

And What of Indexes?

Patrick Durusau, Snowfall Software

Topic Maps 2007

21 March 2007, Oslo



Disclosure

- Standards Lead, Snowfall Software
- ISO: Co-Editor of ISO 13250-1, 5 (topic maps), Acting Convenor of SC 34/WG 3
- ODF: Co-Editor of ISO 26300, ISO Project Editor, Co-Editor of ODF 1.2, Chair and Co-Editor, ODF Metadata Subcommittee
- Chair, V1, Technical Advisory Group (TAG) to US National Body



Why Index?

- If you can't find information, it may as well not exist
- Full text searching? Sure, if you like Google Glut results
- Same subject, different descriptions
- Related information (and how related)
- A “view” of the text and its subjects



Topic Maps

- Independent Information Resources
- Web portals
- Other examples...
- But what of indexes?
 - TaxMap is one example



What to Index?

- Interesting
- Structured Content (to ease automatic construction)
- Useful for further integration



Hittite Text Archive?

- Interesting?
- Structured content?
- Further integration?
- < 200 full time Hittitologists world wide



Alt.sex Newsgroups?

- Interesting?
- Structured content?
- Further Integration?
- But export/import laws and squeamish marketing department



OOXML, DIS 29600?

- Interesting?
- Structured Content?
- Further Integration?
- Under review, lots of interest!



Current Indexing in OOXML

- Part 1: 173 pages – 3 1/3 page index
- Part 2: 129 pages – 1 page index
- Part 3: 472 pages – No index
- Part 4: 5,219 pages – No index
- Part 5: 43 pages – 1 page index
- So, 6,036 pages and a total of 5 1/3 pages of indexing, or 5,691 with no indexing at all.



Subjects in OOXML

- Markup elements
- Relationships between markup elements
- Examples of markup elements
- Use of markup elements in other elements
- Attributes of markup elements
- Values of attributes
- Other rules



More Subjects and OOXML

- But, people are writing about OOXML
- Comments by National Bodies
- Instructional materials (with their own subjects as well)
- Email discussions
- Endless blog postings
- Noted for completeness, here working with the published text of OOXML



Building a topic map of OOXML

- Advantage: OOXML is available in OOXML (a markup format)
- Permits the automatic building of associations between sections and subjects within them
- Human authoring can refine or add more subjects not extracted automatically
- Becomes applicable to OOXML documents generally



First Pass Subjects

- Part
- Section
- Sub-Sections
- Elements in all of the above
- Attributes
- Relationship between element and where it occurs (for navigation)



Skip to Demo!

- Added here to cut off boring theory for jump to comparison of standard indexing in Adobe and Topic Map approach





User: Patrick Durusau

Help Logout Contact Us

Search

Create Account T/M Space Self Forestgate Family Neighborhood Work Extended Friends

Development

FROS

Book Study

spaces:ooxml

OOXML

- Foreword
- Introduction
- Scope
- Conformance
 - Goal
 - Issues
 - What this Standard Specifies
 - Document Conformance
 - Application Conformance
 - Interoperability Guidelines
- Normative References
- Definitions
- Notational Conventions
- Acronyms and Abbreviations
- General Description
- Overview
 - Packages and Parts
 - Consumers and Producers
 - WordprocessingML
 - SpreadsheetML
 - PresentationML
 - Supporting MLs
- Packages
 - Relationships in Office Open XML
 - Constraints on Office Open XML's Use of OPC
- Markup Compatibility and Extensibility
 - Constraints on Office Open XML's Use of Markup
 - PreserveElements and PreserveAttributes
 - Office Open XML Native Extensibility Constr
- WordprocessingML
 - Glossary of WordprocessingML -Specific Term
 - Part Summary
 - Package Structure

Conformance

- Syndicators +
- Communities +
- Individuals +
- Experiences +
- Bundles +

Syndicators Communities Individuals Experiences Bundles





User: Patrick Durusau

Help Logout Contact Us

Search

Create Account T/M Space Self Neighborhood Forestgate Extended Family Work Friends



Development



Book Study

spaces:ooxml

OOXML

OOXML

Document Template

Source Relationship:

<http://schemas.openxmlformats.org/officeDocument/2006/relationships/attachedTemplate>

A [document template](#) can be represented by an instance of a WordprocessingML package, and contains styles, numbering definitions, and so on that are made available when documents based on that template are edited. A WordprocessingML document can refer to another document as its document template, by having a Document Settings part (§11.3.3) that contains an explicit relationship to the file location of the necessary document template using the id attribute on the [attachedTemplate](#) element.

[Example: Consider a document specifying a document template located at c:\template.docx:

```
<Relationships xmlns="..."> <Relationship Id="rId1" Type="http://.../attachedTemplate" Target="file:///c:/template.docx" TargetMode="External"/></Relationships>
```

The document's Document Settings part contains an [attachedTemplate](#) element that explicitly references this relationship:

```
<w:settings ... > <w:attachedTemplate r:id="rId1"/></w:settings>
```

end example]

Syndicators

- productivity.snowfallsoftware.com

Communities

- Self
- Family
- Forestgate
- Extended
- Work
- Neighborhood
- Friends

Individuals

- Patrick Durusau

Experiences

- OOXML
- Shared Resources
- Calendar
- Email
- Note
- Task
- Chat
- Contacts

Bundles

Syndicators

Communities

Individuals

Experiences

Bundles



Development



Book Study

spaces:ooxml

OOXML

- OOXML
 - P1
 - P2
 - Genera
 - + Notatic
 - + Packag
 - Norma
 - Scope
 - Definiti
 - + Physic
 - O vervi
 - A crony
 - Forew
 - + Digital
 - + Thumt
 - + Core P
 - P3
 - P4
 - P5

- ISOSpecs
 - ISO 639 Codes for...
 - ISO 8601:2004 Da...
 - ISO/IEC 10118-3 I...
 - ISO/IEC 8632 Infor...
 - ISO 216 Writing pa...
 - ISO 8879 Informati...
 - ISO/IEC 15445:20...
- Elements
- Terms

- Syndicators
- Communities
- Individuals
- Experiences
- Bundles

Syndicators

Communities

Individuals

Experiences

Bundles

Summary of Screen Shots

- Navigation by term and by relationships
- Dynamic display of relevant content (not a feature of topic maps but of this interface)
- Can add subjects and relationships between subjects
- Automatic generation augmented by human authoring (without syntax)



Forthcoming Features (commercial)

- Export to XTM (other XTM application would have to support OOXML for viewing)
- Browser and standalone versions
- Either version can synchronize with others, all or part of the “local” topic map
- Integration of other information resources, email, local code, etc.



Conclusion

- Augmented indexing for texts
- Integration of multiple sources of comments on one text
- Integration of other information resources with the text
- Integration of a text with local resources (like programming code with reference manuals)



Questions?

