

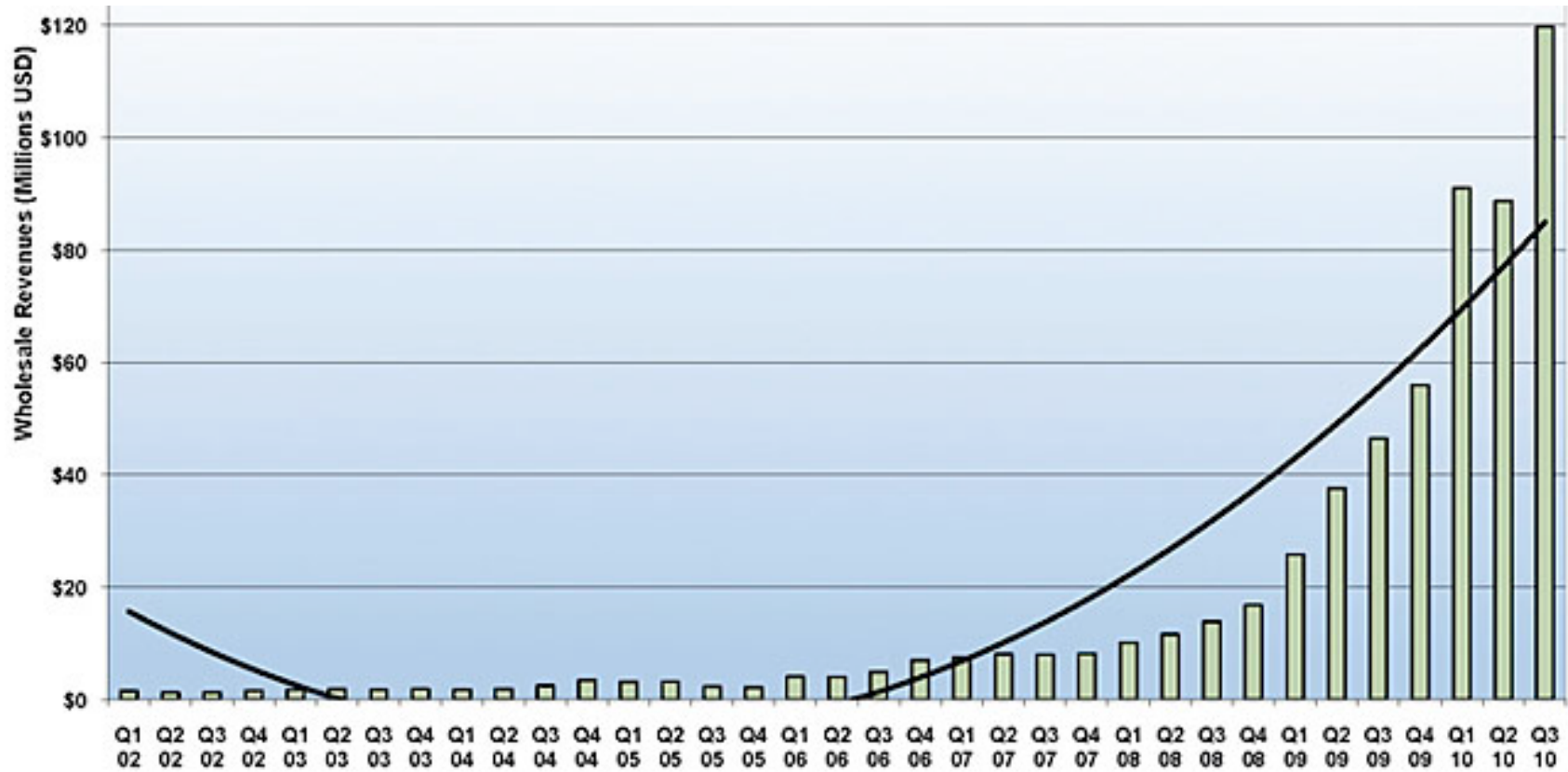


Ebook Indexes

Changes are happening fast!

