

THE INDEXER

Volume 30 Number 1, March 2012

ISSN 0019-4131 (Print)
ISSN 1756-0632 (Online)

The International Journal of Indexing

CONTENTS

- | | |
|--|--|
| 1 Editorial: digital trends | <i>Jan Wright</i> |
| 2 The tyranny of the page | <i>Mary Coe</i> |
| 6 New technology and public perceptions | <i>Bill Johncocks</i> |
| 11 The devil is in the details: indexes versus Amazon's X-Ray | <i>Jan Wright</i> |
| 17 Hand-helds as ereaders: exploratory thoughts on hand-held devices and indexes | <i>Pilar Wyman</i> |
| 25 Visualizing back-of-book indexes | <i>Ceilyn Boyd and Mitch Wade</i> |
| 38 The semantic web: an introduction for information professionals | <i>Matt Moore</i> |
| 44 Publishing, XML and indexers | <i>Nic Gibson</i> |
| 47 XML indexing | <i>Michele Combs</i> |
| 52 Reviews: breaking the page | <i>Edited by Christopher Phipps and Michael E. Jackson</i> |
| 53 Web resources for indexers | <i>Pierke Bosschieter and Jan Wright</i> |
| 56 Around the world: indexing-related awards | |
| 57 Indexes reviewed | <i>Edited by Christine Shuttleworth</i> |
| 60 Reviews | <i>Edited by Christopher Phipps and Michael E. Jackson</i> |

THE INDEXER

The International Journal of Indexing ISSN 0019-4131

Editor: Maureen MacGlashan

Production Editor: Susan Curran

Around the World Editor designate: Sanet le Roux

Reviews Editor: Christopher Phipps

North American Reviews Editor: Michael E. Jackson

Indexes Reviewed Editor: Christine Shuttleworth

Centrepieces Editor: Noeline Bridge

Guest Editor (March 2012): Jan Wright

© 2012 The Indexer and individual contributors.

The Indexer is published by the Society of Indexers on behalf of the societies and networks subscribing to the International Agreement of Indexing Societies. It appears four times a year (March, June, September, December), in print and electronic versions.

Subscriptions: 2012

Institutions: £130 (online only); £130 (print, single copy); £165 (online and print).

Individuals (non-society members): £44 (print, including priority overseas delivery and online access).

Individuals (indexing society members): £28 (print, including priority overseas delivery and online access).

Full details of subscription rates, including online subscription and special offers, are available on the website (www.theindexer.org). Further information may also be obtained from the Subscriptions Manager (subscriptions@theindexer.org).

A cumulative index, at present running from 1996 to the most recent issue, is available at www.theindexer.org.uk

Full-length articles are subject to peer review. The views expressed by contributors are not necessarily those of the Society of Indexers or of any of the other signatories to the International Agreement, nor do we guarantee the claims of advertisers.

Apart from fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the Society of Indexers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency.

Abstracted in *Information Science Abstracts*, *BUBL Information Service*, *Library and Information Science Abstracts*, *Library, Information Science and Technology Abstracts*, and *Current Awareness Abstracts*. Indexed in *Managing Information*. Contents pages reproduced in *Documentaliste*. Full text available in Ebsco's electronic databases.

The Society of Indexers is a non-profit organization founded in 1957 as an autonomous professional body for indexers to promote indexing, the quality of indexes and the profession of indexing. Further details of the Society of Indexers and the other signatories to the International Agreement will be found on the inside back cover.

CONTACT ADDRESSES

Editor: Maureen MacGlashan

16G Main Street, Largs, Ayrshire KA30 8AB, Scotland

Email: editor@theindexer.org

(For all matters of editorial policy and for submission of articles, regular features and shorter items)

Production Editor: Susan Curran

Curran Publishing Services Ltd, 2 St Giles Terrace,

Norwich NR2 1NS, UK

Email: production@theindexer.org

(For all matters related to production)

Reviews Editor: Christopher Phipps

51 Elsham Road, London W14 8HD, UK

Email: reviews@theindexer.org

North American Reviews Editor: Michael E. Jackson

6530 N. Greenview Avenue, Chicago, Illinois 60626-5008, USA

Email: reviews-america@theindexer.org

(For all North American publications for review)

Around the World Editor designate: Sanet le Roux

10 Sam Gordon Drive, Brackenfell, 7560, Western Cape, South Africa

Email: world@theindexer.org

(For all news items and information on recent and forthcoming events in the indexing world)

Webmaster: James Lamb

Email: webmaster@theindexer.org

Indexes Reviewed Editor: Christine Shuttleworth

Flat 1, 25 St Stephen's Avenue, London W12 8JB, UK

Email: indexesrev@theindexer.org

Browser Bar Editor: Pierke Bosschieter

Email: browserbar@theindexer.org

Centrepieces (names) Editor: Noeline Bridge

Email: names@theindexer.org

International Advisory Board members

ASAIB: Madely du Preez (asaib@theindexer.org)

ASI: Pilar Wyman (asi@theindexer.org)

ANZSI: Alan Eddy (anzsi@theindexer.org)

CSI: Guo Lifang (csi@theindexer.org)

DNI: Jochen Fassbender (dni@theindexer.org)

ISC/SCI: Ruth Pincoe (isc-sci@theindexer.org)

NIN: Caroline Diepeveen (nin@theindexer.org)

SI: Pam Scholefield (si@theindexer.org)

Advertising: advertise@theindexer.org

(For advertising in *The Indexer* and other SI publications)

Subscriptions Manager: Paul Machen, Society of Indexers,

Woodbourn Business Centre, 10 Jessell Street, Sheffield S9 3HY, UK

Email: subscriptions@theindexer.org

CONTRIBUTIONS

Proposals for contributions should be made to the Editor. Contributors' notes can be found at www.theindexer.org

ADVERTISING RATES

Full page: £210 Half-page: £120 Quarter-page: £84 Eighth-page: £60 Inserts by arrangement. Generous discounts for block bookings.

The Society of Indexers is a company limited by guarantee, Registered in England and Wales. Company Registration No: 6303822.

Registered address: Woodbourn Business Centre, 10 Jessell Street, Sheffield S9 3HY.

