The Knowledge Engineering Review, Vol. 21:1, 93–95.
doi:10.1017/S0269888906000816 Printed in the United Kingdom

Book Reviews

Spinning the semantic web edited by Dieter Fensel, James Hendler, Harry Lieberman and Wolfgang Wahlster, MIT Press, 479 pp., $23, ISBN 0-262-56212-X
doi:10.1017/S0269888906210816

The concept of the semantic web is a noble one. In a nutshell, the idea is that adding semantic meta-information to the existing content of web pages—that is adding some kind of explanation of what the information on a webpage *means* rather than just how to display it—it will be possible (in the words of the subtitle of this book) to:

(bring) the World Wide Web to its full potential.

What is this full potential? Well, the usual answer is to give an example of the kind of thing that the semantic web will make it possible to do, and here is such an example.

Consider that you are a young, untenured (and therefore both very busy and very eager to impress your colleagues) faculty member who has just been assigned a course to teach, a course that covers material that you are not completely familiar with. Well, if you were not so young and keen, you might feel you could get away with picking a random textbook on the subject and teaching from that. But you are not. You have to impress the students so they will give you good feedback that will, in turn, impress your tenure committee. So you go out and research the subject in great detail, reading lots of background material, and come up with a wide range of different views and approaches to give your students. Of course you use the Web to do this—you go to your favorite search engine and use that to pull down pages and pages of information that are more or less (often less) relevant, and you wade through these, building up a refined list of sources that you then go and read in detail. You order up sample textbooks from publishers' websites, and of course you look at how your peers have covered the same material, browsing their websites and reading their blogs.

Much of the work that you have to do would have been much harder, even impossible, without the Web, but there is still a lot of hard slog. The promise of the semantic web is that it will remove a lot of that slog.

The argument is that the reason investigating topics on the Web can be so much of a slog is that the Web itself is a vast ocean of text and multimedia, with no structure, and precious little information that can be used by search engines beyond the text itself. It is very hard to accurately classify a book from a sample of the text it contains, much easier by looking at chapter headings and easier still if you have access to some kind of catalog that includes the book and identifies its contents explicitly, placing them in a taxonomy of all possible subjects. The idea of the semantic web is, in short, to provide that explicit identification and the taxonomy.

If such an infrastructure were to be provided, then your job in generating material for your new course would be much simplified. Of course, you would still have to read the material and prepare the classes, but the collection of the material would be simplified. Rather than using a search engine and then wading through the mass of links it returns in order to find the material you need to read, you would use some kind of advanced personal assistant agent. This piece of software would read the meta-information provided by the semantic web, and use it to home in on precisely the right books to order—and would be able to order them, being able to distinguish the publisher's page and knowing the difference between ordering a sample textbook and buying the book at an online store—to identify what other sources are worth downloading, and to determine the blog entries that are relevant to the course.

This, then, is the vision of the semantic web—or at least my version of the vision—and it is a vision that is slowly turning into a reality. In the past few years there has been a convergence of opinion on the underlying infrastructure that will be provided to support the semantic web, and much activity in, for example, the generation of ontologies, necessary to provide 'meaning' (in some sense) to web pages that can be manipulated by machine by providing a taxonomy into which the terms that they contain can fit.

Spinning the semantic web is not really the place to find out about this recent work. The book is a reprinting of a book that was originally published in 2003, and which is based[1] on material presented at a conference in 1999. Of course, this in no way invalidates the material or the book. Many of the papers from that 1999 conference, at least in the form in which they have made it into *Spinning the semantic web*, are still valid, and it is interesting to look at them with the perspective of more recent work, especially interesting because they are from relatively early in the history of the semantic web. This makes the book, if not essential, a useful addition to the collection of anyone interested in the semantic web, and probably the second book that a newcomer to the area should buy after (Antoniou & van Harmelen 2004) (which, unlike *Spinning the semantic web*, is intended as both an introduction and a textbook).

Reviewed by Simon Parsons
Brooklyn College, City University of New York, USA

Reference

Antoniou, G and van Harmelen, F, 2004, *A Semantic Web Primer*. Cambridge, MA: MIT Press.