Jan Wright
Wright Information Indexing
Services
www.wrightinformation.com

Statistics on growth



International Digital Publishing Forum/Association of American Publishers
Quarterly US trade Retail eBook Sales

BookStats from AAP



- e-books have grown from 0.6% of the total Trade market share in 2008 to 6.4% in 2010. This translates to 1274.1% in publisher net sales revenue year-over-year with total net revenue for 2010 at \$878 Million.
- Net unit sales growth for e-books was equally impressive, increasing 1039.6% for the same three-year period. In 2010, e-book net units were 114M.
- Beyond the top-level format figures, the explosive growth of e-books is even more visible when considering certain categories. In Adult Fiction, e-books are now 13.6% of the net revenue market share.
- <http://publishers.org/bookstats/formats/>

Amazon eBook sales



- Kindle eBook sales surpassed Amazon's hardcover sales back in July 2010
- Kindle eBook sales surpassed paperback sales in January of 2011.
- In May 2011, eBooks have bested both of them combined.
- Note that this does not include the free books that many Kindle users take advantage of — these are all paid books.
- (*Tech Crunch <http://techcrunch.com/2011/05/19/that-was-fast-amazons-kindle-ebook-sales-surpass-print-it-only-took-four-years/>*)

App books



- Disney - 1 million downloads of children's book apps for the Apple iOS (iPad, iPhone)
- iPhone and iPad apps are a financial hit. Apple announced in January 2011 its 10-billionth downloaded app.
- Bundles: eBook with associated app or website are commonplace now, and will only grow and be expected.
- Sales of print editions go up: The Elements: 186,000 app copies sold at \$14.00 = \$2,604,000. First print ed. = 100,000 copies; after app, sales the print edition went to 300,000
- (<http://www.teleread.com/ebooks/the-elements-ipad-app-helps-sell-printed-books/>)

Academic eBook findings

- Undergrads prefer print books by 58%
- Ability to search within and across content is the primary advantage of e-books
- Annotating and highlighting is vital and not working
- Borrowing or purchasing print copies of is common.
- Students use eBooks for search and discovery tasks (TOC and browse), then move to print editions for reading, note taking, text comparison, deep study, better navigation.
- Undergrads want print copies available for borrowing from the library, with 66% rating it as important.
- 41% rate the option to buy a “print-on-demand” copy as important



Not a pretty picture



- Textbooks: hard to use if there aren't page numbers displayed and good chapter breaks: instructors and students have issues matching assignments
- Taking notes is not easy yet, figures are hard to see, looking up references a pain.
- Less than 40% of UW students used the Kindle regularly for classroom reading.
- And no one is talking about the index.
- <http://blogs.forbes.com/alexknapp/2011/05/19/what-do-amazons-e-book-sales-mean-for-the-future-of-books/>

Current eBook indexes



- Missing
- Static
- Linked, but not to precise paragraph
- Index as chapter at end, not integrated navigation tool
- Hard to get to
- Hard to return and search again
- Hard to browse, especially if left in 2-column format

Using indexes now



Courtesy of Joshua Tallent

AbeLincoln

T

Tandem mass spectrometry (TMS),
genetic testing, [264-265](#)

Tanzi, R., [239](#)

Tay-Sachs disease
carrier screening, [266-267](#)
prenatal diagnosis, [310](#)

Tellegen, A., [123-124](#)

β -Thalassemia, gene therapy, [252](#)

Tiwari, B., [189-190](#)

TMS. See [Tandem mass spectrometry](#)

Tobacco, transgenic plant production of
therapeutics, [166-168](#)

