

Clark, K., Parsia, B. and Hendler, J. (2004). *Will the Semantic Web Change Education?* Journal of Interactive Media in Education, 2004 (3) [www-jime.open.ac.uk/2004/3]

Published
21 May 2004
ISSN: 1365-893X

Will the Semantic Web Change Education?

Kendall Clark, Bijan Parsia and Jim Hendler

Abstract:

To say that the Web has affected many societies and cultures is to understate its impact along several dimensions. The Web is a technology which not only affects, but in some sense encompasses societies, cultures, and certainly institutions. Higher education -- at least in the cluster of ways in which it is practiced in the US, the EU, and Japan -- is one such bundle of social institutions affected and encompassed by the Web.

While it is possible to overstate or mis-state the Web's effect, whether on higher education or on other institutional clusters, the encompassing reach of the technology, used in every country on Earth by literally tens of millions of users, makes it clear that the Web truly has a revolutionary effect. However, exploring what the Web has affected and continues to effect is a necessary element of any accurate estimation of how the newly emerging Semantic Web may, in its turn, effect societies, cultures, and institutional clusters like higher education.

Commentaries:

All JIME articles are published with links to a commentaries area, which includes part of the article's original review debate. Readers are invited to make use of this resource, and to add their own commentaries. The authors, reviewers, and anyone else who has 'subscribed' to this article via the website will receive e-mail copies of your postings.

Kendall Clark, Bijan Parsia, and Jim Hendler, *Maryland Information and Network Dynamics Laboratory, University of Maryland, <http://www.mindlab.umd.edu>*

Page 1

1. Hypertext: Beyond Text

There are many models of hypertext, each of which has various richnesses and affordances and degrees of expressivity. The Web's hypertext model is a relatively impoverished one, especially as compared with models which include bidirectional links, resource versioning, default genre and document structures, guided paths through resources, resource and resource-part annotations, and so on. [The Dexter Hypertext Reference Model¹; Gopher²; Xanadu³; Serving Information to the Web with Hyper-G⁴] Despite, or some would argue because of, the Web's simple model, it has had a greater impact than any of the more expressive technologies that preceded it.

The primary reason that the Web has been so much more widely accepted is that it was designed to allow two key capabilities: openness and scalability. The first resulted from a conscious design decision that one should be able to link to other people's resources without any need for permission, and that one could make a resource available on one's own server in such a way that others could link to it. The second is achieved by the Web's capability to take advantage of a network effect: If I create something, and you create something, we can point to each other's resources, rather than having to duplicate resources. This is the Web in the World Wide Web -- and the network effect is where most of the power comes from -- it is often easier to create content (with pointers) on the Web than to duplicate that information elsewhere. Thus, the one of the fundamental design goals of the Web was to use a relatively impoverished hypertext model that was open and scalable, rather than to use or develop a more expressive hypertext model that was more restrictive.

1.1 Academic Practice

One way to judge the impact of the Web on higher education is by judging the distance between text and hypertext, particularly with regard to academic practice. Historically the various printed page technologies and the university, as well as the research library, have been co-evolving, interdependent institutions.

¹ <http://citeseer.nj.nec.com/balasz94dexter.html>

² <http://www.zvon.org/tmRFC/RFC1436/Output/>

³ <http://www.xanadu.com.au/>

⁴ http://www.igd.fhg.de/archive/1995_www95/proceedings/papers/105/hgw3.html

Because text is, in one sense, a static cultural artifact, dynamic cultural institutions, like the modern university-situated research library, were developed, out of pre-existing institutions, of course, in order to nurture -- that is, to organize, categorize, preserve -- texts. Texts can point, in a variety of ways, to other texts, a point often made by various post-structuralist theorists who talk of intertextuality as a fundamental cultural force. (See, for example, the work of Roland Barthes⁵.) In some ways intertextuality is the precursor of modern, digital hypertext and hypermedia. And, more to the point in this context, there are various kinds of scholarly apparatus designed to create links between scholarly texts; in the modern era, the footnote and endnote are the primary means.

