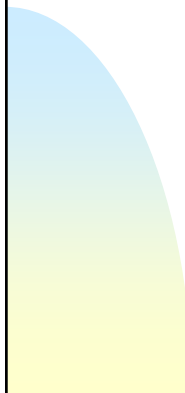


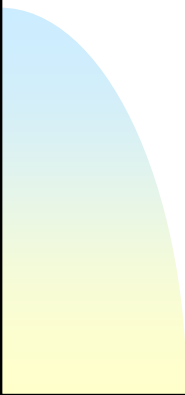
Web Indexes

David K. Ream
© Leverage Technologies, Inc.
www.LevTechInc.com



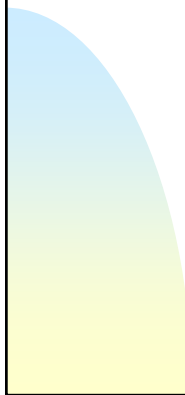
Overview

- Introduction
- Print vs. Web
- Demonstrations
- Summary



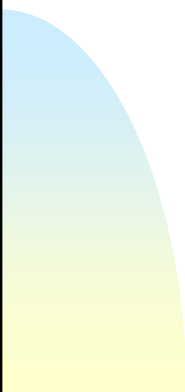
Introduction

- Definition
- Design Considerations
- Contrasts
- Technologies




Web index definition

- preparing a print index for the web
 - ◆ with no links
 - ◆ linking cross references
- indexing the content of a web site
- indexing multiple web sites
- database/keyword indexing
- full text indexing or MAI
- classification/metadata



Index design

- structural elements
- space
- style & layout
- navigation
- process
- locators
- scripting



Structural elements (print)

- title & headnotes
- letter/group breaks
- continuation headings or guide words
- headings
- editorial notes
- cross references
- locators

Structural elements (web)

- approach pages
- separate pages for each letter/group
- heading context clues
- headings
- editorial notes
- cross references as internal links
- locators as external links
- handling of multiple locators

Space (print)

- two-page spread
- page size & margins
- number of columns
- point size & line spacing
- publisher determined
- cost constraints: printing & distribution

Space (web)

- screen resolution
- browser window
- single column
- frames
- proportional scaling
- window sizing
- varies from user to user
- cost constraints: index creation

Style (print)

- leading before breaks & headings
- fonts
- headings: run-in or indented
- cross references:
 - ◆ attached or separate lines
 - ◆ multiples: in paragraphs or on individual lines
- locators: single or multiple

Style (web)

- blank lines
- fonts, colors, display attributes
- headings: avoid run-ins
- cross references: separate lines
- locators: what is the link text/value?
- avoid multiple locators if possible

Navigation (print)

- placement within publication
- turning pages
- guide words
- continued headings
- letter breaks
- alphabetic context
- visual overview

Navigation (web)

- how to get to the index?
- separate pages for each letter?
 - ◆ previous/next letter links/buttons
 - ◆ letter lists
- cross reference linking
- find (on this page)
- full text search of index
- loss of context
- frames

Process (print)

- get publisher's requirements
- get manuscript or file
- create index
- locator sort/number check
- return file or pages
- composed by publisher

Process (web)

- is style known/different from print?
- access to web pages?
- link values known?
- link testing?
- return HTML page files
- uploaded by publisher

Locators (print)

- locator is the link!
- page: 16, 23-45, 58*n*, 67 *et seq.*
- chapter/page: 3-7, 6-2 to 6-17
- section: 1.3, 1301.55, 6.17[A][1][a]
- serial
 - ◆ volume issue: 4:11
 - ◆ chronological: June 2000

Locators (web)

- link text (user) vs. link value (computer) or unlinked
- page filename
 - ◆ with a target
 - ◆ to a specific point within an HTML page
 - ◆ to invoke a database query
 - ◆ transformable locators
 - ◆ unique/arbitrary number
- Universal Resource Locator (URL)

Page numbers (web)

- pages are non-existent
- reconnect entries to tops of paragraphs
- differentiate lists of page numbers
 - ◆ add link text
 - ◆ add subheading levels for each
- ranges can be meaningless

Scripting

- switching between multiple indexes
- navigation (between index pages)
- searching & viewing hits
- context indicators
- database queries
- security
- “subscribe now”

URL syntax

- scheme (http://)
- authority (www.LevTechInc.com)
- path (/directory/subdirectory/...)
- filename (support.html)
- target
 - ◆ #anchor
 - ◆ ?query
- inheritance or basing

URL scheme

- http (hypertext transfer protocol)
- https (secure http)
- ftp (file transfer protocol)
- mailto (invokes e-mail)
- etc.