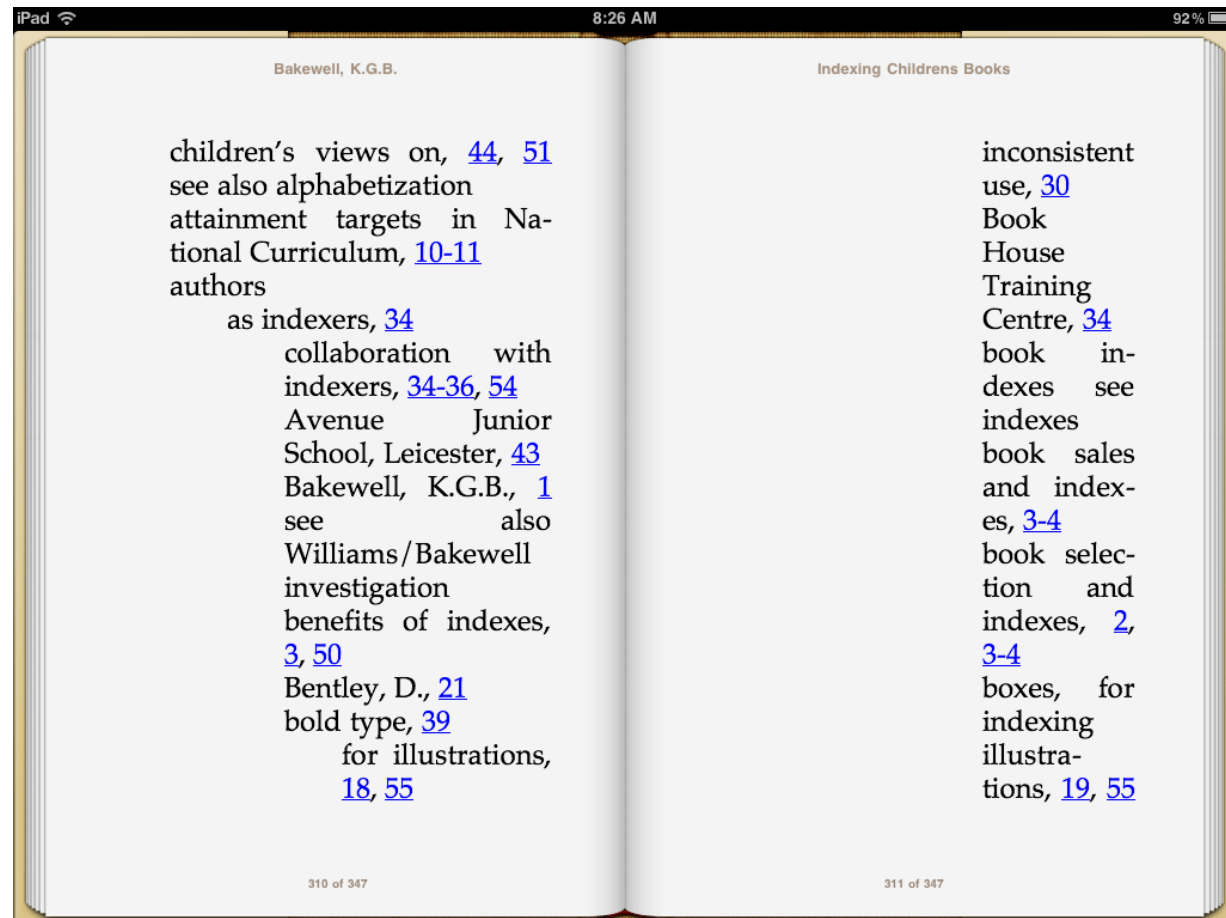
Tolins, J., [145](#)

Toulouse-Lautrec
alcohol abuse, [33](#), [36-37](#)

.....

Locations 5648-55 OFF Menu ▶

Things can go wrong



New paradigm



- As important as “Find”
- Navigation in nonfiction – critical!
- Adds browsability and helps readers expand their ideas
- Memory and recognition aid, and returnability
- Needs to be included in eventual interfaces for eReaders
- Re-use - Source of metadata and aboutness at a paragraph level

Browsing an index is important



- Amazon's "Look Inside This Book" as sales tool
- Browsing all of the index
- Mode one: full chapter-like browsing display
- Mode two: combined with "Find" and results shown first
- Mode three: underlying semantic guide

Indexes provide reader assistance



- Pre-analysis
- One of main routes in
- Conversation with the reader
- Helping hand for phrasing concepts
- Assistance when skimming is difficult

Example of critical need

- Grammar Girl's books



that

as redundant, 170–71

which versus, 42–43

who versus, 43–44

the

double, 176

pronunciation of, 7–8

then versus *than*, 42

there is versus *there are*, 44–45

starting a sentence with, 81

Reader's mental patterns and search behavior



- Readers come to texts at differing times
- Never read the book, and need to know if it contains their search concept
- Read the book, and knows the concept is in there, but where?
- Asking a question, and when reading information, rephrases or narrows or changes the question
- Search has stages, and we need to serve all the stages – it is as if we are conversing at each stage, not just feeding more Google results