But since these links are conceptual, rather than implementational, the university needs an institution that serves as a nurturing repository of all such relevant texts, such that enacting or activating these conceptual links becomes a matter of physically manipulating -- locating, paging through, reading -- other texts. Many fields of academic inquiry in the modern research university today are still centered around these practices, which are largely unchanged over the past two or three centuries. In many fields a common scholarly practice is to enter an area of study by finding a text which serves as a guide, and then by following all of its various links to other texts, to journal articles, to conference proceedings, and the like.

In order to make that a realistic practice, such that it could underwrite and support other scholarly practices which together form academic, inquiring communities, the university took on the role of a cultural and scholarly repository of texts, together with various attendant practices: information space organizational schemes (Dewey Decimal, Library of Congress, etc.); scholarly resource sharing schemes (inter-library loan); preservation of non-scholarly but otherwise formal or official texts (for example, federal government and other public interest archives).

These social practices are constrained by the technologies (printed books, libraries, card catalogs, footnotes, endnotes, indexes) which make them possible and call them forth; likewise, these technologies are constrained in that they are used in *these* social practices and not in others. A parallel sort of relation between social practices, embedded in communities of inquiry, and technologies has been developing for as long as the Web has had a presence in higher education. We should expect, therefore, that the Web may make possible different modes of scholarly practice and discourse, including different modes of publication, citation, and information organization, because it is based on a technology -- distributed, decentralized hypertext -- with a different set of affordances than printed text.

⁵ <http://we.got.net/~tuttle>

Indeed, this is exactly what we see happening on the Web. Some of the emerging practices include less costly forms of academic publication, including Web-only journals, virtual conferences, purely ad hoc, geographically distributed study and affinity groups, distance education, preprint paper and research sharing patterns, personal scholarly publishing, the diminishment of journal and press editors as arbiters of academic standards and taste. (In addition, collaboration on the Web, and the reach of the Internet technology that supports it, has led to a proliferation of other collaboration technologies like Internet Relay Chats and Instant Messenger Services, but we do not address their effect in this paper.)

Let's consider for a moment a very concrete example. The foot or endnote is a significant element of scholarly discourse. But the Web's hypertext model actually contains no concept which is strictly equivalent to the printed page, at the foot of which one might add a note. One of the discursive differences the move from footnotes to hypertext links makes possible is a more indirect scholarly style of expression. In a scholarly text, replete with footnotes, one directly expresses the linkage between the present text and another one. These direct expressions run from the concise -- a bare footnote, "See also", "Cf." -- to the verbose -- "As the influential C.P. Snow argued in his landmark essay, ..." -- but in each case the linkage is only peripherally related to the text itself. The linkage itself cannot be easily associated with an arbitrary sequence of text, as it can in a Web publication.

Thus, rather than creating concise or verbose linkage markers, scholarly discourse on the hypertextual Web is able to interleave and interweave such linkages within the main text itself. We can -- arbitrarily or elegantly -- make any text, within a scholarly hypertext, link to any other Web resource (or even to named parts, or fragments, of other Web resources). That difference in technology, which is admittedly quite subtle, calls forth and makes possible a change in the way that scholarly discursive practices are created and enacted.

This shift in the style of citation, by itself, would not be as significant without the enormous amount of material published on the Web and the growing ubiquity of Web use and expertise. Consider two examples. First, CiteSeer⁶, the Scientific Literature Digital Library, is an interesting example of the Web as a helpful supplementary system to established, existing academic practice. Using an Autonomous Citation Index, CiteSeer takes an existing academic practice, the citation, and supplements it with the Web by treating citations as hypertextual links. When searching for some relevant scientific literature at the CiteSeer site, citations of papers are turned into hypertext links, with CiteSeer indexing and providing some

⁶ <http://citeseer.nj.nec.com/>

modest keyword metadata services as well. Thus, without any additional effort on the part of researchers and scholars -- beyond, that is, publishing papers on the Web -- CiteSeer turns the research literature of a scientific field into a kind of hypertext, through which scholars and other interested parties may wander in the pursuit and support of their own research interests and projects.

