EPUB Tools for Techcomm

tcworld conference 2013 - Wiesbaden, Germany

Scott Prentice, Leximation, Inc.

Introduction

- Scott Prentice, President of Leximation, Inc.

- Specializing in FrameMaker plugin development as well as structured FrameMaker conversions, consulting, and development. FrameMaker user/developer since 1991.

- Developed DITA-FMx, a FrameMaker plugin for efficient DITA authoring and publishing.

- Consulting for custom Help systems, creative/functional web applications, and EPUB solutions.

What is EPUB?

- EPUB is an ebook file format that can be read on almost every device and platform

- Requires a "reader" application or dedicated reader device

- A collection of XHTML, XML, CSS, and media files wrapped up in a "zip" archive

- Defines the content and formatting of the ebook

Brief history of EPUB

EPUB specification is maintained by IDPF

Initial EPUB format approved in 2007, superseding Open eBook standard

EPUB 2.0.1 approved 2010

EPUB 3 approved October 2011

EPUB 3 Fixed Layout format approved May 2012

EPUB 3 Dictionaries/Indexes Working Groups, in progress

EPUB and techcomm?

- Probably not for everyone .. yet

- Best for "linear" content

- Table-heavy docs may not be a good candidate

- Likely not the "primary" deliverable, but nice to offer

- If you're single sourcing already, why not add one more deliverable?

- No better way to make your docs available everywhere

EPUB 2 or EPUB 3?

- EPUB 3 is current spec, but full support still weak
- EPUB 3 adds interactivity, multimedia, more (via HTML 5)
- To take full advantage of EPUB 3 you'll have to hand code, no tools provide dynamic features (yet)
- If you do include EPUB 3 features, it will limit the readers
- Use simple EPUB 3 or EPUB 2

Formatting

- Rule #1 .. Keep It Simple!

- Your goal is to make the content "available" not to make a finely crafted work of art .. right?

- The least formatting produces the best looking EPUB on the widest array of readers

- Embedded fonts? Don't bother (IMHO).

- Use CSS selectors not @style attributes

Reader support

- Dedicated ereader devices
  - Kobo, Kindle, Nook, Sony, ...

- Mobile devices (phones/tablets) with reader apps
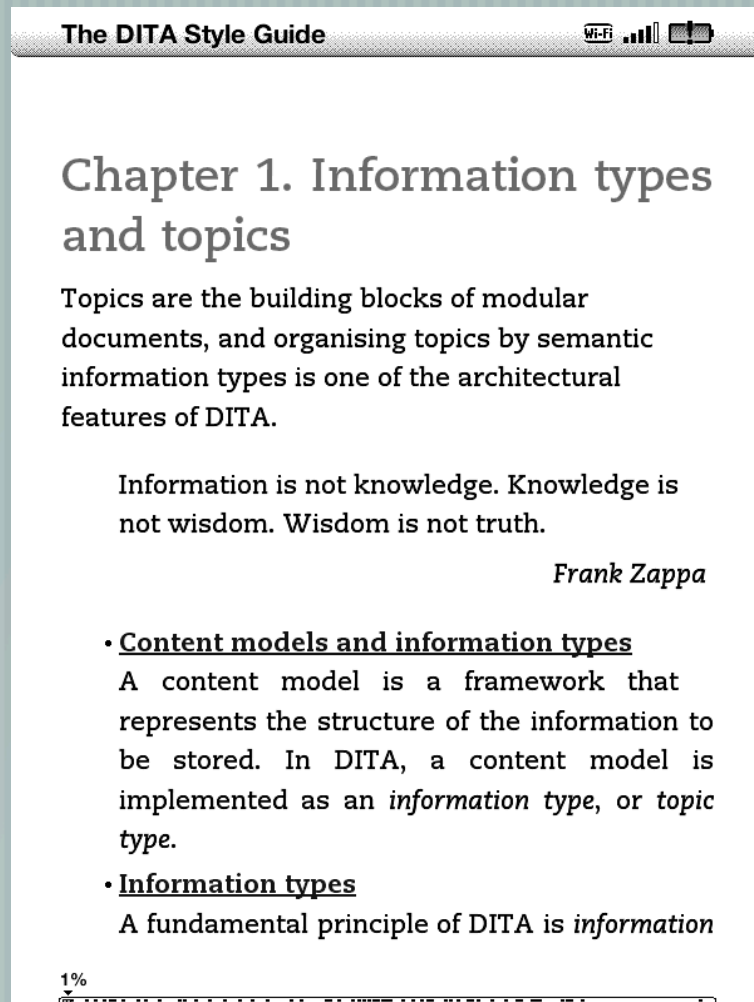 - iBooks (iOS), Aldiko (Android), Kindle, Kobo, Nook apps, ...

