usable standards, even for simple items such as gene and protein names, where a name is often imprecise, changes over time or has several synonyms pointing to different entities. Some progress is being made but significant strides forward need new strategies in many disciplines. We suggest two specific areas of bioinformatics where semantic web ideas could prove useful:

(1) Knowledge representation and systems biology. Information stored as facts and relationships could support simulation of larger systems and complex interactions. This requires ontologies and logic. The Gene Ontology effort represents a first demonstration of what could be achieved in knowledge representation in the biology domain. With the advent of increasingly advanced high-throughput data-harvesting techniques, biology is rapidly being transformed to knowledge handling and conceptually to mathematically based systems biology.

(2) Knowledge extraction. It is an important but difficult task to extract existing knowledge. Semantically annotated (tagged) documents could significantly ease the extraction of facts and/or relationships, and would provide a way to increase the precision and specificity of digitally represented biological knowledge. Today, most scientific journals provide digital versions of their content. However, semantic content annotation has not been achieved and would be difficult to realize with the present barriers to literature access. The advent of the Open Archive Initiative (http://www.openarchives.org/), which aims to free the full body of scientific literature from the Gutenberg restraints, could provide opportunities for obtaining useful tags, possibly in association with publishing houses. To our knowledge, no literature annotation initiatives have been declared so we take this opportunity to declare the need for such an initiative.

With the advent of such work, the hope is that the semantic biology web will become more than an idea for the future.

Tor-Kristian Jenssen and Eivind Hovig
Dept of Tumor Biology
Institute for Cancer Research
The Norwegian Radium Hospital
0310 Oslo, Norway

Promises of text processing: natural language processing meets AI ▼

We were pleased to see the timely review by Mack and Hehenberger on methods for analyzing biomedical

1359-6446/02/$ – see front matter PII: S1359-6446(02)02458-3