THE DISCREET CHARMS OF ELECTRONIC TEXT

ABSTRACT

Here, the authors discuss ways in which new technologies can enhance our reading practices. They define and explain the versatility and usability of what they call 'metadata' or data about data, which can be used to tag electronic texts so as to allow readers instant access to a vast array of information. The parameters of these metadata tags are set by the readers, and can then be captured and converted into machine-controlled information-seeking dynamics. These dynamics may have little to do with curling up with a good novel. But they do speak to information access. And while novels, short stories, poetry, biography, philosophical treatises, histories, inspirational texts, and cookbooks in print form are all important, so is access to information using sophisticated search and user parameters.[1]

There is every reason in the world to celebrate reading. In fact, it is arguable that the centrality of reading and of literate values has served as a foundation for the development of western society. Combined with such technologies as print on paper and its derivatives – books, magazines, newspapers, scholarly journals and the like – reading and literacy have served as handmaidens to the primacy of reason combined with empiricism and hence the development of the scientific method, codified law, and even institutionalized democracy. In short, we owe much to reading and its necessary pre-condition, writing.[1]

Western society has built a variety of institutions in homage to reading. Educational institutions from the primary school through to the most prestigious universities are, to a significant degree, spin-offs of reading. To encourage authors to provide readers with access to written material on virtually any subject, western society has extended the dynamics of the marketplace into the expression of ideas. It has invented the notion of intellectual property and copyright so that authors can benefit from the fruits of their intellectual labour. In addition, many western societies have created public libraries to ensure that communities have wider access to written materials than they might have were each member required to purchase every item he or she might wish to read. So revered is reading material that both state and private enterprises foster reading. Private enterprise does it by subsidizing the dissemination of reading material in the form of magazines and newspapers.

But what if we were able to go further than reading? It is true that the technologies associated with reading are well ensconced in society and that they fit into social behaviour rather well. Hair salons are filled with magazine readers. Subways, buses and trains are filled with readers of all kinds. Beds and beaches are strewn with book readers. … But what if we could make reading more powerful, perhaps by using technologies other than the book (or magazine or newspaper)?

The purpose of this paper is to awaken you to the possibility that the conjoined production and reading of printed texts on paper leaves room for development and for improvement. Indeed, one might claim that it leaves room for a whole new creativity. The ideas and information we offer for consideration focus on metadata, a familiar term among librarians and computer scientists, but generally not a word that pops to mind when one wants to describe the pleasures of benefits of reading.

First, some definitions: metadata are data about data; information about information. And XML (Extensible markup language) is a coding technique (not the only one but the easiest for computers to handle) that brings metadata into existence. Metadata and XML presume databases of content. XML tags content, thereby creating a second level of content and allows for the separation of content and presentation. Let me put it another way:

Content is the text: for example, the words of this article. Content is stored in a database that can be accessed as required.

Structured content is content that is tagged or labeled denoting the status of the text in relation to the whole and to other tagged elements, for example, the title, authors, and headings of this talk provide a first level of structure.

Presentation is what you see – the first three letters of the old concept of WYSIWYG. But metadata turn WYSIWYG on its head, inasmuch as structured content can combine with a variety of presentation parameters to create multiple forms of the real (i.e., ways of presenting the text).

ROWLAND LORIMER

Rowland Lorimer is Professor of Communication at Simon Fraser University. He is the author of countless Major Commissioned Reports, books, and articles on Canadian mass communication and on the Canadian publishing industry.

ROWLAND LORIMER

John Maxwell is Instructor and research associate at Simon Fraser University. He is writing his dissertation, which traces the history of ideas in educational technology.

To carry on just one further step, grafted onto this foundation of a database of tagged and hence structured content, separated from presentation, is *scripting* – that is, commands that access certain tagged content, convert it "on the fly" (as they say) and present it in a particular form, a template for example.

The significance of the separation of structured content from presentation is perhaps not in the same league as the invention of zero, but it does have major significance. We will illustrate the significance in three main ideas.