Cognitive mapping



- Reading styles for learning or pleasure differ
- For learning, readers use physical cues such as location on page or position in book to go back and find sections of text
- eReaders disrupt that mapping
- Skimming is another behavior that is disrupted
- Marking up text in textbooks helps memory – difficult but becoming easier on eReaders
- Navigation for mathematicians will differ from historians or bird watchers
- <http://www.washington.edu/news/articles/college-students2019-use-of-kindle-dx-points-to-e-reader2019s-role-in-academia>

Ideal eBook indexes



- Accessible from every page
- Mode 1: full index browsability as chapter
- Mode 2: combined in a search window with a type-in field, with index results shown first
- Mode 3: semantic underpinnings – map of topics
- “Find” should be reflecting the best hits, as identified by index
- Show results as snippets of text in context
- Searched word should show entries below, subheads, closely-related words
- Cross references to help refine terms

Returning to the index for next step of search

- Index remembers last lookup
- Snippet feature reduces need to return to index



Index

Search for the keyword:

polydactyl

pleasure

body language 39-40
and purring, 39

points, Siamese, 12

polydactyl, 49, 50 79

ponytail plant, 65

pothos, 65

potpourri, 64

predictability and cats 106

pregnancy

cats, 323

humans, toxoplasmosis, 197

prey. See hunting

private breeders. See breeders

probiotics, 316-317

protein

cat food, low quality, 222-223

in diabetic diet, 240

kidney disease diet, 315

kitten needs, 242

senior cat needs, 243

49

Polydactyl Cats Cats normally have five toes on their front paws, but some American **polydactyl cats** have as many as ... **Polydactyl cats** can be found in any breed, but American **polydactyl cats** are bred specifically for their extra toes. ...