A second example is the arXiv.org e-Print archive⁷, a site which archives scholarly articles in physics, mathematics, nonlinear sciences, computer science, and quantitative biology. The focus of arXiv.org is to make papers in these quickly moving fields available as quickly and as easily as possible, in advance of, but not as a substitute for, the costly and time-consuming process of peer review. There is little doubt that peer review, in some form, is absolutely essential to progress in fields of academic inquiry. But, as it is most often practiced in many fields today, peer review is essentially unchanged since the post-WWII generation, it can hinder fields undergoing rapid or exploratory advances., and it may be reconfigured to more ideally fit contemporary realities (See P. Ginsparg, *Winners and Losers in the Global Research Village*⁸). By providing an initial clearing house for (primarily) physics and mathematics papers -- which are very often submitted simultaneously to both arXiv.org and to a peer-reviewed journal -- arXiv.org supplements existing academic practice by providing a ubiquitously reachable archive of relevant materials.

For the general audience, the Web has replaced the encyclopedia as the entry point (and more) into arbitrary topics of inquiry. Aside from classic reference materials -- encyclopedias, dictionaries, and scholarly paper indexes -- republished on (and enhanced for) the Web, and even aside from standard scholarly material -- articles, monographs, proceedings -- published or republished on the Web, there are masses of interconnected *lightweight* commentary, both individual and collaborative, freely available, often easy to find, and typically trivial to create. Lecture notes, class notes, email exchanges, presentation slides, syllabi and reading lists, study questions and answers -- all of these were once primarily shared only via direct personal contact, with only a small fraction of this marginalia of academic life published in collections and treatises. As the Web becomes a primary medium of academic and pedagogic interaction, all that was once ephemeral, parochial, and largely hidden becomes more permanent and universally available.

⁷ <http://arXiv.org/>

⁸ <http://arxiv.org/blurbl/pg96unesca.html>

1.2 Pedagogic Practice

Aside from a trickled down effect from the ongoing transformation of academic practice, the Web has directly changed education, most obviously in the way classes are organized and taught. There are innumerable classes *about* the Web, from simple "how to browse the Web and write HTML" to complex Web-based information design. Many schools now teach advanced web search techniques, as opposed to physical library search methods, to junior high school students. There are also classes which use the Web to disseminate course material and collect assignments. Interestingly, classes about the Web are not a subset of classes that at least minimally use the Web. There are classes which significantly incorporate the Web, e.g., where course materials and assignments aren't merely transmitted by the Web, but are enduring, ongoing Web sites, or where significant class discussion occurs in Web based or reflected fora. And, finally, there are classes which are conducted entirely on the Web without the requisite for physical presence. As with many technological revolutions, in the early days of the Web, any class which wanted to make significant use of the Web had to also be a class about the Web, at least to the extent of providing minimal sufficient training in browsing and publishing Web pages. As Web literacy spreads, this portion of general-topic Web-using classes has been reduced to dealing with idiosyncratic Web applications used by the class (say, a custom discussion board or WikiWikiWeb) and tips on finding subject-specific good information on the Web.

The transition from academic communities focused on text to academic communities also focused on hypertext has matured and borne real fruit (reference). We suggest that with the next transition, from hypertext to knowledge representation on the Semantic Web, that new social practices and institutions are likely to appear.

2. Semantic Web Changes

If the Semantic Web means anything, it means changing the Web's infrastructure such that information exchanges between computers alone become as ubiquitous, cheap, and easy as exchanges between humans, mediated by the Web, are already. One vital goal, however, is to make inter-machine exchanges possible without doing permanent damage to the ecology of the Web: inter-machine exchanges are not meant to replace or supplant inter-human ones, merely to supplement them.

