

A DTD for Indexes

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

Overview

- History of tagging (abridged)
- Index DTD goals
- Embedded DTD & example
- Compiled 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, Pagemaker, Tex/LaTeX, RTF
- Inconsistent conversions
- Not portable across platforms (OS, S/W, 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

- Public
 - ◆ Does not dictate tag names
 - ◆ AAP, UCP (structure tags)
 - ◆ ISO (character entities)
 - ◆ DTD fragments (ie, for an index)
- Private
 - ◆ Publisher specific
- Portability?

SGML Tools

- Editor
- Parser
- Processing programs
 - ◆ Extractions
 - ◆ Reports
 - ◆ Media creation (CDs, eBooks)
- 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)

Index DTD

- Goal: single DTD
- Variety of index styles
- Print & all Electronic media
- Compiled & Embedded forms
- Software support
 - ◆ editing tools
 - ◆ index compiler

Compiled Index DTD

- <index type=, id=
 - ◆ <info
 - ♦ <title
 - ♦ <creator
 - ♦ <head.notes
 - ♦ <release.date
 - ♦ <revision.number
 - ♦ <coverage.info
 - ♦ <content.descriptors
 - ◆ <body

Compiled Index DTD

- <body
 - ◆ <group id=
 - ♦ <level number=, id=
 - <locator
 - <cite ref=
 - <see
 - <cross ref=
 - <see.also
 - <cross ref=
 - <ed.note