Typeset by Curran Publishing Services, Norwich, UK

Printed in the UK by Barnwell Print Ltd, Aylsham, Norfolk

Editorial: digital trends

Jan Wright

As I write this in February of 2012, we have received the exciting news that the International Digital Publishing Forum (IDPF) Board unanimously approved the charter of the Indexes Working Group. Then, during a two-week voting period, the entire IDPF membership cast their votes, and the Working Group was approved by a vote of 69 for, 1 against, 1 abstain. This means that the Indexes Working Group will take on the challenge of writing ePub specifications for indexes so that they will become active parts of ebooks. An ePub 3.0-compliant ebook will declare that it has an index available in one or more formats, and will make navigation to the index much easier. Indeed, we hope that the index becomes one of the major navigational tools available for an ebook. When implemented by ereader manufacturers, an ePub 3.0 index will include exciting ways of viewing index data in an ebook:

- A traditional browsable back-of-the-book chapter view, with active hyperlinks and easy navigation, both to the index and within the index.
- A second pop-up view that will allow the reader to highlight a word or phrase, and have the index auto-scroll to that portion of the index. While in the pop-up view, the user can type new terms and look at those index sections, and snippets of contextual text surrounding the locators can be previewed.
- A third non-traditional contextual view of all the index markers in a portion of highlighted text, allowing the user to follow new paths to other index terms and to other portions of the book. We can expose index metadata for use by readers at the point of reading.

The approval of the charter is an exciting step forward toward the active inclusion and use of indexes in ebooks. The team had extremely thorough and detailed input from indexers, nonindexers, and international societies. The dreaded phrase ‘Can’t they just use Search?’ was never heard. Please do check on the status of the initiative as it moves forward at <http://bit.ly/uqKwD7>

There are still two elephants in the room, and their names are Amazon and Apple iBooks Author. So we still have work to do, as the Kindle does not support ePub. Its newest format, Kindle8, needs to be studied, and indeed supports some ePub and HTML5 thinking. Apple made a huge announcement about its new authoring tool for the iBookstore. Its new tool has conflated the glossary with the index, and changed the textbook game. Nooks, Kobo, Google books, iBooks and many other readers support ePub, but we need Amazon’s and Apple’s ear so that we can talk about indexing that works on the screen and for the reader.

We are experiencing changes in publishing technologies on every front: how books are read, prepared for publication, sold, distributed, and marketed. How do we get index data created and available for all kinds of formats? We face

a battle not only to meet our deadlines, but also to learn new technologies. We need to understand the level of technical expertise our clients have. They are having to learn new techniques at the same time we are, and are often not sure how to go forward. Discussing the hows and whys, and the eventual outputs they want to create, becomes crucial for indexers now. We all need to learn more about the technologies so that we can discuss project processes, know the implications of decisions, and understand the final displays of our indexes.

Paper now is just one more interface, and although it is the most free and luxurious in its display capabilities, we might need to index to a narrower, more stringent screen display. We may face limitations in numbers of subheadings, width of entries, numbers, symbol use, and coding of locators. We might need to change our phrasing styles and figure out how devices alphabetize and sort. ‘Embedded,’ ‘unique IDs,’ ‘anchors,’ ‘dynamic indexing,’ ‘tagging,’ ‘XML,’ ‘DocBook,’ ‘HTML5,’ ‘Kindle 8 format’ – all these terms could enter your life in the next email from a client. We have to be able to describe processes, workarounds, pros, and cons to clients. We have to think about the interface the index is appearing in, and write for that interface’s strengths and weaknesses.

This issue of *The Indexer* is dedicated to technical aspects of indexing in many arenas. We have long been working with set pages, around changing pages, or without page boundaries, and Mary Coe and Bill Johncocks address the tyranny of the page. What happens to both reading and indexing when there is not a set page boundary? Does anyone miss it? Matt Moore gives us an overview of the semantic web and what it means to us in understandable terms. Nic Gibson and Michele Combs address XML in broad and specific terms, allowing us to understand the impact of XML processes in the publishing world. They give us a close-up view of XML indexing in practice.

Hand-held devices demolish the concept of a page, and rearrange our thinking about best ways to get a reader to a topic on a very small screen. Pilar Wyman discusses the development and implementation of indexes for medical works in hand-held format, and how they relate to printed edition indexes. Ceilyn Boyd and Mitch Wade take us on a fascinating journey into visualizing indexes as data, emphasizing relationships between topics and locations in a graphic format. Can we use some of these visualizations to show readers related topics and locations? Amazon has implemented one form of visualization: is it truly helpful, or not? I then take a closer look at the new Amazon X-Ray feature, which attempts to pull in index-like data, visualize its locations, and present Kindle readers with supplemental crowd-sourced information, with varying success. Is this better than an index? Can we convince people that it is not?

continued on page 46

The devil is in the details: indexes versus Amazon's X-Ray

Jan Wright

The X-Ray feature available on new editions of Amazon's ereaders could be viewed as a replacement for an index. Comparisons of the details available to readers in X-Ray and in a traditional index show that X-Ray is missing a lot of valuable information, and does not show us, as Amazon claims, 'the bones of the book.'

The devil in the white city by Erik Larson (2003) was a best-selling history of the 1893 World's Columbian Exposition in Chicago, Illinois. Larson combined the story of the architects designing and creating the beautiful buildings for the fair with the tale of a serial killer who used the crowds and festivities of the event to shield his murders. The first Ferris wheel was built for the fair, and many renowned characters passed through the grounds or helped create them, such as Buffalo Bill Cody, Frederick Law Olmsted, Susan B. Anthony, and Franz Boas.

The printed editions of *Devil* had a index of approximately 1,032 lines. The Kindle edition of the book does not include that index. For those who read the book on a Kindle purchased before November 2011, their only access to the contents is through a hyperlinked table of contents, or through the Search feature.

After the release of the Kindle Touch, however, Amazon implemented Kindle X-Ray, a new search and information feature that allows you to find information about characters, events, and topics in books. According to the Amazon website:

Amazon invented X-Ray, a new feature that lets customers explore the 'bones of the book.' With a single tap, readers can see all the passages across a book that mention ideas, fictional characters, historical figures, places or topics that interest them, as well as more detailed descriptions from Wikipedia and Shelfari, Amazon's community-powered encyclopedia for book lovers.