- Desktop reader apps
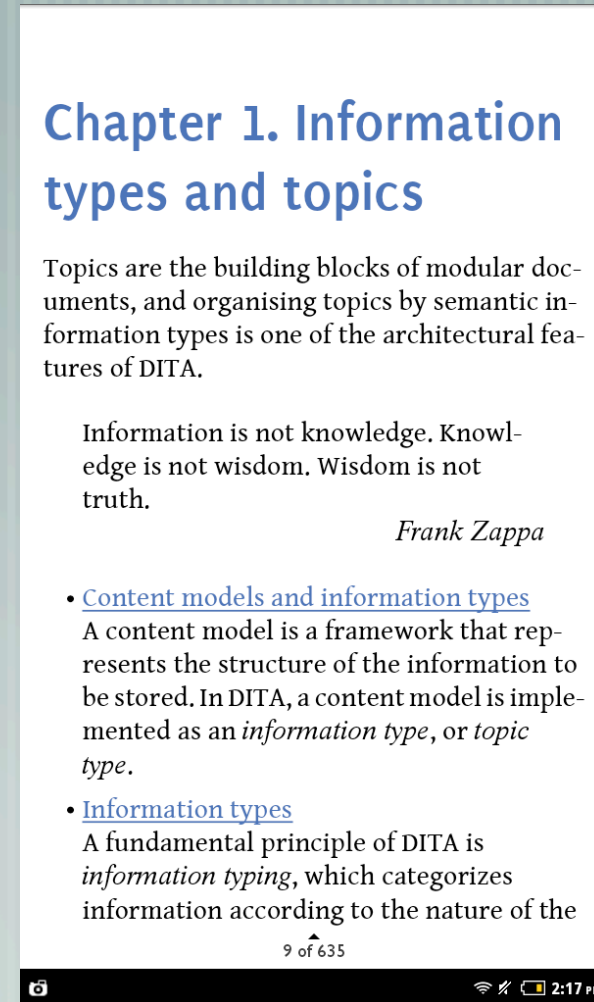 - Firefox EPUBReader, Chrome Readium, Nook, ADE, Calibre, ...