In terms of the production or publishing of text, this separation of context from presentation leads to what is called *single source publishing.* In terms of reading, or access to text, this same separation allows for *rich or contextualized access.* At the juncture of publishing and reading, separated content and presentation make feasible *individualized texts.* Let us explain.

The significance of the separation of structured content from presentation is perhaps not in the same league as the invention of zero, but it does have major significance.

On the publishing side, one can produce tagged content that can be used in a variety of applications. The same content (the single source) is used (sometimes via scripting) for myriad applications. For example, the tags allow the user to list all labs (which then can be used to book the required rooms and technology). In the book world, an author's bio can be placed in a database and tagged as such and then used for a catalogue, a publicity piece, the back cover of a book, a citation in *Who's Who,* as part of a list of authors published by a house, and so on. Single source – multiple use. Imagine: Command = "Get bios 2002". Result: Bios of all authors published by the firm in 2002. Similarly, conversion scripts can take structured content files and present them in HTML as text on a screen or generate a PDF to be sent to a printer.

Metadata, that is to say the structuring elements of content, move us into a post-Google world. While Google (and other search engines) bring us searchability (based on the occurrence of words and phrases in any place in a text), metadata allow searches restricted to certain types of articles whose titles contain the words "reading in Canada." Metadata, however, allows an article that happens to use the phrase "reading in Canada" to be *distinguished from* an article that is "about" (in some formally defined sense) reading in Canada. Searching and selectivity are suddenly much more valuable. Our texts become, then, machine-understandable. This doesn't have to imply anything as esoteric as artificial intelligence; it simply means that the semantic structure of the text has been made explicit, so that it can be acted upon by software, so that the machines can be part of the sense-making process.

The added value that metadata and hence machine understandability yields is what is termed "the semantic web" – not merely the vast array of information and disinformation of the web itself but an *array of ordered of filtered* content. Metadata, then, can be seen (for computers) as a kind of Rosetta stone of web content and an immensely powerful reference tool for human beings using the web.

Added to the above, and again, derivative from databases and metadata, is an additional possible array of automated control and interactive functions connected to database operations. For example, in a neat little module we created for the *Canadian Journal of Communication,* we invited graduating Communication Ph.D.'s to enter their thesis abstracts into a database. Included in the information requested is the name and email address of their supervisor. When they click on "submit," an email is generated from the editor to their supervisor asking him or her to confirm the validity of the abstract. When the supervisor validates the submission by replying to the editor, the editor accesses the submission, clicks once to change the status of the submission from "inactive" to "active" and presto, the abstract is published on the CJC-Online.ca. By the way, a lot of science article publishing is incorporating procedures such as the one described above with the added function of online editing by both reviewers and professional editors. It both speeds up the process and renders it less expensive.

To move on to the second idea, on the side of reading, tagging or metatagging allows for contextualized access. Apart from the issue of a much-improved signal-to-noise ratio in the results we see when searching the web, the separation of format from content means that a text is not the same thing to all people. While literary critics have made this point for some time, consider the multiple presentations of a work in formats suited to (a) sustained reading, as at the beach or on the bus, (b) close reading for reference, perhaps onscreen with a search tool, (c) text-to-speech conversion for the hearing-impaired, (d) quick browsing on a cell phone or PDA, and so forth. Further, as collections of tagged texts grows, new possibilities for reading styles and research appear: for example, a citizen can examine the time-ordered statements of a politician on a particular topic alongside his or her voting record, by browsing legislative records tagged according to such criteria. This is an illustration of creating or accessing ontologies "publishing or recorded histories and contexts."

At the juncture of publishing and reading on the web, the above dynamics are combined with reader ontologies. The resulting process is inspired by the much maligned and much abused technology called "cookies." Imagine if your computer provided you with the following when you accessed a report you had been reading online:

> In your last session you highlighted the following text passages. (A list would then appear.) Further, you noted the following. (The list is then changed to an annotated list.) At the end of your session you added some thoughts of you own. (A second list appears.)