Amazon built X-Ray using its expertise in language processing and machine learning, access to significant storage and computing resources with Amazon S3 and EC2, and a deep library of book and character information. The vision is to have every important phrase in every book.¹

As indexers, we have seen many concordance tools and proposed replacements for indexing come and go over the years. We have heard the question 'Wouldn't you just use Search?' a few too many times, and we all have our elevator speeches ready to explain (yet again) why indexes are a superior navigation tool. X-Ray will require us to have yet another version of the speech ready, so we need to examine its workings.

Is X-Ray yet another concordance tool, or is there some real value for readers in this technology? Let us see what

kind of results it gives the reader, and compare them with the printed index.

X-Ray functionality

X-Ray, as of this writing, is only available on the Kindle Touch model. If successful, we are sure to see it implemented in the baseline software for other models of the Kindle as well. The new technology was introduced when the Touch was issued in November 2011, and some books were immediately available with the feature implemented. The feature is implemented by the use of a sidefile, a file that is downloaded with the book from Amazon. Books can be updated for X-Ray when the Kindle is connected, and books the reader owns might suddenly gain the X-Ray feature, as more and more files are added to the X-Ray system.

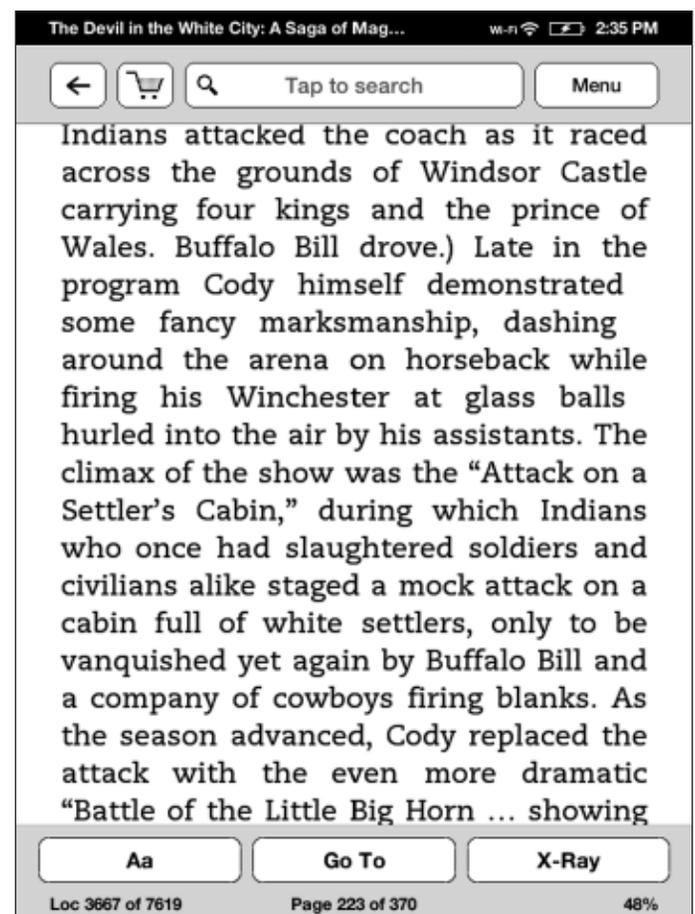


Figure 1 Accessing X-Ray

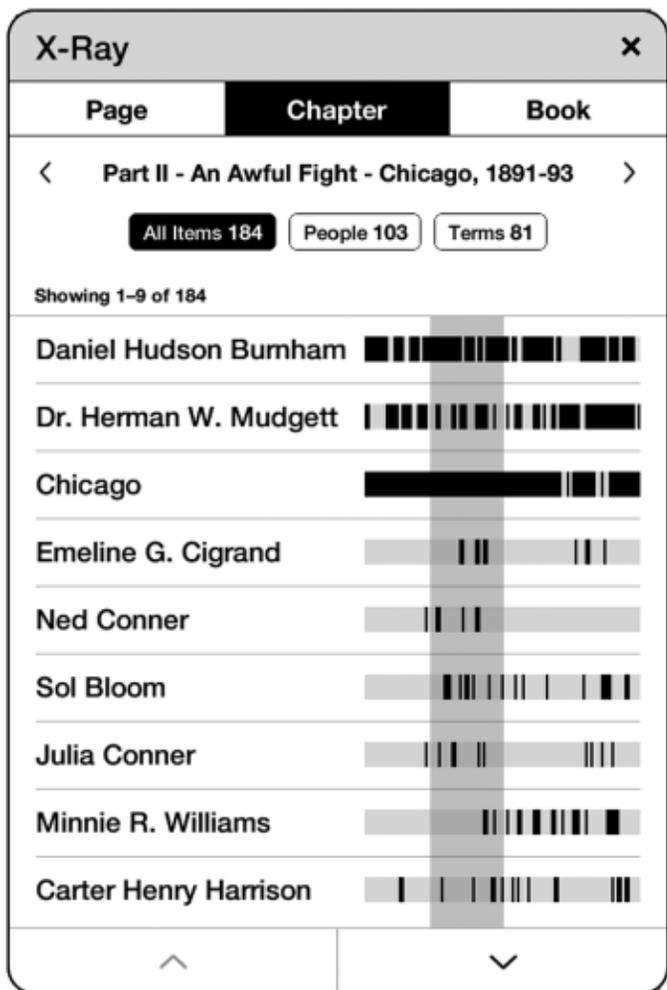


Figure 2 X-Ray’s choices for page, chapter, or book filtering, as well as choices between people and term filtering. This screen shows choices for a chapter, and for all entries, both people and terms. The sparkline-like graph shows occurrences throughout the book. The vertical gray bar shows which occurrences are in the current chapter. Note this screen was shot directly after Figure 1.

The sidefile, if available, downloads automatically. This file contains links and information about the book, allowing the reader to access Wikipedia information, Shelfari information, and other terms that have been determined (by human or machine means, we do not know which) to have importance.