2.1 Beyond Hypertext

So far we've argued that the Web's hypertext model, though expressively impoverished in comparison to other hypertext models, has been widely successful in and across a great many parts of society, including higher education. The differences between text and hypertext have called forth and made possible interesting differences in the way academic communities constitute themselves and enact their scholarly practices.

The success of the Web suggests, however, that the network effect is more important than the expressivity of the hypertext model. In some sense the fact that millions of people are engaged in a wide diversity of interesting projects and activities using the Web overwhelms the fact that the Web's hypertext model is relatively inexpressive. It is rather astonishing to explore the rich webs of signification and linkage which have been created on the Web with only the lowly, unidirectional link. The algorithm which powers Google, Page Rank, is based on the unidirectional link, as well as some assumptions, which turn out to be mostly correct, about popularity and relevance. That is, we end up getting a lot of power out of a relatively inexpressive hypertext model, with its untyped, unidirectional link, and the network effect.

Thus, as we begin to see some of the building blocks of the Semantic Web put into place, we anticipate that there will be new practices and institutions that are called forth by these new technologies (just as these new technologies are themselves being called forth by a different set of practices and institutions). As we've focused so far on the transition from text to hypertext, we'll now take up the transition from hypertext to hypertextual knowledge representation or *hyperkrep*.

2.1.1 RDF as a Foundational, Enabling Technology

There are at least two technologies, in addition to the existing Web infrastructure itself, which are key to the Semantic Web: RDF⁹ and OWL¹⁰. RDF, the Resource Description Format, which is an XML vocabulary, is an assertional knowledge representation language, allowing anyone to say anything about anything. How does it accomplish this? The first point to make is that RDF is based on a formally specified semantics, grounded in model theory.

⁹ <http://www.w3.org/RDF>

¹⁰ <http://www.w3.org:2001:sw:WebOnto>

The main idea behind RDF is that knowledge can be represented as a graph of directed, labeled arcs; one makes assertions about a thing by means of associating subjects and objects by way of predicates. Put the other way around, RDF graphs are full of things called "triples", which are three-tuples, or assertions, containing subject, predicate, and object terms. What makes RDF particularly useful in the context of the Web and the Semantic Web is that the value of these terms -- subject, predicate, object -- may each be a URI. "URI" stands for Universal Resource Identifier; it is the term most commonly used for what was formerly called a URL or Universal Resource Locator.

Let's take a concrete, if contrived and simplistic example. You are a philosopher of science and a member of the (mythical, as far as we know) C.P. Snow Society. The society maintains a presence on the Web at <http://www.cpsnow.org/>, which includes a few notable resources: a page about C.P. Snow himself, <http://www.cpsnow.org/cpsnow/>, and a page about his famous little book, *The Two Cultures and the Scientific Revolution*, <http://www.cpsnow.org/two-cultures>. Imagine, further, that you would like to represent some knowledge; for example, "C.P. Snow wrote a book called *The Two Cultures and the Scientific Revolution*".

How might you go about encoding some bits of knowledge such that Semantic Web agents could interpret them. Let's begin by rewriting our simple sentence in a longer but slightly more literal form: "There is a book that is titled 'The Two Cultures...' and its author is 'C.P. Snow'". More awkward, more wooden, and more verbose, but this version of our sentence is semantically equivalent.