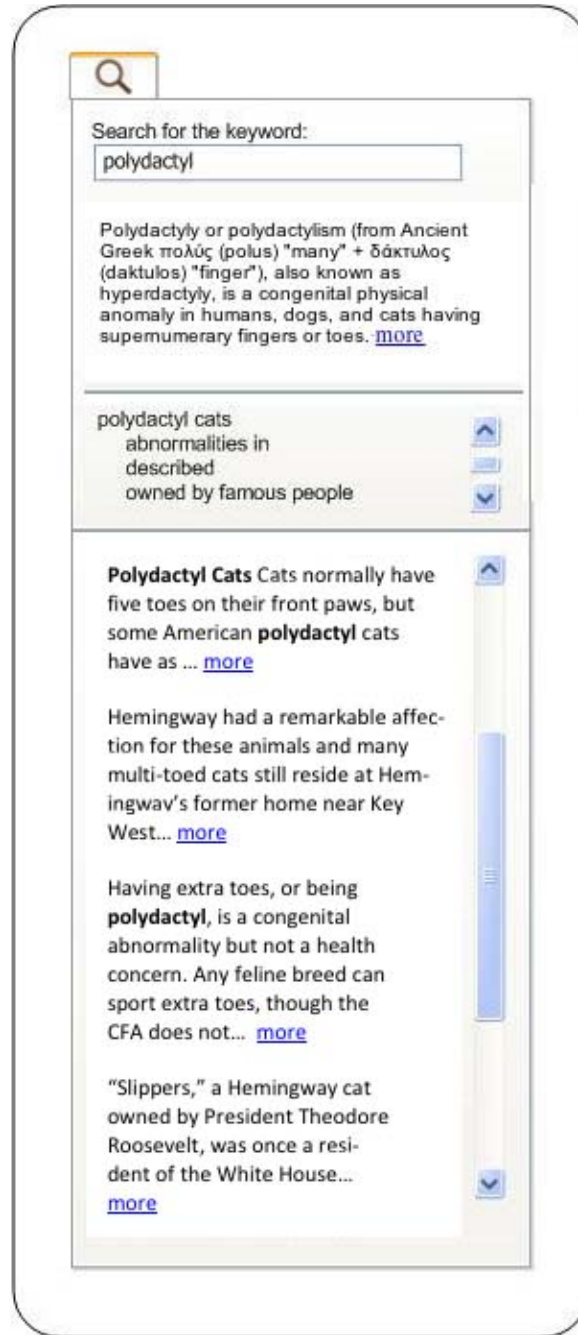
50

He had a remarkable affection for these animals and many **polydactyl cats** still reside at Hemingway's former home near Key ... "Slippers," a **polydactyl cat** owned by President Theodore Roosevelt, was once a resident of the White House. ...

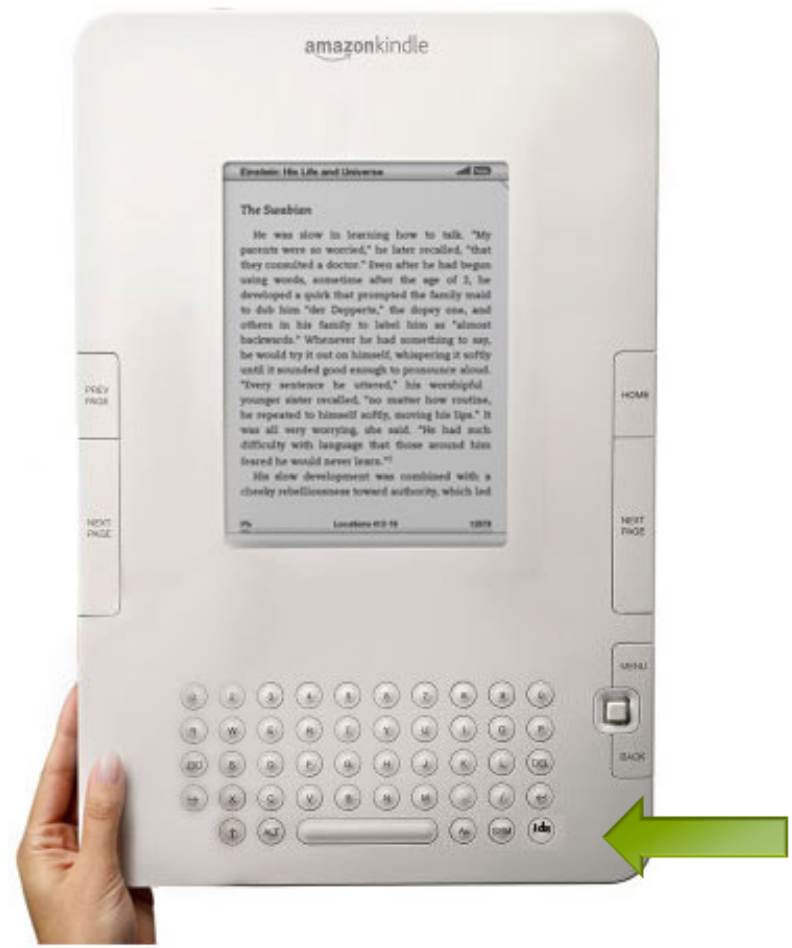
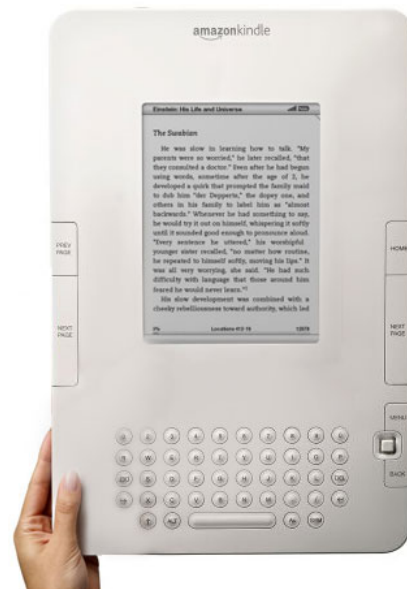
79

Having extra toes, or being polydactyl, is a congenital abnormality but not a health concern. Any feline breed can sport extra toes, though the CFA does not register **polydactyl cats** for showing. Legend has it that seventeenth-century ...