You may wish to proceed with your analysis. However, since your last session the following scholarly articles, government reports, and newspaper clippings have been identified as potentially relevant to your topic. Here are abstracts of those texts. You may access them now, or later, or you may delete the derived list.

My sense is as these three dynamics, web publishing, contextualized access, and tracked, smart user histories come together, we seem to be talking about something significant.

In overview, at the simplest level, metadata allow for user-selected and user-sensitive information access, or, put differently, user-defined databases, assuming, of course, that the information exists online and that user-responsive software continues to be developed. Placing metadata at the fingertips of the reader combines the power of print on paper with the range of and ready accessibility to the vast array of content on the web, and with the tools of a reference librarian. Instituting smart tracking extends or projects forward the efforts of user potentiality quite powerfully.

In overview, at the simplest level, metadata allow for reader-selected and reader-sensitive information access, or, put differently, reader-defined databases and reader-responsive information seeking. Bumping up the significance somewhat, the full implementation of metadata combines the power of print on paper with the range of tagged web content, and with the tools of a reference librarian. Instituting smart tracking potentially extends or projects forward the efforts of the reader quite powerfully.

Let us inch toward a conclusion with another example to illustrate a slightly different point. Here are some facts about book purchasing that come from two studies undertaken by faculty of the Canadian Centre for Studies in Publishing in 1997 and 1999:

- Sixty percent of bookstore visitors depart with a purchased book that they did not intend to buy when they entered the store.
- Bookstores are the main source of awareness of books.
- Sixty percent of book purchases are made by women.
- Eighty-one percent of children's books are purchased by women, mostly of 'motherhood' age.
- Canadian book buyers are mostly up-scale citizens. Two-thirds have graduated from university and over 50% have a household income of over $55,000.
- While women purchase more books than men, men spend more per book.
- The average buyer purchases 2.3-2.5 books.
- The average amount paid per book in 1999 was $15.95, males $17.05 and females, $14.95.

These are points that are germane to any discussion of reading. Were they readily available on the web, and were teachers, publishers, scholars and librarians sophisticated and constant users of the web, these facts would be well known and they would likely inform discussions of reading and book purchasing. But because they only exist in print within two studies done for two different Canadian publishers' organizations, they are effectively hidden from view. And so we tend to work with a lack of information; to wit: that books are not like opera tickets, purchased primarily by the well-to-do; that people most often purchase one book when they go to a bookstore; that the majority of people do not buy on impulse; that people find out about books from a wide variety of sources, none of which are dominant.

While conceptualization is extremely important and appears to be greatly facilitated by literacy, and conceptualization has brought much to western society, in part because western society has embraced conceptual thought, there is, at certain times and in certain circumstances an all-too-readiness to substitute conceptualization for information.

Let us close with a speculation. While conceptualization is extremely important and appears to be greatly facilitated by literacy, and conceptualization has brought much to western society, in part because western society has embraced conceptual thought, there is, at certain times and in certain circumstances an all-too-readiness to substitute conceptualization for information. The example of the book purchasing data was meant in part to illustrate how we can carry around a concept of the typical book purchaser that may not be borne out by the facts. Because electronic text allows ready access to a vast increase of ordered information, and user dynamics can be instituted to assist the user in a defined task, we may find that the tendency to conceptualize in the face of limited information may lessen. Put differently, the balance between reason and empiricism may shift somewhat towards empiricism. The matter is not just easy access. The web addresses the real in sound, colour, and still and moving images. Electronic text may lead us out to the courtyard (or at least to an account, complete with visuals, of someone who went out into a courtyard) to count the number of teeth in a horse's mouth rather than encouraging us to merely speculate or repeat our well-worn conceptualizations.

Notes

[1] We wish to thank Richard Smith for his comments on earlier drafts of this paper.

[2] Derrida aside, for the moment.