How might we encode this strange set of sentences in RDF? That is, how might we encode it as a set of three-tuples of the form (subject, predicate, object)? First we will give the encoding, then we will explain it:

```
(http://www.cpsnow.org/two-cultures, rdf:type, cpss:book)
(http://www.cpsnow.org/cpsnow, dc:author,
 http://www.cpsnow.org/two-cultures)
(http://www.cpsnow.org/two-cultures, dc:date, "...")
(http://www.cpsnow.org/two-cultures, dc:title, "The Two Cultures and
the Scientific Revolution")
```

What have we done here? First, we've said that the web resource, <http://www.cpsnow.org/two-cultures> is (or, more accurately, represents a thing which is) a book. The term form "xxx:yyy" is a kind of abbreviation, known as an XML

qualified name or "qname". It means that we're using a term from an existing vocabulary or set of terms, rather than making up our own. The RDF specifications from the W3 Consortium, specify that "rdf:type" is a term which means, roughly, "is-a". You can read that first triple as, roughly, "the web resource, <http://www.cpsnow.org/two-cultures>, is of the type cpss:book". Perhaps the CP Snow Society doesn't know or approve of existing sets of terms which define "book", so it's defined its own, using the prefix "cpss".

The second triple can be read as saying that "there is a web resource, <http://www.cpsnow.org/cpsnow>, which is or represents the entity which is the author of another web resource, <http://www.cpsnow.org/two-cultures>". We know this second web resource, the one in the object position in the second triple, is a book, because that was the assertion made in the first triple. Putting these together, we've now said that there is a book, identified by such-and-such a web resource, which was authored by some entity, identified in turn by such-and-such a web resource.

Lastly, the two final triples says that there is a web resource, which we now know to be a book, that has the title "The Two Cultures..." and a specific date. Rather than making up our own terminology for date and title, we use the well-known *Dublin Core* meta-data standard using its common qname prefix "dc:" to denote it.

That's not so difficult. We've expressed a helpful bit of knowledge, and we've done so in a way that can be easily turned into a format that Semantic Web agents can understand -- a format backed by a rigorous, formal semantics. Now, suppose we want to say a bit more? Suppose we want to say a bit more about C.P. Snow, the natural person, himself? We can start to see a bit of the promised power of the Semantic Web by taking this question a little further.

Even though all of the web resources discussed so far are mythical, there is a good chance that you have been assuming a particular thing about them, namely, that if there were such resources on the Web, what you would find when you used your web browser to visit them would be some HTML. That's a perfectly reasonable assumption, given the past 10 or so years of history and experience with the Web. That is, if you pointed your browser at <http://www.cpsnow.org/two-cultures> you would expect to see a page describing the book in HTML.

But, in another sense, it's dead wrong. And here's why. The existing Web works because web resources represent (and, sometimes, just are) interesting things in the world. And these resources, standing in for (or being) interesting things in the world, often point to other resources, which in turn stand in for (or are) other interesting

things in the world. Imagine, then, that instead of finding HTML, meant for human consumption, at those web resources, one could find RDF meant for machine consumption. So, instead of (or in addition to) finding an HTML page giving the biographic details of C.P. Snow, one may find an RDF document which includes the following triples:

```
(http://www.cpsnow.org/cpsnow, rdf:type, foaf:Person)
(http://www.cpsnow.org/cpsnow, foaf:name, "Charles Percy Snow")
(http://www.cpsnow.org/cpsnow, foaf:img,
 http://www.cpsnow.org/cpsnow.jpg)
(http://www.cpsnow.org/cpsnow, foaf:gender, "male")
```

You can read the first triple as saying, roughly, that "there is a web resource, <http://www.cpsnow.org/cpsnow>, which represents a natural person". In this case we're using the term `foaf:Person`, which means we're using the term "Person" drawn from a vocabulary called "Friend of a Friend", a common way to represent information about natural persons on the Semantic Web. Next, "there is a web resource, which represents a natural person, that is named 'C.P. Snow'"; third, "there is a web resource, which represents a natural person of the male gender".

Note the network effect is once again present! The CP Snow society let the Dublin Core folks define facts about publication metadata and let the Friend of a Friend vocabulary define facts about people. DC and FOAF, in turn, may link to other documents that represent other types of information and so on and so forth. Instead of every document making up its own representation, they are linked into a Web of semantic representation.