New combined interface



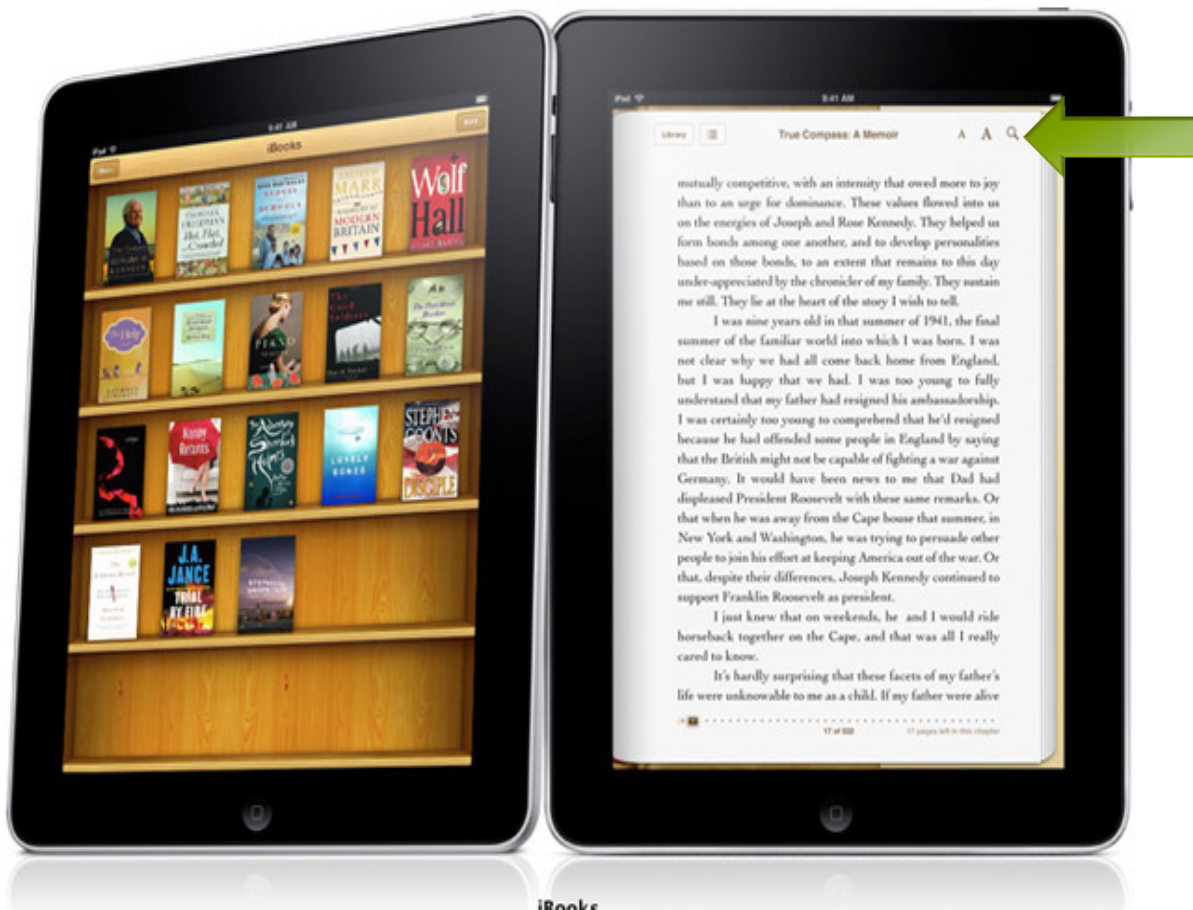
Some interface ideas



If only the Kindle would add a nice button for the Index.....



LetsGoMobile



This is the current iPad eReader interface. Notice there is a search icon on the upper right. Let's use that for more than search!

