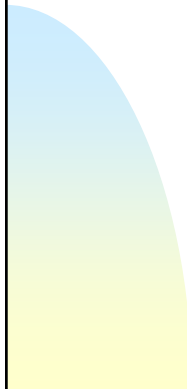


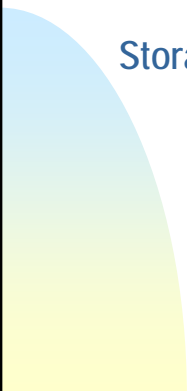
An XML DTD for Indexes

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



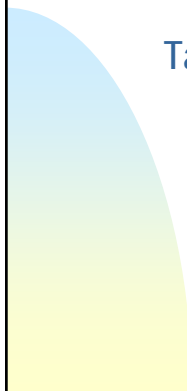
Overview

- History of tagging (abridged)
- DTD goals
- Display DTD & example
- Data DTD & example
- Summary



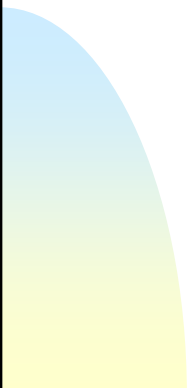
Storage of Publishing Data

- Fixed / fielded formats
- Tagging
 - ◆ Within fields
 - ◆ Entire documents
- Types of tags
 - ◆ Structural (containing the data)
 - ◆ Integral (within the data)



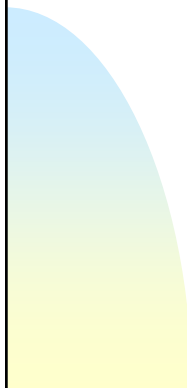
Tagging Schemes

- Diverse: product-specific & custom
- Quark, Ventura, Tex/LaTeX, RTF
- Inconsistent conversions
- Not portable across platforms
 - ◆ operating systems
 - ◆ software
 - ◆ hardware



SGML

- Standard
- Generalized
- Markup
- Language
- ISO 8879 (1986)



SGML

- Defines types of documents
- Each type specifies its tags
- Consists of
 - ◆ Document Type Definition (DTD)
 - ✦ Elements and their attributes
 - ✦ Entities (non-keyboard data)
 - ◆ Document instances
- Instances can be parsed

SGML tagging models

- Meta language
- Public
 - ◆ AAP, UCP (structure tags)
 - ◆ ISO (character entities)
 - ◆ DTD fragments (ie, for an index)
- Private
 - ◆ Publisher specific

SGML Tools

- Editor
- Parser
- Processing programs
 - ◆ Extractions
 - ◆ Reports
 - ◆ Media (CDs, eBooks, Web)
- Composition (print)
- Arbortext, Framemaker, Omnimark

Markup Languages Today

- Product-specific & legacy schemes
- RTF (de facto standard)
- SGML (content)
- HTML (hypertext; display)
- XML (extensible; content & display)
 - ◆ late 1990s
 - ◆ set of tools & specifications

DTD for Indexes

- Goal: single DTD
 - ◆ Variety of index styles
 - ◆ Print & Electronic media
 - ◆ Display & Data forms
- Software support
 - ◆ preparation software
 - ◆ cdXml conversion application

XML DTD

- defines a tree of nested elements
- <idx:index>
 - ◆ creator
 - ◆ created.datetime
 - ◆ sort.method

XML DTD

- <idx:display>
 - ◆ <idx:group>
 - ◆ <idx:level>
 - ◆ <idx:entry>
 - <idx:heading>
 - <idx:locator>
 - <idx:xref>
 - <idx:ednote>

XML DTD

- <idx:data
 - ◆ <idx:record
 - ✦ attributes
 - creator
 - created.datetime
 - editor
 - edited.datetime
 - ✦ <idx:field

XML DTD

- namespaces
- <idx: applies to structure
- <gen: applies to "data"
 - ◆ <gen:style (bold, italics, underline, subscript, superscript, small caps)
 - ◆ <gen:font
 - ◆ <gen:sort (suppression, inclusion)

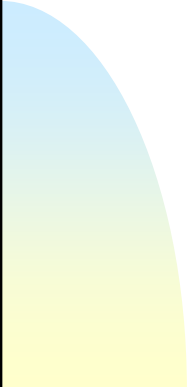
XML DTD

- ASCII (7-bit, keyboard) for basic data content
- character entities
 - ◆ UNICODE numeric entities
 - ✦ é or é
 - ◆ ISO named aliases
 - ✦ é
 - ✦ & < >

Display Index Example

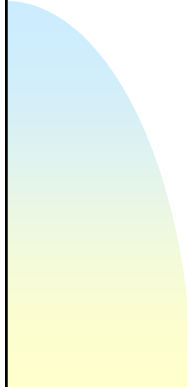
Data Index Example

DTD Walkthrough



It's just a model!

- Punctuation
- Cross reference placement
- Structures vary across indexes
- Full locators vs. suppressed parts
- Integration with publisher's DTD
 - ◆ Unique tag names
 - ◆ Link tags and attributes



Summary

- *SGML, An Author's Guide*, Martin Bryan, Addison-Wesley, 1988
- www.w3c.org (HTML)
- www.itwriting.com/xmlintro.php (XML intro with list of resources)
- Slides, DTD, samples, utility on www.LevTechInc.com under Resources/Presentations