One may quickly see, or so we think, that if a great many affinity groups within higher education -- study groups, learned societies, scholarly conferences and colloquiums, departments, colleges, seminars, groups of students, groups of students and a faculty member, and so on -- develop in the next five years even one hundredth as many RDF resources as they have created HTML resources in the past five years, then the Semantic Web will become a thing very rich in knowledge, that is, in knowledge discoverable and consumable by machines and agents.

2.1.2 Adding a Web Ontology Language (OWL)

OWL is a newly developed ontology language for the Web. An ontology language is a means by which one can formally describe a knowledge domain, with the goal of

enabling computers to provide various kinds of reasoning services about that domain, and about the knowledge described by an ontology for that domain. In our current, technical usage, an ontology is a formal specification of a knowledge domain: what individuals and classes of individuals there are in that domain, the relationships which obtain between these individuals and classes, their proper and apparent parts, and so on. Thus, using OWL one can formally specify a knowledge domain, describing its most salient features and constituents, and then use that formal specification to make assertions about what there is in that domain. You can feed all of that to a computer which will reason about the domain and its knowledge for you. And, here's the most tantalizing bit, you can do all of this on, in, and with the Web, in both interesting and powerful ways.

Two brief points: First, we all spend some amount of our brain power -- almost entirely without consciously knowing that this is what we are doing -- dealing with informal, implicit ontologies. In order to act meaningfully at all within particular social contexts, we need to have understood something roughly like an ontology of that context. In any situation or context there will be features which we attend to, because they just are the salient features of that context, and an even larger number of things about the situation which we do not attend to, which we cannot even call features, because they are the background noise against which salience emerges. Second, unlike humans, computers can only provide reasoning services over a knowledge domain because the domain and the knowledge have been formally and rigorously specified in advance and because some human has implemented various reasoning algorithms in a way which that computer can apply.

From these two points we may be able to conclude that ordinary people, with the right support and motivation, can learn to use the formal tools of computerized ontology languages, like OWL, to represent the things which they already know in a way which computers can then reason about, as a supplement and aid to human interests. It's worth noting that the alternative, expecting the computer to understand and reason with human concepts and language, is far beyond the current state-of-the-art, if achievable at all.

So far nothing we have said about ontology languages and reasoning systems is specific to OWL as an ontology language for the Web. However, OWL has been specifically crafted out of its Webbish forerunners, particularly SHOE¹¹ and DAML+OIL¹², to take advantage of some of the interesting things about the Web.

¹¹ <http://www.cs.umd.edu/projects/plus/SHOE>

¹² <http://www.daml.org/>

OWL is intended to be an ontology language that has the following features: it should operate at the scale of the Web; it should be distributed across many systems, allowing people to share ontologies and parts of ontologies; it should be compatible with the Web's ways of achieving accessibility and internationalization; and it should be, relative to most prior knowledge representation systems, easy to get started with, non-proprietary, and open. In short, OWL was based on the same principles we mentioned about the Web itself much earlier in this discourse -- openness and scalability to allow a network effect.

Insofar as OWL accomplishes or will accomplish these goals, it will do so by virtue of the fact that it was designed by a collection of Knowledge Representation and Web experts, with the explicit goal of making a formal knowledge representation (KR) language work on the world's first globally distributed hypermedia system. This is a relatively new thing to aim at in the history of KR systems. In some ways, the OWL Working Group (WG) is among the most ambitious of the W3C's many WGs. It is often said of W3C WGs that they are not meant to do new work, that is, to do new research into some field; rather, they are meant to standardize and specify things which are already known in such a way that makes open computing possible and proprietary vendor lock-in improbable. In the case of the OWL WG, however, this general rule was broken. While OWL has precursors, the most important of which is DAML+OIL, it took a non-trivial amount of real, new technical work to make OWL into a practical ontology language for the Web.

Despite our enthusiasm for OWL, we have to temper it with a dose of realism. OWL can be and probably is everything good which people have said about it; if so, that in and of itself will not mean that the Semantic Web visions will be widely achieved. Whether or not the Semantic Web ever happens, in as robust and important a sense as the original Web happened, depends on a complex set of factors and their interactions, only some of which are under anyone's direct control.