To access X-Ray, the reader taps near the top of the Kindle screen, and then chooses the X-Ray button in the lower right (see Figure 1). The feature then shows choices for the reader: examine this page, examine this chapter, or examine the entire book (Figure 2).

One immediate issue is that you cannot sort the results in alphabetical order. You can filter them by chapter or page, or by people or terms, but they are still in non-alphabetical order. The small up and down arrows at the bottom of the screen allow you to navigate through the list. The sparkline-like bars next to each subject show you the occurrences of the topic in the text, and allow you to tap on a vertical line to access the location. The results do not seem to be ordered by importance.

X-Ray has identified 341 items for this book, 205 people and 136 terms. This works out to be about a third of the number of lines in the printed index, so it provides about a third of the details.

Clicking on a topic (Buffalo Bill, or Europe) leads you to a screen showing additional information pulled from Wikipedia or Shelfari, and snippets of text from locations in the book where the topic appears. Let’s explore Buffalo Bill (see Figure 4).

X-Ray’s treatment of people

As you can see from Buffalo Bill’s entry (Figure 4), X-Ray has some value in giving you background on a character and his role. And if the person has not appeared a lot in the text, it is relatively easy to read through and pick a location to read about him or her further in the text. The printed index has a heavy component of personal names, listing architects, planners, developers, relatives, and other important or interesting people who attended, had a part in, or were murdered during the fair.

William ‘Buffalo Bill’ Cody was one of the famous people involved, and the indexer included two sections related to him in the index: ‘Buffalo Bill’s Wild West Show’ has 11



Figure 3 This screen shows all items filtered for just the current page

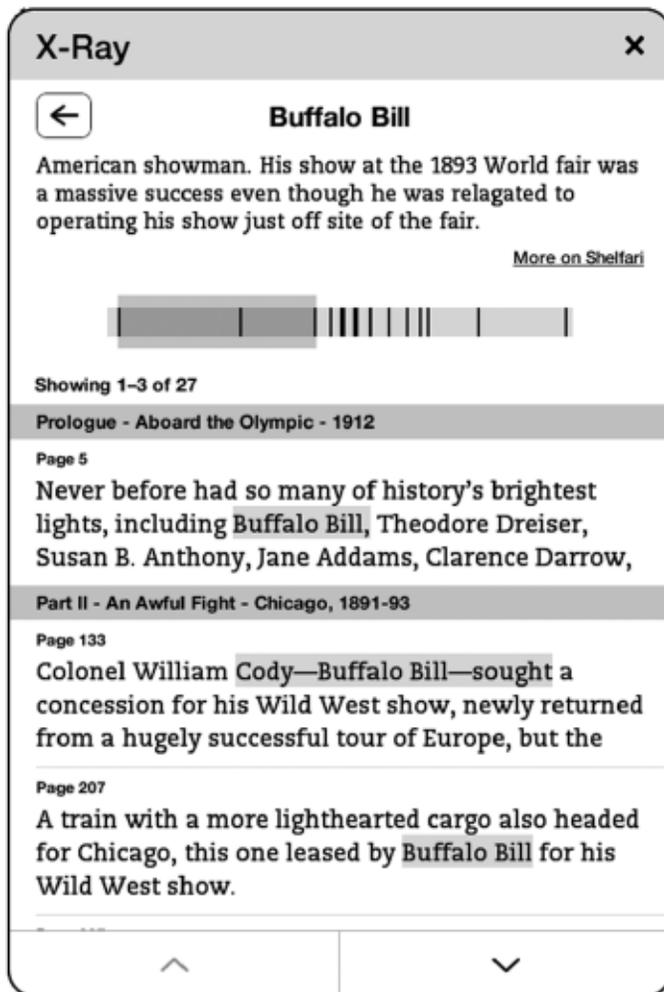


Figure 4 Buffalo Bill’s entry in X-Ray, linked to Shelfari information and pulling it word for word. Shelfari’s full entry is ‘Colonel William Cody (Buffalo Bill)’ but that information has not informed X-Ray.

locators, and ‘Cody, William “Buffalo Bill”’ has 13 locators. The indexer has kindly broken out the show’s specific entries from Bill himself.

In X-Ray, when we search for Buffalo Bill, we see that the feature has failed to grasp that Buffalo Bill and William Cody are the same person. Re-examine Figure 3, and note that X-Ray has them listed separately. The Wild West Show does not appear at all in X-Ray. If we choose the X-Ray entry for William Cody, it is drawn from Wikipedia, and gives a much fuller description than the Shelfari one for Buffalo Bill (compare Figures 4 and 5).

X-Ray appears to be guilty of ‘scattering’ references. If you compare the sparklines, you can see you are getting differing locations.

Daniel Burnham, a major figure in the book, is thoroughly indexed in the print edition and has many X-Ray locators as well (547 instances). The indexer carefully noted his many accomplishments and events in the subheadings of the printed index, as well as distinguishing him from his son, Daniel Burnham. The distinction between father and son did not survive in X-Ray: there is only ‘Daniel Burnham.’

Many people are referenced in X-Ray with only their first or last names, providing the reader with little or no information. In

Figure 6, you can see the plight of a Mr McElroy, who appears only as ‘McElroy’ in X-Ray, and has no other information available. It sounds as though he is involved with a cannon, so surely he deserves a first name.

X-Ray’s treatment of subjects

Like the Wild West Show, many subjects are ignored in X-Ray, and the subjects (or ‘terms’) that are included seem to have been developed by the Mad Hatter. ‘Murder’ does not appear, nor does ‘serial murder’ in any form. Architecture is not in X-Ray, although ‘landscape architecture’ made it in. Considering the egos of the architects involved, if they were alive, nasty notes would be in the mail. The Ferris wheel, which debuted at the fair, is not in X-Ray, although George Washington Gale Ferris is available. The Kindle screen however, cuts off his name so that the reader can see only the beginning: ‘George Washington G . . .’. Any reader who did not know his first name would not pick this entry, and would have to resort to Search, or the entry for Mrs Ferris, no first name.