URL authority

- www. support. ...
- site name
- .com .org .edu .gov ...
- .us .ca .uk .es .de ...

URL examples

- http://www.LevTechInc.com
 - ◆ default.htm, index.html
- .../htmlprep.htm
- .../greeks.pdf
 - ◆ .jpg, .asp, .xml
- .../mjixa.htm#mh6
- http://www.socind.demon.co.uk

URL examples *cont.*

- http://www.microsoft.com/data
- .../content.asp?id=690?prod=66
- mailto:info@LevTechInc.com
- http://www.wellingtonnewzealand.co.nz/glacierguides/photography/foxglacier.html

Link text

Canines. *see* [Dogs](#)

Dogs

[Viszlas](#)

[Weimaraners](#)

Link value

```
<p><b>Canines.</b> <i>see</i>
<a href="#dogs"><b>Dogs</b></a></p>
```

```
<p><a name="dogs"></a><b>Dogs</b><br>
&nbsp;&nbsp;&nbsp;<a href="vizz.htm">Vizslas</a><br>
&nbsp;&nbsp;&nbsp;<a
href="weim.htm">Weimaraners</a></p>
```

Multiple locators

Cats, [♦](#), [♦](#), [♦](#), [♦](#)

Dogs, [1](#), [2](#), [3](#), [4](#)

Fish, [4:3](#), [6:15](#), [11:8](#)

Horses, [jumping](#), [showing](#),
[training](#)

Llamas

[description](#)

[pictures](#)

Technologies

- markup languages
- web browser
- HTML editor
- dedicated indexing software
- HTML indexing software
- custom software

Markup languages

- SGML (structural)
 - ◆ Document Type Definition (DTD)
 - ◆ document instance
 - ◆ HTML (display)
- XML (structural/display)
- scripting

DTD for a book

- title page
- table of contents
- introduction
- chapter(s)
 - ◆ paragraphs
 - ◆ footnotes
- index
- about the author

DTD for a journal issue

- title page
- table of contents
- editorial
- article(s)
 - ◆ title
 - ◆ author(s)
 - ◆ paragraphs
 - ◆ footnotes
 - ◆ references
 - ◆ biographies

Instantiations

- book
 - ◆ dogs
 - ◆ cats
 - ◆ horses
- journal issue
 - ◆ January
 - ◆ February

Tag types

- structure or display tags (<>)
 - ◆ mark begin and/or end of document elements
 - ◆ optional attributes
- entity references (&;)
 - ◆ special characters
 - ◆ accented letters
 - ◆ greeks

Tag examples

- paragraph: <p> ... </p>
- meta data:
 - ◆ <meta name="KEYWORDS" content="indexing software, CINDEX">
- anchor:
 - ◆ ...
 - ◆ ...

Entity reference examples

- ° °
- Á Á
- ¿ ¿
- no-break space

Tagging structure

- HTML (display)
 - ◆ <h1>
 - ◆ <h2>
 - ◆ <h3>
- XML (structure)
 - ◆ <chapter>
 - ◆ <section>
 - ◆ <subject>

Tagging content

- HTML (display)
 - ◆ `<i>in situ</i>`
 - ◆ `<i>Smith v. Jones</i>`
 - ◆ He `<i>really</i>` meant it!
- XML (structure)
 - ◆ `<forphr>in situ</forphr>`
 - ◆ `<casnam><plaintiff>Smith</plaintiff>`
v. `<defendant>Jones</defendant>`
`</casnam>`
 - ◆ He `<emph>really</emph>` meant it!

Demonstrations

- journal index
- fiction index
- history index
- legal index
- article index
- "3D" web directory

Summary

- various kinds of web indexes
- difference between print and web
 - ◆ appearance & style
 - ◆ print locators and web links
 - ◆ process to create indexes
- requirements
- implications

Requirements

- more tools
- more skills & knowledge
- more time
 - ◆ initially training/learning
 - ◆ job turnover
- more involvement up front
- team work

Implications

- new service to existing clients
- new markets with new clients (intrawebs)
- more frequent revisions, frequency of cumulations
- more income!!