- Each EPUB will look different in each reader

The following images use the "DITA Style Guide" by Tony Self, published by Scriptorium Press (used with permission)
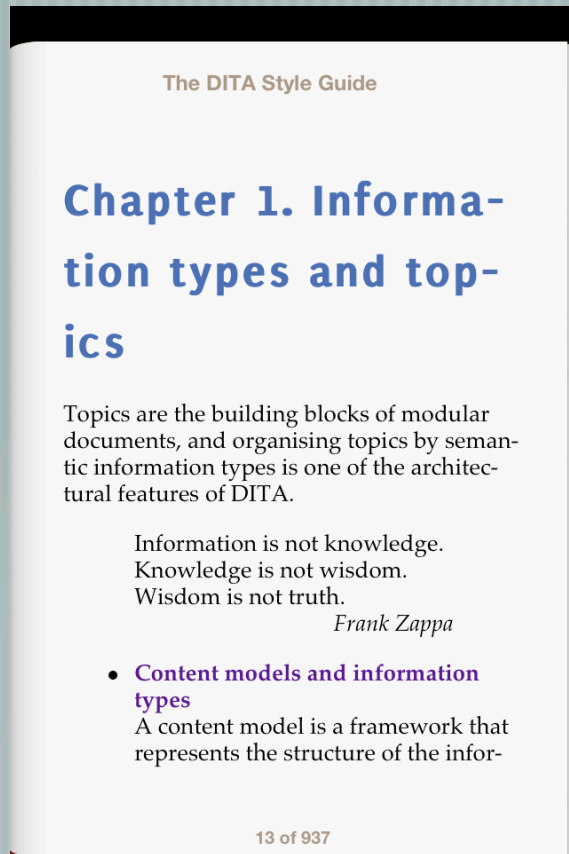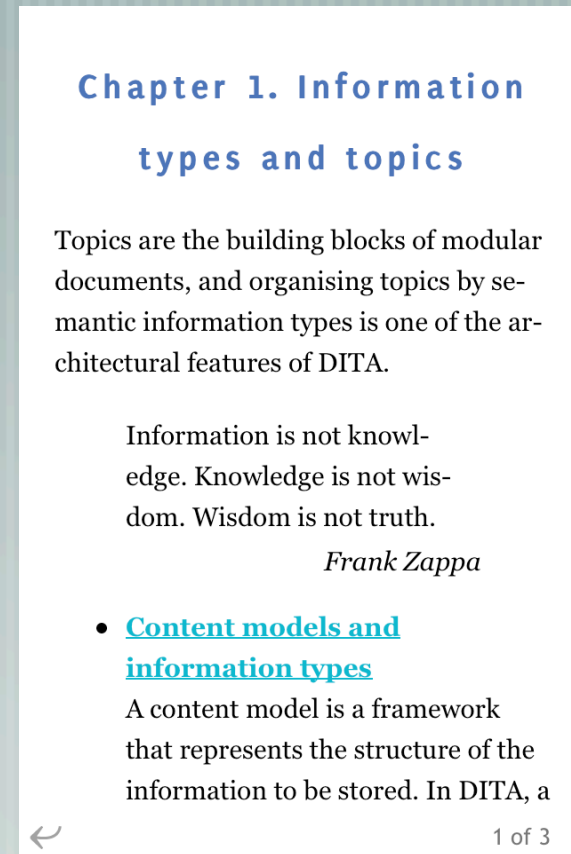
Dedicated ereaders

Chapter 1. Information types and topics

Topics are the building blocks of modular documents, and organising topics by semantic information types is one of the architectural features of DITA.

> Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth.
>
> *Frank Zappa*

- **Content models and information types**
  A content model is a framework that represents the structure of the information to be stored. In DITA, a content model is implemented as an *information type*, or *topic type*.
- **Information types**
  A fundamental principle of DITA is *information*

1%

Kindle 4 basic

Chapter 1. Information types and topics

Topics are the building blocks of modular documents, and organising topics by semantic information types is one of the architectural features of DITA.

> Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth.
>
> *Frank Zappa*

- Content models and information types
  A content model is a framework that represents the structure of the information to be stored. In DITA, a content model is implemented as an *information type*, or *topic type*.
- Information types
  A fundamental principle of DITA is *information typing*, which categorizes information according to the nature of the

9 of 635

2:17 PM

Nook Color

Mobile device + reader app

iPhone + iBooks app

iPhone + Kobo app

Mobile device + Kindle app

Galaxy S3 + Kindle app

Chapter 1. Information types and topics

Topics are the building blocks of modular documents, and organising topics by semantic information types is one of the architectural features of DITA.