In the printed index, George G. W. Ferris, his wife Margaret, his Ferris Company, and the amazing wheel are all luxuriously detailed by the indexer.

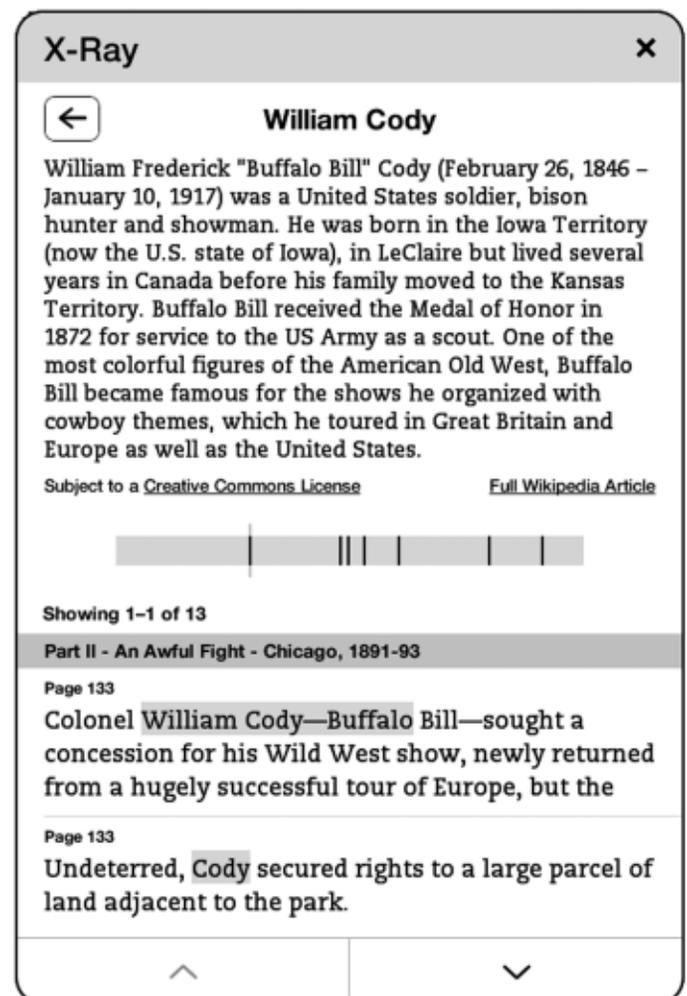


Figure 5 William Cody’s fuller biography

Ferris Wheel, 208, 305, 373
 in construction, 193-94, 218, 228, 236, 239, 240, 255, 280
 final days of, 380-81
 operation of, 279-81, 284, 287-88, 299, 300, 302, 327
 romance on, 306-7
 test runs of, 258-61, 269-73

What a lovely section! Any reader who has any romance in their soul will immediately turn to p. 306 with anticipation.

When filtered by ‘terms’ or subjects, instead of displaying the romance or the lurid instances of murder, X-Ray’s version of the ‘bones’ of this book is very dry and uninviting (see Figures 7 and 8).

Jackson Park, the site of the fair, is not important enough to be in X-Ray, but ‘belly dancers’ and ‘plate glass’ are (Figure 8). Jackson Park was laid out by Frederick Law Olmsted and Calvert Vaux, the landscape architects who created New York’s Central Park. The park has a large Wikipedia entry, and is quite important in landscape architecture’s history. The indexer skipped the belly dancers, mentioned one location for the development of plate glass, and fully described Jackson Park:

Jackson Park, 35, 53-56
 architects’ visit to, 94-96

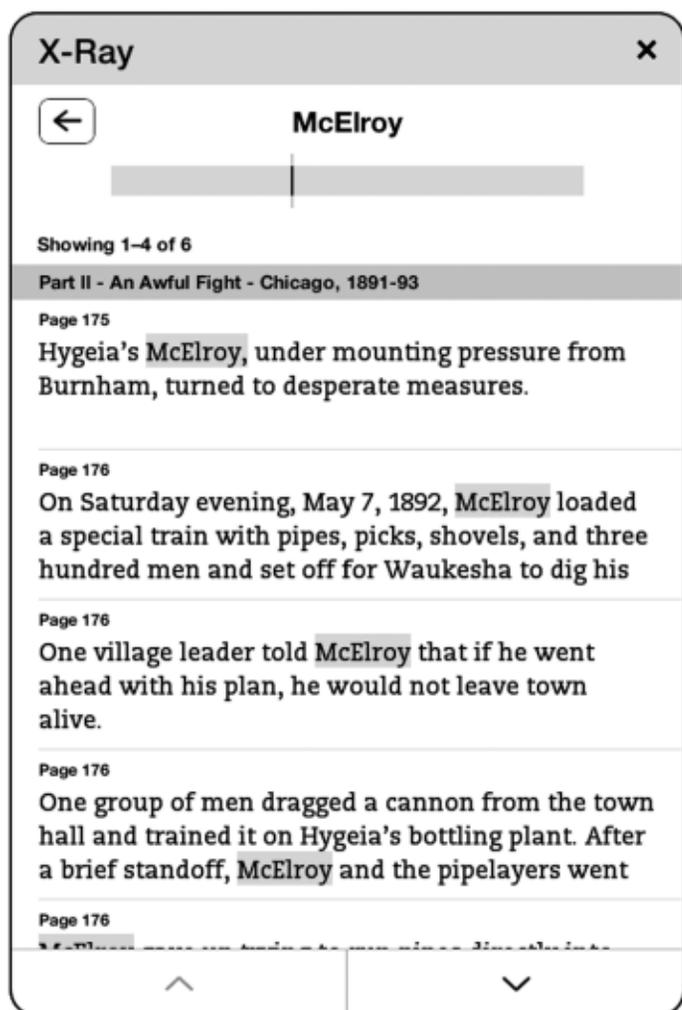


Figure 6 Note that poor Mr McElroy has no first name, no links to information about him, and no description