Advances in minimum description length by Jae Myung and Mark A. Pitt, edited by Peter D. Grünwald, MIT Press, 444 pp., $50.00, ISBN 0-262-07262-9

doi:10.1017/S0269888906220812

Induction is vastly important in science. Scientists carry out experiments, collect data and then try to establish—by induction—theories that underlie the data. These theories can be considered as general rules that, depending on your point of view, either generate the specific data items, or predict what those data items will be. In particular, scientists seek those theories that explain the data in a way that is most elegant (which is often the same thing as saying they explain the data in the most economical way). Machine learning, as a discipline, is often concerned with exactly this problem of induction,[2] seeking algorithms that will do the theory generation.

When establishing such algorithms, a key problem is what takes the place of the scientists' notion of elegance. How is it possible to determine, in the form of an algorithm, which of two theories is most economical? One possible answer to this question is to use the minimum description length (MDL) principle. Broadly speaking, this principle says that the best theory (expressed in terms of a generative model) for a set of data is the one that allows the greatest compression of the data (which is the same thing as saying the model that explains more of the regularities in the data than any other model).

This book provides an excellent starting point for anyone who wants to know about MDL, despite the fact that the book started out as a collection of papers at a workshop at the Neural

[1] As Tim Berners-Lee points out in his foreword, the bulk of which itself is a reprinting of an article written in 1997.

[2] Some forms of machine learning, evolutionary methods for example, are more about inventing new artifacts and so have more to do with deduction than induction.

Information Processing Systems conference.[3] This is because the editors, and in particular Peter Grünwald, have thoughtfully provided some very accessible introductory material designed to get the novice up to speed before they tackle the more advanced papers. The result, while not as useful to the wider machine learning community as a comprehensive textbook on MDL, is sure to be widely read until such a textbook comes along.

Reviewed by Simon Parsons
Brooklyn College, City University of New York, USA

Imitation of life: How biology is inspiring computing by Nancy Forbes, MIT Press, 176 pp., $8.95, ISBN 0-262-06241-0
doi:10.1017/S0269888906230819

Anyone active in computer science over the last few years cannot help but have noticed the growing overlap between biology and computing. Part of this is because, as some have put it, 'computing is the hand-maiden of the sciences'. That is, as the quantities of data gathered increase, more researchers across the full span of the sciences are using techniques developed in computer science, techniques like unsupervised machine learning, to help them make sense of the results of their work. This effect has led, for example, to much of the field of bioinformatics. However, the growing overlap is not just due to techniques from computer science being deployed in the service of biology—there are also a number of ways in which biology is contributing to computer science. These ways are typically ways in which ideas from biology have inspired new approaches to computer science, and it is these ways that are the subject of *Imitation of life*.

Some of the ways that computer science has drawn from biology will be familiar to many readers—they include evolutionary computation, for example, and artificial life models like ant colony optimization.[4] Other techniques, like molecular self-assembly and DNA computing, are less likely to be well known. All are fascinating, and it seems likely that many will be important to the future of computer science.

The chance to get a look at the future of computer science, or at least the future of some significant part of computer science, is a good reason to read *Imitation of life*, especially considering that it is priced as a mass market paperback. However, do not expect a scholarly text. This is a book written to give a broad overview of an area, with some pointers for further reading, rather than the kind of detailed monograph that MIT Press usually publishes. That does not make it a bad book—it makes no pretence to be anything other than it is, and it is a perfect book to use as the basis of an undergraduate class on the interface between biology and computer science, but it does mean that anyone interested in the detail of biologically-inspired computing is going to have to look elsewhere.

Reviewed by Simon Parsons
Brooklyn College, City University of New York, USA

References

Dorigo, M and Stützle, T, 2004, *Ant Colony Optimization*. Cambridge, MA: MIT Press.
Reynolds, CW, 1987, Flocks, herds, and schools: a distributed behavioral model *Computer Graphics* **21**(4), 25–34.

[3] I say 'despite' since most books that start life as a collection of papers from a workshop, including those I have edited myself, are usually aimed at experts in the area rather than novices.
[4] At least I thought the book mentioned ant colony optimization but, on checking back through it, I find that though it covers related models like Craig Reynold's Boids (Reynolds 1987), it does not cover work on ant colonies (Dorigo & Stützle 2004).