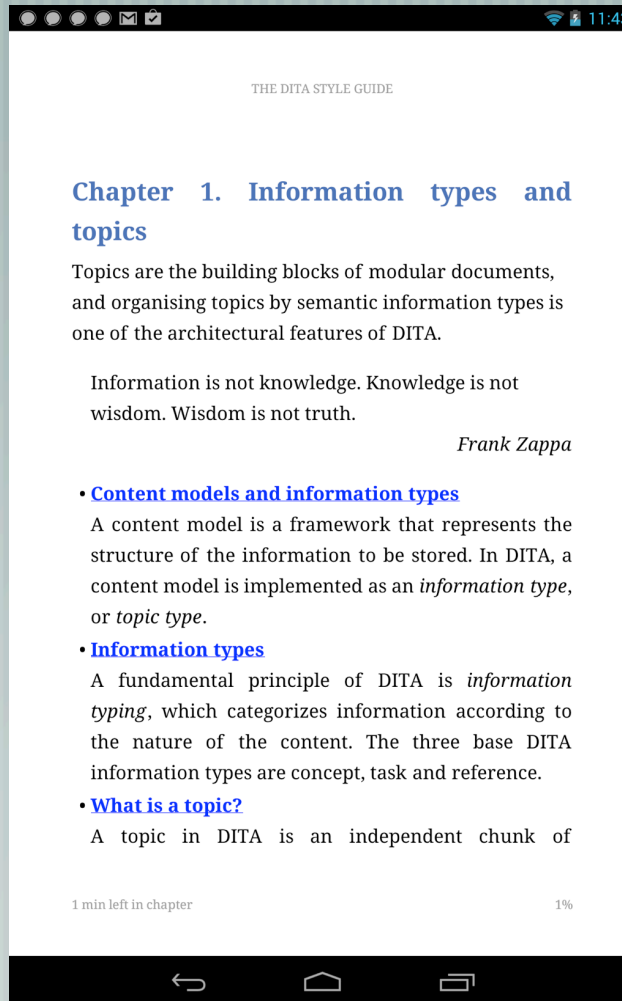Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth.

Frank Zappa

- **Content models and information types**
  A content model is a framework that represents the structure of the information to be stored. In DITA, a content model is implemented as an **information type**, or **topic type**.
- **Information types**
  A fundamental principle of DITA is **information typing**, which

Nexus 7 + Kindle app

Chapter 1. Information types and topics

Topics are the building blocks of modular documents, and organising topics by semantic information types is one of the architectural features of DITA.

Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth.

Frank Zappa

- **Content models and information types**
  A content model is a framework that represents the structure of the information to be stored. In DITA, a content model is implemented as an *information type*, or *topic type*.
- **Information types**
  A fundamental principle of DITA is *information typing*, which categorizes information according to the nature of the content. The three base DITA information types are concept, task and reference.
- **What is a topic?**
  A topic in DITA is an independent chunk of

iPhone + Kindle app

Chapter 1. Information types and topics

Topics are the building blocks of modular documents, and organising topics by semantic information types is one of the architectural features of DITA.

Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth.

Frank Zappa

- **Content models and information types**
  A content model is a framework that represents the structure of the information to be stored. In DITA, a content model is implemented as an *information type*, or *topic type*.

Desktop reader apps

Chrome + Readium

Firefox + EPUBReader

Desktop reader apps

AZARDI

AZARDI :: The DITA Style Guide – Tony Self

Chapter 1. Information types and topics

Topics are the building blocks of modular documents, and organising topics by semantic information types is one of the architectural features of DITA.

> Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth.
>
> *Frank Zappa*

- **Content models and information types**
  A content model is a framework that represents the structure of the information to be stored. In DITA, a content model is implemented as an *information type*, or *topic type*.
- **Information types**
  A fundamental principle of DITA is *information typing*, which categorizes information according to the nature of the content. The three base DITA information types are concept, task and reference.
- **What is a topic?**
  A topic in DITA is an independent chunk of information covering a single idea or task, authored as a unit, and stored in its own file. Topics are categorized by information type. DITA's topic-based architecture enables the same topic to be used in different publications.
- **Working with topics**
- **Specialized information types**
  Specialized information types provide a topic architecture that closely matches the structure of your content.