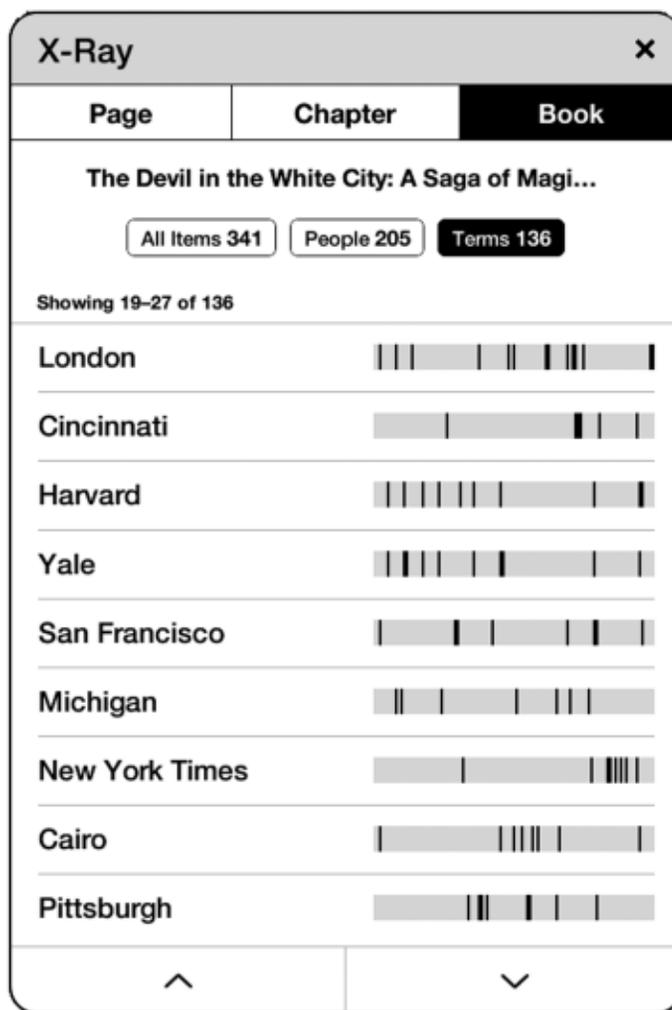


Figure 7 Most of the ‘terms’ appear to be state and city names

entrances to, 253
 as fair site, 73-74, 82, 95-96; see also World’s Columbian Exposition
 permanent structures of, 375
 plans for transformation of, 117-18
 soil of, 129-30, 132, 144, 193

Again, the reader is probably going to be intrigued by ‘soil,’ and wonder why that was so important. Ah, landscape architecture! It is hard to immediately bring a white city into beautiful floral bloom if the soil is bad. If the soil is swampy, as Jackson Park was, buildings and heavy traffic could be a large concern. We can also see that Jackson Park has a relationship to the World’s Columbian Exposition, important enough for a *see also* reference. An X-Ray-reliant reader could miss Jackson Park and its importance completely.

Shelfari’s information

Amazon says that it is ‘using its expertise in language processing and machine learning, access to significant storage and computing resources with Amazon S3 and EC2, and a deep library of book and character information. The vision is to have every important phrase in every book.’ Thus far, we see no real deep access, nor every important phrase revealed.

Looking at the Shelfari² information for the book, it is

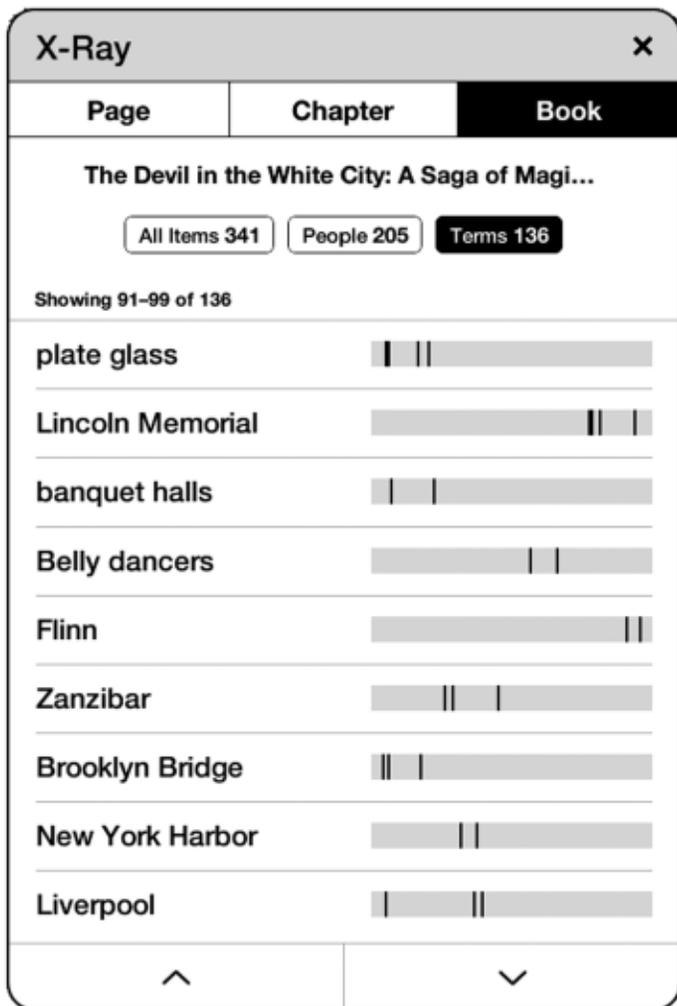


Figure 8 Plate glass and belly dancers do sound enchanting, but Flinn is actually a reference to a name included in the bibliography. Most of these items have very few occurrences. Every ‘important phrase in every book’ does not appear to be here.

clear that Shelfari knows that Buffalo Bill and William Cody are the same person, but that information is not informing X-Ray’s topic choices. Shelfari’s entry for the book makes no mention of the Ferris wheel, belly dancers, or plate glass, but does talk about ‘swampy Jackson Park,’ the Gilded Age,

and murder. It seems as though X-Ray is not making full use of Shelfari’s ‘deep’ knowledge for this particular book, and we have to wonder who chose to have the belly dancers make the cut, while leaving out the Ferris wheel. Shelfari also does not list Mr McElroy, so we have to wonder how he wound up in X-Ray.

Copyright, piracy, and file development?

