Three Digital Publications: Carrying the Printer Home

Wybo Wiersma
King’s College London
mail@wybowiersma.net

Abstract

In this paper we compared three widely differing digital editions of Adam Smith’s *Wealth of Nations*: those offered by the Gutenberg Project, the Library of Liberty, and Google Books. While there are notable variations between these resources in terms of their suitability for various ways of reading, we found their usefulness for serious research to be rather limited, mostly for lack of transformative possibilities (such as adding notes to them). Thus instead of replacing the printed version, digital editions are mostly to be seen as an additional format, especially suitable as input to other programs and processes such as searching or home printing.

1 Introduction

In this paper we will look at three on-line versions of Adam Smith’s *An Inquiry into the Nature and Causes of the Wealth of Nations*. They will be examined for their readability and their usefulness for research. First we will explain the limitations of this paper and the study it is based on, then we will justify our choice of resources and editions, followed by a short sketch of the different modes and types of reading that we judged them for. When the approach is thus clarified, we will describe the three resources in more detail, and give a reasoned critique of them. Finally we say something about possible improvements, and likely current uses of the resources.

1.1 Constraints and Limits

In this paper we limit ourselves to on-line versions, and will not be looking at other software for reading electronic versions of books. We will only use the standard, unextended (no plugins), Javascript enabled browser (Firefox 3.5.x) using it’s default settings (default fonts & font-sizes). In a sense this could be seen as an artificial limitation — and maybe even begging the question —, as nowadays there is both specialized software for reading (various .pdf readers), and some specialized software for research available (such as John Bradley’s Pliny, the XLibris annotation tool, the PKP Reading Tools, and others).¹

However first of all we had to limit the scope of this paper for practical reasons. Secondly we wanted to stay as close as possible to the experience of the regular reader and researcher (plugins and specialized software are still hardly used). Naturally, while very relevant and important, the reading-experience on dedicated hardware-devices such as the Kindle, or Netbooks (for reading on the bed) are also not taken into consideration.

1.2 Choice of Resources and Editions

The three editions of the Wealth of Nations we are looking at are: the Project Gutenberg plaintext version (Gutenberg), the HTML (by-chapter) version offered by the Library of Liberty (Liberty) and the oldest scanned (Harvard 1789) edition on Google Books (Google).2 The reasons for choosing this set of resources are first of all that they are large, (relatively) well known projects, thus many readers and researchers would be accessing the digital version of this book through one of these sites, if at all.

Secondly they are different in 4 aspects: the edition they provided, in how it was created, what format it is in, and are also done by very different kinds of organisations. Therefore an interesting comparison can be made: The Gutenberg Project does not give any information about the edition(s) used or offered, uses OCR with manual proofreading, offers simple ASCII plaintext, and is a web-community project. The Library of Liberty provides a licensed academic edition, where differences are outlined in footnotes, makes use of HTML and hypertext, and is a nonprofit foundation. Lastly the Google Books project presents scans of specific, well-described editions, offers facsimile .pdf’s of the books pages for reading, and Google is a commercial corporation digitizing millions of books. Thus the choice of resources ensures a broad, varied view of mainstream digital resources (table 1).

1.3 Two Modes of Reading

We will be looking at two modes of reading. The first mode, is leisurely reading. This kind of reading is the ‘light’ kind that most people do for entertainment. It is the most common for books (especially for novels). The second mode is scholarly reading. It is quite a diverse range of more serious ways of reading, such as done by students, academics, and professionals when they want to research a text, or want to learn from it.3 These modes can then be of two types: fragmented, or attentive. Fragmented reading is reading while browsing the web, interacting on a forum, searching, or looking things up. Attentive reading on the other hand is linear, and lasts over a long time. It is heavily influenced by, and suitable for, traditional print-media.4 Both attentive and fragmented reading can be done for leisurely and scholarly purposes (table 2).

---


2 Project Gutenberg Plaintext Version

The Gutenberg Project is the oldest digital library. It began as a one-man project, founded by Michael S. Hart. It mainly contains public-domain texts. Initially these texts were hand-crafted by volunteers, but now most of them are OCR-ed and then manually proofread by volunteers via the Distributed Proofreaders project (in page-sized chunks). The aim of the project is to ‘provide as many e-books in as many formats as possible’, but the plaintext format (US-ASCII or ISO-8859-1) is preferred by Hart, because it would be more universally readable, also far into the future.

Therefore we chose to focus on the plaintext version on this site. From the download-page other formats of this text are offered (both compressed and uncompressed): HTML (nice free flowing 10% margins, and css justified text), Unicode, and various e-book formats. Most notably missing is the .pdf format (commonly used for screen-reading). In addition a ‘recoding service’ is offered, which allows one to transcode the text to about three dozen different character encodings.