Nook desktop

The DITA Style Guide

Chapter 1. Information types and topics

Topics are the building blocks of modular documents, and organising topics by semantic information types is one of the architectural features of DITA.
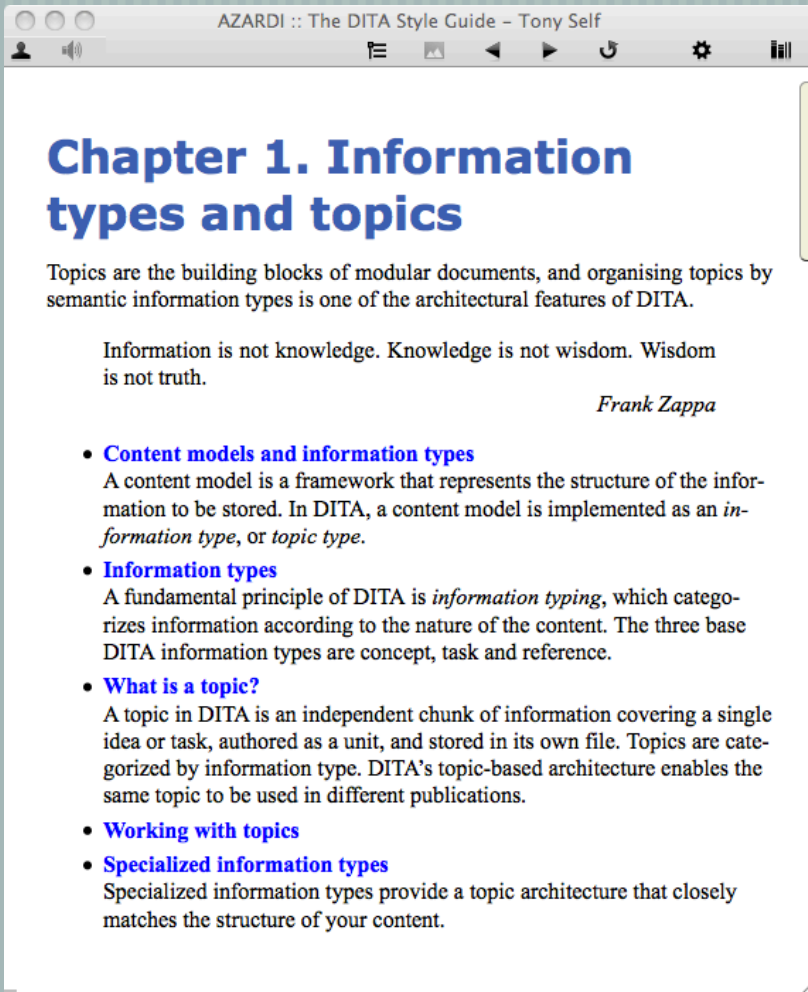
> Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth.
>
> *Frank Zappa*

- Content models and information types
  A content model is a framework that represents the structure of the information to be stored. In DITA, a content model is implemented as an *information type*, or *topic type*.
- Information types
  A fundamental principle of DITA is *information typing*, which categorizes information according to the nature of the content. The three base DITA information types are concept, task and reference.
- What is a topic?
  A topic in DITA is an independent chunk of information covering a single

Page 9   (9 of 635)