The X-Ray feature has raised interesting issues beyond the reader’s benefit or lack thereof. Who owns the X-Ray file: the author, Amazon, the publisher? Within days of its release, two hackers figured out how to create your own X-Ray file (sans sparkline graph) and download it to your Kindle. Will that be a violation of some form of copyright? Will Kindle remove these handmade files? Or replace them when they develop their own for a book?

Going forward, who is responsible for ensuring that a book has a good X-Ray file? Again, the author, the publisher, Amazon, or an entrepreneur who might develop them for sale on the web? This is all unclear ground.

A small ten-minute experiment reveals the details of the file. Having decided that Mr McElroy deserved to have at least some information, I edited a copy of the .asc file containing the data (Figures 9 and 10).

After I replaced the .asc file on my Kindle Touch, and reloaded the book, Mr McElroy became a full-fledged character with information (Figure 11, overleaf).

Ramifications for indexers

I was prepared to enjoy the X-Ray feature, especially while reading Steve Jobs’ biography and playing with the feature thoroughly in the process. In the Jobs book, no characters went without their full names, and there were many more meaningful term entries. X-Ray also helpfully offered up search results from the index, allowing readers to see that there was indeed an index, that it was hyperlinked and that it was ready for use. For high-profile books, it is obvious that care has been taken to make the X-Ray experience a good one. And it could be good – it is easy to edit these files. But with lower-profile books, or older bestsellers like *Devil*, care

```
{ "type": "character", "term": "McElroy", "desc": "", "descSrc": "wiki", "descUrl": "http://en.wikipedia.org/wiki/McElroy", "locs": [[438060, 92, 16, 8], [438233, 184, 34, 7], [438795, 99, 24, 7], [439045, 247, 199, 7], [439553, 196, 0, 7], [439891, 123, 101, 7]] },
```

Figure 9 Raw X-Ray data for the McElroy entry before editing. We can see the term, that he is a character, and his locations. The Wiki information leads to a general search in Wikipedia for the word McElroy.

```
{ "type": "character", "term": "J. E. McElroy", "desc": "J. E. McElroy, an entrepreneur heading up the Hygeia Mineral Springs Company. Hygeia contracted to pipe water to the fair.", "descSrc": "wiki", "descUrl": "http://en.wikipedia.org/wiki/McElroy", "locs": [[438060, 92, 16, 8], [438233, 184, 34, 7], [438795, 99, 24, 7], [439045, 247, 199, 7], [439553, 196, 0, 7], [439891, 123, 101, 7]] },
```

Figure 10 I have filled out his initials, and given a brief description of his role in the book

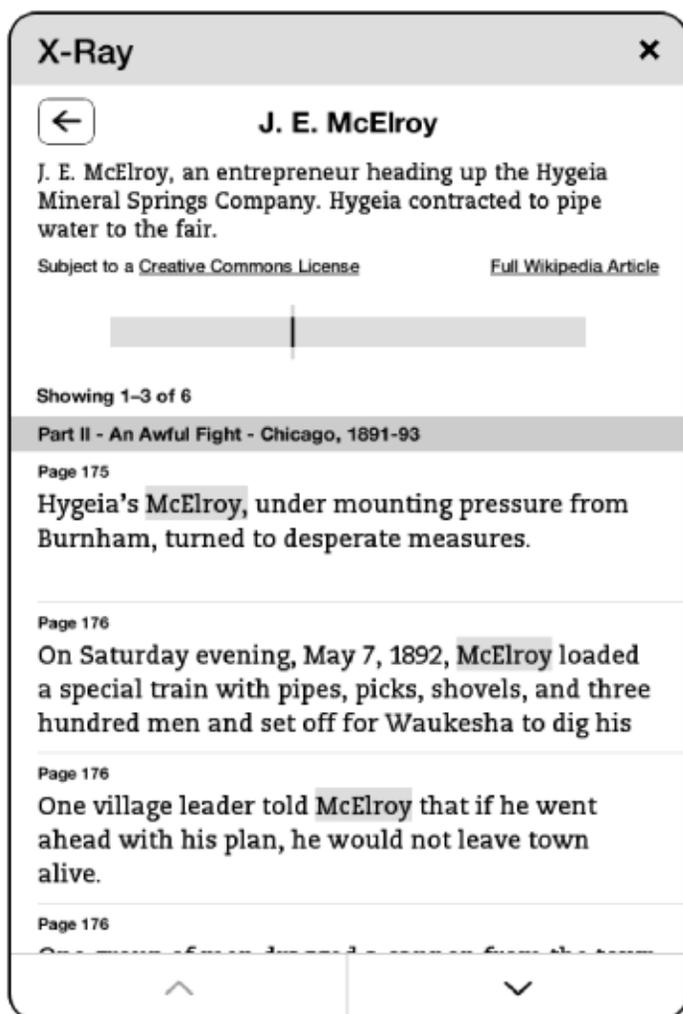


Figure 11 Now J. E. has come alive. So have, mysteriously, links to Creative Commons and Wikipedia. Unfortunately, the Wikipedia link leads still to a generic McElroy search: further editing could resolve this issue. The Creative Commons link leads to general terms and conditions for Creative Commons materials, and nothing about McElroy himself. I can only hope I have not violated something by editing this entry.

is not being taken, and this could leave the reader with far fewer navigational aids. Leaving the index out, and relying solely on Search and X-Ray, gives the reader a substandard experience for *Devil in the white city*. Since the print editions had an index, it is hard to imagine why it was not used to inform X-Ray choices for topics. Time and money may be the devils here.

Amazon Kindles are closed systems, using their own specified formats for ebooks, .mobi and .azw. The coolness factor of X-Ray may be appreciated and used at first by readers, but unless the details get better for all books, readers will be missing some of the picture. Indexes are one of the best navigation tools for a book, print or not: not only does the reader get directly to the piece of information needed, they also may learn a bit, find out differences and details, and get lured into other locations by romance and drama.