For the plaintext format there is both the possibility to open the whole (2.2mb) file in one browser-screen, and the ‘On-line Reader’, which splits up the text in pages, and allows one to bookmark the page one is reading. Interestingly it assumes one finishes a page before placing a bookmark, because when one returns to a bookmark, it opens at the page following it. No table of contents is provided, nor are there any hyperlinks or footnotes included. Plain plaintext, in short, even in the on-line reader.

If we open the whole .txt file in one browser-window (figure 1), it loads relatively quickly (much faster than reaching out for a book on a shelf). One is presented with the whole text (all 575 ‘pages’ of it). It is almost like having a giant papyrus scroll in front of oneself, except that it scrolls much easier, and is a lot lighter. The typesetting of the text is quite primitive (default 12pt monospace font, which people less used to Linux-consoles might consider ugly) and the ends of lines are rigged (not justified). Reading it for more than a few minutes is quite unpleasant.

Because of the proofreading by the community the quality of the text itself is quite good, which makes it suitable for leisurely reading. Lack of information on its edition makes it less useful for scholars. And mainly because of the poor layout, it is not ideal for attentive reading, while it works fine for leisurely fragmented reading such as short reads, searching or looking for quotes (table 3). It seems that the main aim of the Gutenberg Project of making texts available, but not necessarily readable or printable (in this format at least), seems to shine through in this.

<table>
<thead>
<tr>
<th>Table 3: Suitability of Gutenberg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leisurely</td>
</tr>
<tr>
<td>Fragmented</td>
</tr>
<tr>
<td>Attentive</td>
</tr>
</tbody>
</table>

3 Liberty Fund HTML Version

The Liberty Fund is a private foundation whose aim is to educate people from a Libertarian point of view. It was founded in 1960 by an Indianapolis businessman and lawyer, and currently has a fund of about 200 million dollars in assets. It publishes books and organizes conferences, and one of these books is Smiths Wealth of Nations.

Their edition is from 1976, edited by Roy H. Campbell and Andrew S. Skinner (both academics) and licensed from Oxford University Press. It is based on the first 6 editions of the Wealth of Nations, the first 5 of which were published during Smiths life-time. In

---

1Gutenberg, Project Gutenberg - The Wealth of Nations.
addition to Smith’s own footnotes, and footnotes by the editors, the differences between the 6 source editions are also lined out in special notes. It thus is a proper scholarly edition.

The site it is on (figure 2) is called the Online Library of Liberty (a Joomla CMS instance). Right at the top of the page one sees information on the edition. At the right is a box showing the categorization of the book, and allowing one to click through to other books in the same categories. Below that is a listing of the formats in which the book is available (HTML, HTML by chapter, facsimile PDF and e-book PDF). Then follows more information about the edition, concluded by the table of contents of the book. We are using the HTML by chapter, as it seems to be the suggested way of reading.

The first chapter consists of a detailed introduction to the book, and Smith’s other works. At the top of it is again the information about the edition, and the categorization box. Then follows the table of contents, followed by the text itself. Paragraph-numbers are shown to the right of each paragraph, but strangely enough, while they do contain anchors, they are not links, and so cannot be bookmarked (nor are they used in footnotes to get back to the text; they can only be found by inspecting the source). In addition the text of each chapter is quite long 5,000 to 35,000 words, and because there is no link to each next chapter at the bottom, it is quite a scroll back up to the table of contents (from which one is supposed to pick the next chapter). This all makes it seem that, compared to all the effort invested in the edition itself, no thought was given at all to the on-line readers experience.

The rigorously scholarly edition is suitable for scholars. And because the footnotes are put at the bottom, it is possible to read it in a leisurely way. However the lack of ‘next page’ links, and the impossibility of making bookmarks to places inside the text, make it not very suitable for attentive reading. As far as fragmented reading is concerned, it is fine...
for the passive, mostly leisurely kinds such as searching, but there is no way for adding annotations or taking clippings linked to the original text. For any serious, attentive, scholarly use, printing it out seems to be the expected action (table 4).

<table>
<thead>
<tr>
<th>Table 4: Suitability of Gut. and Liberty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leisurely</td>
</tr>
<tr>
<td>Fragmented</td>
</tr>
<tr>
<td>Attentive</td>
</tr>
</tbody>
</table>

4 Google Books full view

Google Books (officially Google Book Search) is a project by Google, aiming to make books searchable, and available worldwide. It currently contains 10 million books (Oct 2009). It started out as a partnership between Google and 5 universities (among which Harvard, Stanford and Oxford), and currently it has over two dozens of participating institutions. The books are scanned, and then OCR-ed (OCR used only for search), while users are presented with images of the pages for reading.