Desktop reader apps

Adobe Digital Editions

Adobe Digital Editions – ditastyleguide.epub

Library

Table of Contents | Bookmarks

- About The DITA Style Guide
- Chapter 1. Information types and topics
- Chapter 2. DITA map files
- Chapter 3. Syntax and mark-up
- Chapter 4. Language and punctuation
- Chapter 5. Graphics and figures
- Chapter 6. Cross-referencing
- Chapter 7. Content re-use
- Chapter 8. Metadata, conditional processing...
- Chapter 9. The DITA documentation process
- DITA authoring concepts
- Index
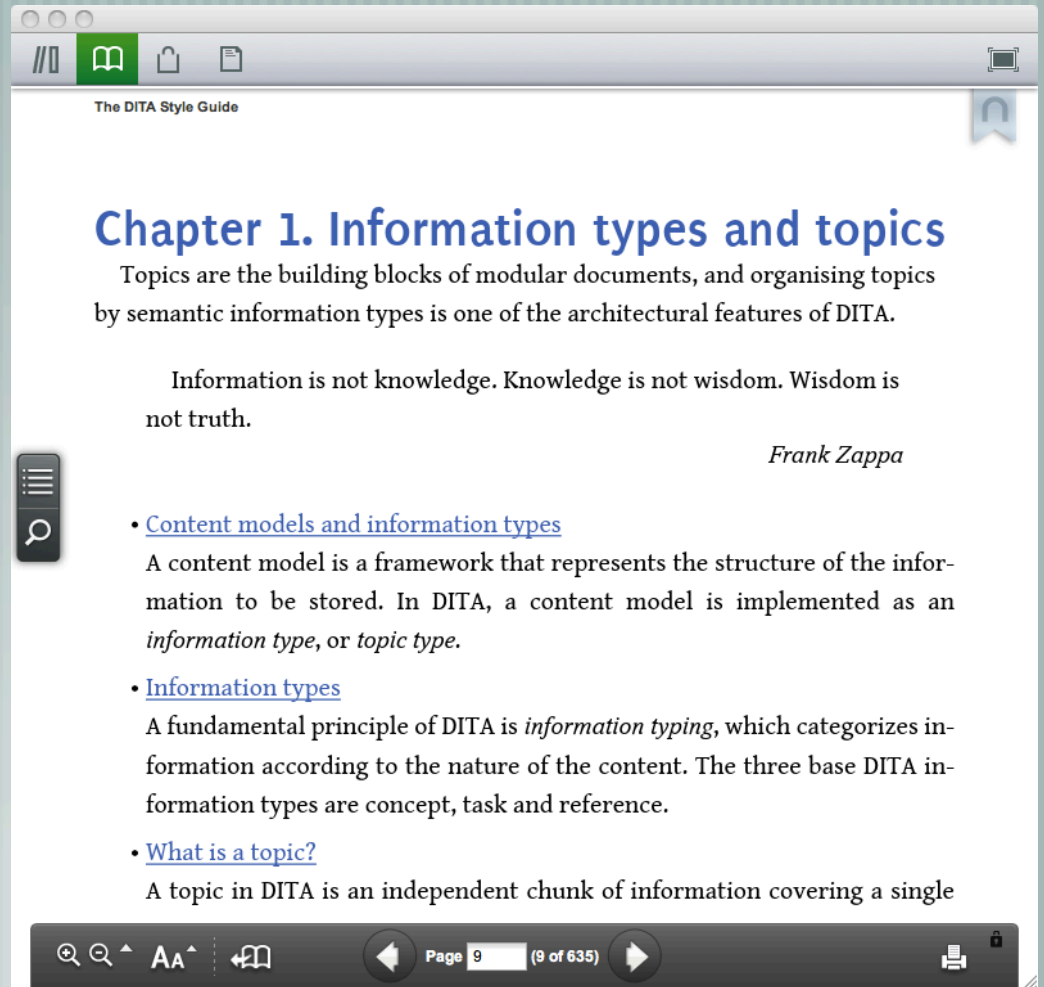- Copyright

Chapter 1. Information types and topics

Topics are the building blocks of modular documents, and organising topics by semantic information types is one of the architectural features of DITA.

> Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth.
>
> *Frank Zappa*

- **Content models and information types**
  A content model is a framework that represents the structure of the information to be stored. In DITA, a content model is implemented as an *information type*, or *topic type*.
- **Information types**
  A fundamental principle of DITA is *information typing*, which categorizes information according to the nature of the content. The three base DITA information types are concept, task and reference.
- **What is a topic?**
  A topic in DITA is an independent chunk of information covering a single idea or task, authored as a unit, and stored in its own file. Topics are categorized by information type. DITA's topic-based architecture enables the same topic to be used in different publications.
- **Working with topics**
- **Specialized information types**
  Specialized information types provide a topic architecture that closely matches the structure of your content.