New standards



- ePub 3.0 includes scripting: pathway to index functionality
- Index appearance on small screens
- Avoid long strings of locators
- More information in subheads
- Single locator for each subhead may be needed
- Alphabetization and ignoring prepositions may not work
- Page ranging may not work
- Embedding may be needed

What can be done now by publishers



- Get ready for eventual use of scripts and anchor IDs in ePub 3.0
- Develop an anchor ID scheme, and add IDs within books to be ready
- Put in anchor codes at the paragraph level for index entries
- Include active indexes now as chapters
- Look at interactive index interfaces to be ready for reader or app support
- Plan for re-use of metadata: wikis, handhelds, print, web pages

Anchor codes

- ePub3 allows addressing by anchor code (a hidden unique string) at almost any level of the book
- Pinpoint accuracy for references and linking
- Working towards references to other books



Anchor IDs illustrated



```
<a name="parks001" />The history of  
parks in Cleveland is a long and  
complex one. One of the first parks to  
be created was the Emerald Necklace  
in 1776, remarkable considering that  
Cleveland was sparsely populated at  
that date. Later the city added the  
Squire's Castle, known for its unusual  
ghostly creaks and door-slamming  
ethereal visitors. Later, Euclid Creek  
was established
```

Parks

Emerald Necklace [1](#parks001)

ePub CFIs



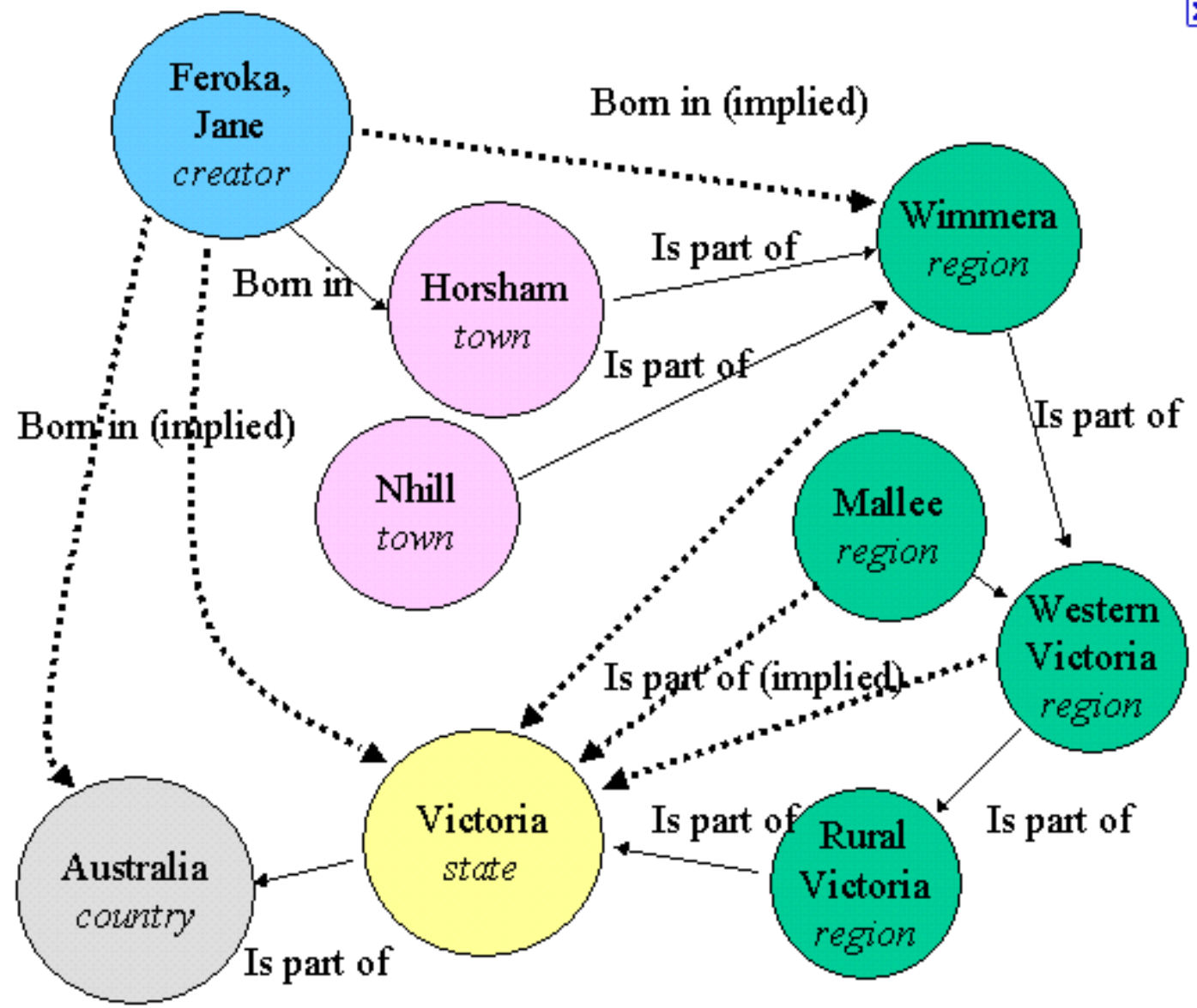
- Precision locations to the word in an epub
- Could be used for highlighting ranges of text when a page range locator is clicked
- Good news is precision
- Bad news is implementation: nothing supports this technology yet on either eReaders or book layout software

```
#epubcfi(/6/4[chap01ref]!/4/10[foozle])
```