If the majority of X-Ray files are not being evaluated by humans before being shipped, it is likely that many books will suffer from lack of X-Ray detail. One promising thought is that Amazon might wake up to see that humans, perhaps

indexers, could make this feature work a lot better, and offer them some work. Publishers might want to ensure that their book is shipped with a good X-Ray file, and not rely on one as poorly developed as this one. In both cases, opportunities for indexers lurk. If not, there are other ebook formats for us to work with, ones that will support indexes.

Throughout this spring of 2012, indexers will continue to be working with the International Digital Publishing Forum on the inclusion of real indexes in the Epub 3.0 ebook format. Some of the features we hope to incorporate are a pop-up index available to the reader at any point in the book, with an area to type in their search, and scrolling features to aid discovery. If implemented as outlined in the Charter Document (see Epub 2011a, 2011b), there could also be a reverse index, showing all the entries for a range of text, much like the X-Ray filter set to Page. But unlike X-Ray, these entries will be index entries, and they will distinguish between Mr Burnham and his son, and they will know that Buffalo Bill is William Cody. Mr McElroy will have his initials. And lastly, they will be sure to include the Ferris wheel and murder in all its lurid details.

If the devil is indeed in the details, X-Ray needs to get much better or it will be bedeviled.

Acknowledgments

Permission to use text from *The devil in the white city* was kindly granted by Erik Larson. Thanks also to Kate Mertes, indexer for the paperback edition of *Devil*.

Notes

- 1 From Amazon's explanation of its new X-Ray functionality: www.amazon.com/Kindle-Touch-e-Reader-Touch-Screen-Wi-Fi-Special-Offers/dp/B005890G8Y/ref=amb_link_359054382_3?pf_rd_m=ATVPDKIKX0DER&pf_rd_s=center-1&pf_rd_r=17F1Q2TRY2AV5QEY1Y68&pf_rd_t=101&pf_rd_p=1337548602&pf_rd_i=507846#xray
- 2 Shelfari page on *The devil in the white city*: www.shelfari.com/books/10065/The-Devil-in-the-White-City

References

- Epub (2011a) Charter document for indexes in ePub 3.0. Available at: <https://code.google.com/p/epub-revision/wiki/IndexesCharterProposal>
- Epub (2011b) Main page for Indexes in ePub 3.0. Available at: <https://code.google.com/p/epub-revision/wiki/IndexesMainPage>
- Larson, E. (2003) *The devil in the white city: murder, magic, and madness at the fair that changed America*. New York: Crown.

Jan Wright has been indexing and taxonomizing since 1991, and has many years of experience in software documentation, help systems, and embedded indexing. An ASI member, she has won several awards from the Society for Technical Communication, and in 2009 won the ASI/H.W. Wilson Award for her index to Real World InDesign CS3, the first time a technical manual had won. She is currently on the IDPF Indexes Working Group Charter team, and is co-chair of ASI's Digital Trends Task Force. Email: jancw@wrightinformation.com

might be in a strong position to propose a solution to this problem by thinking creatively about the appearance of an index in text that has no pages (and, importantly, no page numbering).

The way forward for indexers

If indexers can offer XML-based services, publishers will use them. Learning enough XML to create an embedded XML index will be a prerequisite for joining the digital publishing revolution.

Learning XML is not hard for anyone who can think rigorously. Indexers are good at structure, and fundamentally, XML in publishing is about the imposition of structure on text. Publishers often believe that general text search will overcome the need for an index in digital publishing. This is not the case – and indexers should try to educate publishers to make sure they understand this. In the near future, many books will be published that will never appear in print. Many of those books will need to be indexed.

Nic Gibson is half of Corbas Consulting, which provides digital publishing consultancy and training to publishers. He has worked in and around the media industry since the early 1990s. He started his career in multimedia, doing hypermedia development and progressing into web development and XML. For the past ten years, he has worked on digital publishing projects including workflow transformation projects, eBook development and XML-driven publishing for a variety of trade and academic publishers in the United Kingdom and Europe. Email: nicg@corbas.co.uk

Editorial continued from page 1

The reviews section, not by accident, is oriented towards the indexer’s digital world. And lastly, Pierke Bosschier has pulled together have a lovely section filled with important links and working groups, to help expand your knowledge in all of these areas.

Our biggest challenge is to keep track of book publishing innovations, and quickly develop ways to incorporate indexes as important navigation and metadata in each innovation we see. We add value and metadata. The index can come along for the ride into every interface as long as we understand the technology used for presentation. We can let go of the page, and think of locators as just that: locators. We do not care if they are anchor points in HTML, unique ID strings in an InDesign file, or time codes in a video. The human analysis of the aboutness at that location is our unique contribution. The technology for presenting that location and accessing its content can change, but we are still the best resource for letting readers know whether a location is worth the visit, and where else they can look.

I urge you to find out for yourself how the current crop of ereaders display indexes. Find, buy, or borrow an ereader, whether it is a Kindle, Nook, Kobo, or iPad. Look at book indexes on the readers, such as Steve Jobs’s biography, the SI publication *Indexing children’s books*, the ASI publication *Marketing your indexing services*, Browne and Jerney’s *Indexing companion* at <http://amzn.to/ynrn8Y>, or the OASIS Open Office Specification at <http://bit.ly/wcjsLg>. Each of these chapter-like indexes was developed in a different way, and each has differing success with access: links to the page, or to the paragraph. How should we use locators? What is the simplest path for the reader? Examine the interface your reader will be using to look at your index, and think about how to make it work better. Read the resource documents at <http://bit.ly/uqKwD7>. We illustrate ways forward in these documents, and indexers can use them in their discussions with clients.

Let’s all take a look at what we want indexes to be, and throw out ideas about how to get there. It is still early in the game.

[See too the note from Maureen MacGlashan on page 56.]

indexing specialists

Our team of experts have many years’ experience in undertaking and designing effective indexes

- *Books, journals, manuals* ●
- *Multi-volume works* ●
- *From 100,000 to 100 lines* ●
- *Unfamiliar subjects* ●
- *Design of style and layout* ●
- *Short deadlines met* ●

Phone +44 (0) 1273 416777
email indexers@indexing.co.uk
website www.indexing.co.uk

INDEXING SPECIALISTS (UK) LTD
306A Portland Road, Hove
East Sussex, BN3 5LP, UK