9 / 635

Calibre

E-book Viewer – The DITA Style Guide [EPUB]

10.0 / 636   Go to...   Search

Chapter 1. Information types and topics

Topics are the building blocks of modular documents, and organising topics by semantic information types is one of the architectural features of DITA.

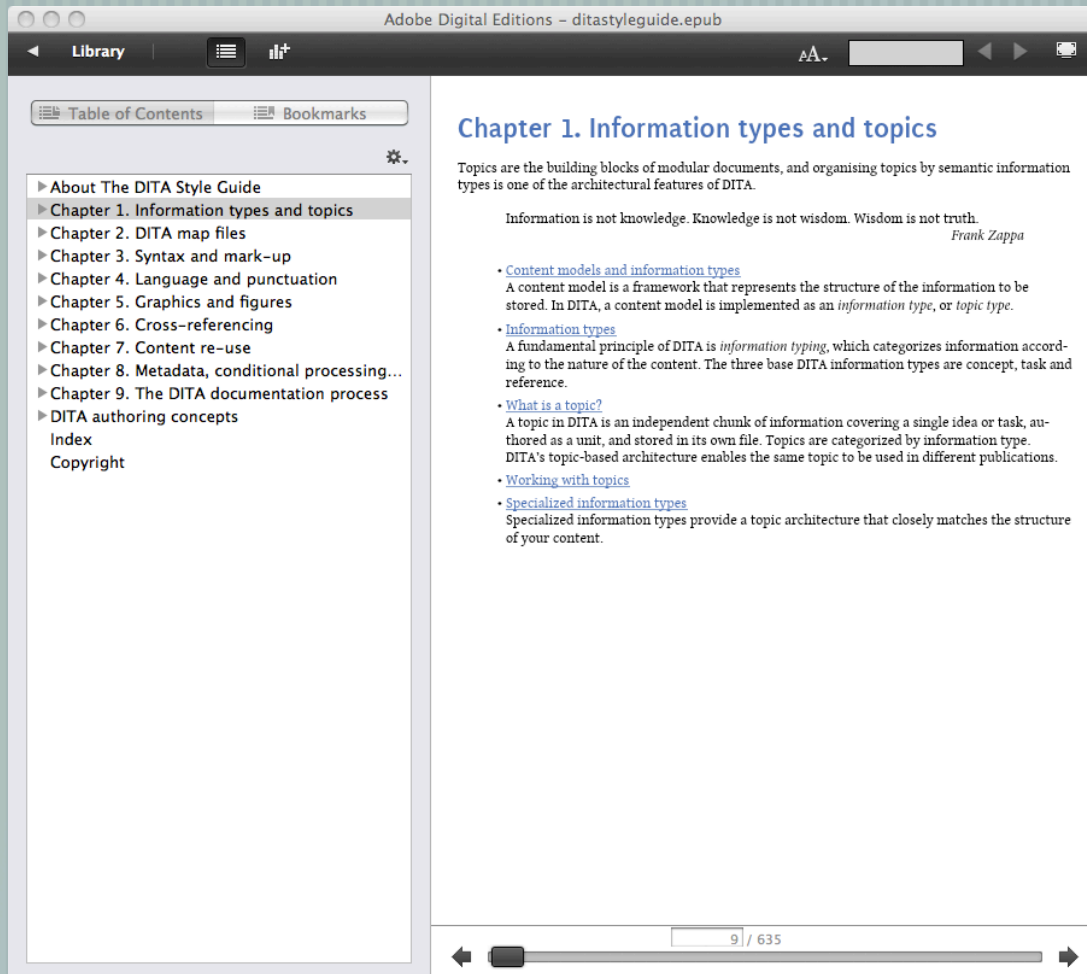> Information is not knowledge. Knowledge is not wisdom. Wisdom is not truth.
>
> *Frank Zappa*