The edition we used is the oldest edition available on Google Books. It is the 1789 edition from the Harvard College library. It is presented on an ‘overview’ page, which contains a lot of useful information. Among which links to sellers of the book, reviews, both by buyers (108) and professional reviewers (none found in this case), related works (also by Adam Smith) and other editions, and references from other books and sites. But there are three more features.

The first of these is a tag-cloud with the most frequently mentioned terms in the book. These terms are links which allow one to view all instances of them highlighted in images of the books pages (concordances). In addition to terms given in the cloud, it also allows you to search for any other terms. The second feature, are mapped out place-names, with links to the pages on which they occur, which is nice to have, though probably most

9Google, Google Books - An inquiry into the nature and...
useful for other kinds of texts (literature, or travel guides). And the third are popular passages from the book. That is; passages quoted in many other documents. This, combined with the other meta-data, is very useful for getting a quick feel for the essence of a book.

Then to the text itself. Because it is out of copyright, it is available as ‘full view’. So the whole text can be read on Google Books. It is also available for download in PDF format (monochrome, lower quality than what one sees while reading it on-line), and in the EPUB e-book format. In addition one can view the plaintext version. This is the output of the OCR process, and it is full of typical OCR-errors (such as ‘TN’ instead of ‘In’).

The standard images-view (figure 3) starts out with the brown cover of the book, and then continues with the cover-pages, and white-pages, before the real text begins. In these pages the notes by the librarians are still visible (one edition even shows stamps with library return-dates). Pages can be navigated both via a drop-down index, and by clicking on the hyperlinked index in the book itself (still in it’s historic font, very nifty), via a thumbnails overview, two or one pages at a time, in full-screen-view, and simply by scrolling down. Especially this page-by-page scrolling down is very convenient and fast, as there is no interruption when ‘flipping pages’. It reminds one of Vannevar Bushes Memex. And indeed doing it through a lever as proposed for the Memex (speeding up as pushed further down), instead of by mouse-scroll-wheel, would be even more convenient.

In short, Google Books is suitable both for

<table>
<thead>
<tr>
<th>Table 5: Suitability of Gut. Lib. and Google</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leisurley</td>
</tr>
<tr>
<td>Fragmented</td>
</tr>
</tbody>
</table>

5 Digital Publishing: Carrying the Press Home

To paraphrase McLuhan, the web carries the press, as the press carried writing, writing carried speech, and speech carried thought. Nevertheless most scholars still carry books from the press, or at least a printer (home-press) into their study. This home-press is then used to print off the digitally obtained article anyway, allowing the web to only carry the pre-press-manuscript of the book or article, not the press itself.

And as noted this is because the examined electronic editions are at best good for leisurely attentive and fragmented reading, and for fragmented scholarly reading. Attentive scholarly reading, whether actually practised, or intended at the time of getting the book, requires the ability to take notes, either in, on the book/article itself, or in a way that easily allows one to link notes to pages. Current digital editions support none of this.

Now while it is true that putting a digital text on the net is distinct from making it interactive, and having to use 30 different site-dependent tools for annotating the same number of editions would not be very helpful, but...
the fact that no well-known, easy to install, responsive, user friendly application for this exist, is — we think — what keeps scholarly use down (the thing that comes closest is Pliny). So scholars are, to a large extent, justly ignoring digital editions.

The ability to take notes on books and articles would best be provided by external software / an external service, in which the source-texts could be imported. Having editions available in a standard format, such as TEL, would help the development and viability of such tools enormously. Pretty websites hosting book-blobs for screen-reading are nice, but until these issues are addressed most scholars will only use them for search, and then go straight to the — to speak with hypertext-pioneer Ted Nelson — PDF ‘papersimulation’ to print it off.

6 Conclusion

To conclude, we have analysed the Gutenberg projects edition, one that was mostly there to make the work available, and was not suitable for any, but fragmented leisurely reading. Then we looked at the Library of Libertys superb scholarly edition, which was simply dropped onto the web without much thought about the readers experience. Thus making it only suitable for scholarly and leisurely fragmented reading. The final resource we looked at was Google Books’ edition. It was the best of the three, being suitable for attentive leisurely reading, and fragmented reading in both leisurely and scholarly modes.

Lack of annotation-facilities was what barred them from attentive scholarly reading. To some extent this was part of the studies setup of not using plugins or external software. Nevertheless far too little, if at all, of such software exists or is easy to install, and usable from a humanities scholars perspective. Until it does, digital editions are most useful as input into search-engines, for search, and for home-printing.

References

Bradley, John. Pliny. 2009. URL: http://pliny.cch.kcl.ac.uk/


MacGregor, James and Michael Joyce. “Revolutionary Reading, Evolutionary Tool-making: (Re)development of Scholarly Reading and Annotation Tools in Response to an Ever-changing Scholarly Climate”. In: (2009).