Semantic metadata



- Allows additional metadata within books
- Additional metadata about books for stores to develop into finding tools
- Add metadata within text to add intelligence – semantic marking: definitions, related terms, related topics
- We could have another view of the index, as a topic map! Visual interactive view



ePub 3.0 Scripting



- JavaScript will be supported for:
 - Interactivity
 - Enhanced layouts (indexes?)
 - Formatting
 - Improved user experience (navigation?)
- TEST TEST TEST: indexers will need to be able to test the interactivity portions
- TEST what happens if the reader doesn't support JavaScript

The help we need

- Software tools
- Programming resources
- Outreach efforts
- Quality assurance/testing procedures for scripted indexes and interfaces
- Buy-in from eBook and eReader industry
- Membership on standards committees
- Knowledge: every indexer needs to know what to suggest to publishers



Resources



- **ASI DTF LinkedIn group**
http://www.linkedin.com/groups?gid=4005509&trk=hb_side_g
- **ASI DTF proposal for ePub 3.0:**
<http://dl.dropbox.com/u/2248375/IDPF%20Index%20functionality%20in%20ePub.pdf>
- **BNA study on index usability:**
<http://dl.dropbox.com/u/2248375/Using%20BNA%20Indexes%20study.pdf>
- **O'Reilly's What is ePub 3.0? Booklet:**
http://dl.dropbox.com/u/2248375/What_Is_EPUB_3_.pdf
- **This presentation:**
<http://dl.dropbox.com/u/2248375/ABQ%20epub.pdf>

More Resources

- The Status of ebook Indexing:
<http://dl.dropbox.com/u/2248375/The%20status%20of%20eBook%20indexing%20%E2%80%93%20Jan%20Wright.pdf>
- eBookarchitects:
<http://www.ebookarchitects.com>
- ePub 3.0 Standard:
<http://idpf.org/epub/30/spec/epub30-overview.html>
- Aptara Seminar: EPUB3: An Insider's Look:
<http://www.aptaracorp.com/resources/category/webinars/>
- http://www.cdlib.org/services/uxdesign/docs/2011/academic_ebook_usage_survey.pdf
- <http://www.washington.edu/news/articles/college-students2019-use-of-kindle-dx-points-to-e-reader2019s-role-in-academia>



Slides available at

<http://dl.dropbox.com/u/2248375/ABQ%20epub.pptx>

A business card with a green background and rounded corners. On the left side, there are several overlapping circles of varying shades of green. On the right side, the name 'Jan Wright' is printed in a large, bold, black font. Below the name, the company name 'Wright Information Indexing Services' is printed in a smaller, bold, black font. Underneath that, the address 'P.O. Box 658' and 'Sandia Park, New Mexico 87047' are printed in a smaller, bold, black font. The website 'www.wrightinformation.com' and the email address 'jancw@wrightinformation.com' are printed at the bottom in a smaller, bold, black font.

Jan Wright

Wright Information Indexing Services

P.O. Box 658

Sandia Park, New Mexico 87047

www.wrightinformation.com

jancw@wrightinformation.com