Having OWL means a few things are no longer true. First, it is no longer true that the Semantic Web can be dismissively written off as a bit of magical, wishful thinking on the part of some Utopian-leaning technologists. OWL provides a real foundation, rooted in the rich research and engineering tradition of KR and DL, for the Semantic Web. Second, it is no longer true that RDF and RDF Schemas are the obvious choices for a certain class of Web applications. OWL will soon be considered in some cases a better choice than RDF alone; it is more expressive and, in the OWL Full variant, upwardly compatible with RDF.

To see how OWL can be used, we return to our earlier example. Suppose the C.P. Snow Society wants to organize its bibliographic information already encoded in RDF. To take a simple example, they would like to distinguish between works *by* Snow and works *about* him. In OWL, we can express these concepts using class expressions, in particular, *restrictions* on the various properties a work has. For example, the class of work by C.P. Snow is just the set of work which have <http://www.cpsnow.org/cpsnow> (the person designated by this URI) as their `dc:author`, while the class of works by C.P. Snow is just the set of works which have <http://www.cpsnow.org/cpsnow> as (one of) their `dc:subject(s)`. We can easily express these definitions in OWL, give names to these concepts (e.g., <http://www.cpsnow.org/WorksByCPSnow> and <http://www.cpsnow.org/WorksAboutCPSnow> and expect an OWL system to correctly infer which works we've already described fall into which class. The C.P. Snow society can build upon these concepts to express the distinction between works and articles solely written by Snow and collaborative works (e.g., by defining `WorksByOnlySnow` as a subclass of `WorksByCPSnow` where there is only one author, and `CollaborationsWithSnow` as the subclass of `WorksByCPSnow` where there is at least one author who isn't snow). While helpful for organizing the C.P. Snow society's Web site, such an ontology only becomes interesting, and only become a true *Web* ontology, when it is published on the Web for all and sundry to examine, use, extend, or dispute, along with the facts (expressed in RDF) the ontology is meant to organize. Anyone, anywhere on the Web could then take the facts and impose an alternative or rival organization upon them, or take both the facts and the ontology and refine the ontology to greater detail. In this way, the Semantic Web enables non-coordinated (and even non-cooperative) collaboration about a domain of discourse, one in which the *conceptual* work is aided and abetted by programs. Not only will our Web Agents find and aggregate information from the Web (and without fragile and error prone "scraping" of HTML pages), but they will be able to give some initial guidance about whether certain aggregations make sense.

2.2 Everyone is a Hyperkrep Hacker?

Traditional knowledge representation oriented development, say, for expert systems, has required a strong division of labor between, at least, the domain expert and the knowledge engineer. Even when these two roles are performed by the same person, knowledge engineering requires a skill set that is not common, and is generally considered difficult to master. Even if the ontology is developed and deployed, adding new information or interpreting claims made by the system can be difficult. In the Semantic Web vision, there is the expectation that hordes of developers, web

masters, page authors, and even casual users will be creating and consuming Semantic Web data. Everyone will be hyperkrep hacker, able to casually create a mix of hypermedia and knowledge representation that fits in to the global hyperkrep system.

Why do we think that it's even possible, much less likely, that everyone can become a hyperkrep hacker? There was a time, not so long ago, when things like hypertext, markup languages, and relational database systems were considered too complex for most programmers or technically-sophisticated people. But these technologies, and the concepts they express, have become the building block of the Web as we know it today. Today more people than anyone every imagined build complex web sites and applications using XML, SQL, and a lightweight, high-level programming language.

Why did so many people learn to use such complex technologies? Because they were highly motivated by and committed to the success of the Web. There's no reason to believe that this same kind of thing won't happen for the Semantic Web. Logic programming, knowledge representation, and ontology modeling sound like very intimidating, complex tools and techniques. And in some ways they are; but no more so, or so we believe, than the technologies powering the first generation of the Web.

For classes, the situation will be must the same as for the Web. Early adoptors will be faced with the tasks of teaching the (Semantic) Web as well as teaching their subject matter. As the Semantic Web becomes more prevalent, as people start getting RDF classes in high school, as more people explore putting up their own Semantic Web pages, it will become very difficult *not* to use the Semantic Web in teaching, learning, research, and related activities and practices.

2.3 Semantics, Ontologies, and Education

So what impact may all this RDF and OWL have on the educational enterprise? Just as it was impossible to predict which features of the Web would have what impacts on institutions of higher learning, it is difficult to guess where the impacts of the Semantic Web will be most deeply felt in academia. However, one area where it seems fair to guess about the impact is in the area of the continued evolution of electronic publishing, continuing the trends we discussed earlier.

Using ontologies, in the next few years, we expect that tools for publishing will automatically help users to include machine-readable markup in the papers they produce. Whereas current tools using XML (Extensible Markup Language) can allow a user to assert that some part of a document is about an 'experiment', the new

languages will let the author express that the experiment uses certain chemicals and reagents; that the system used involved some particular organic matter; that the experiment produced gels with certain DNA information on them (and that the images of these gels are located in particular places on the web); and other domain-specific concepts expressed based on an OWL ontology (early versions of such tools are already becoming available.)

Papers that include this new markup language will be found by new and better search engines, and users will thus be able to issue significantly more precise queries. More importantly, experimental results will themselves be published on the web, outside of the context of a research paper. So a scientist could design and run an experiment, and create an emerging web page containing the information that he or she wants to share with trusted colleagues. Finding out about experiments and studies in progress will be easy, and work will be able to be modified as a result of interaction with peers, with less need to wait for formal publication. Just as preprints challenge established journal publishing approaches, these new 'papers in progress' will change the culture of publishing (and of the pursuit of science).

Additionally, the added expressivity of the Semantic Web, coupled with search and query tools already under development, will allow changes in non-scientific fields as well. For example a number of historians could each annotate the same document to express differences of opinion about its content, creating communities of deconstruction. Filtering mechanisms could provide capabilities for seeing annotations by some particular colleague, by all colleagues, by colleagues from a specific institution, etc. Non-historians could see these annotations and explore the marked up documents in other ways -- perhaps exploring them semiotically or even using pseudo-sciences like handwriting analysis or horoscopic analysis of the dates of publication (remember, it's the Web, everyone can play!)

Thus, it is not unreasonable to assume that in the long run, the Semantic Web will facilitate the development of methods for helping users to understand and to recreate in new contexts the content and knowledge produced by those in other disciplines. On the Semantic Web, one will be able to produce machine-readable content that will provide, say, automated translation between the output of a data collection study (say the cancer risk assessment tables published by the EPA) and the input of a data-mining package developed for some scientific pursuit (perhaps genomic databases). Mechanisms used in one field or discipline become available and linked, in real time, for others, creating a network effect in academic knowledge itself. The very notion of a journal of medicine separate from a journal of bioinformatics, separate from the

writings of physicists, chemists, psychologists, and even kindergarten teachers, will someday become as out of date as print journals are becoming today.

3. Conclusion

In short, we are in the beginning of a new revolution in information management that will make more and more content available to any combination of human and computer processing, allowing new means of collaboration between and across disciplines. However, the structures of teaching and learning, and the structure of the institutions that support them, are largely based on exactly the divisions between course content and discipline, especially in the higher grades. What will be the effect of the Semantic Web on education? It's hard to predict the details, but one thing is certain: if the Semantic Web becomes as ubiquitous as the Web is today, the effects will be profound.

Acknowledgements

Some of the material in this article amplifies ideas from a column in *Nature* entitled "Scientific Publishing on the Semantic Web" coauthored by Tim Berners-Lee and Jim Hendler.