- **Content models and information types**
  A content model is a framework that represents the structure of the information to be stored. In DITA, a content model is implemented as an information type, or topic type.
- **Information types**
  A fundamental principle of DITA is information typing, which categorizes information according to the nature of the content. The three base DITA information types are concept, task and reference.
- **What is a topic?**
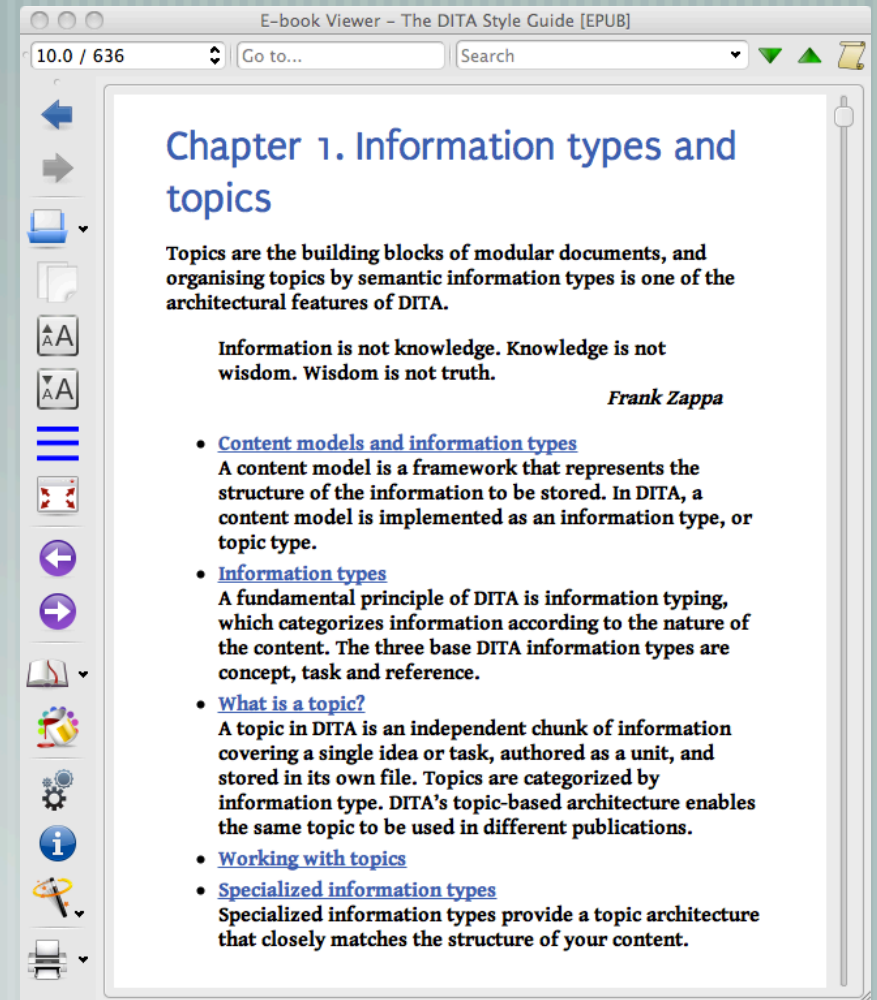 A topic in DITA is an independent chunk of information covering a single idea or task, authored as a unit, and stored in its own file. Topics are categorized by information type. DITA's topic-based architecture enables the same topic to be used in different publications.
- **Working with topics**
- **Specialized information types**
  Specialized information types provide a topic architecture that closely matches the structure of your content.

What about Kindle?

- Kindlegen (from Amazon) converts EPUB to MOBI/KF8

- May require modifications to EPUB for decent Kindle output

- Simple EPUB coding produces the best Kindle output

- KF8 supports some of the EPUB 3 features

- Always test after conversion!

Technologies and tools

- Just a collection of XHTML, CSS, XML and media files
- Techcomm developers will want to single-source from existing content (my assumption)
- Could hand-code .. but don't
- Use a tool (HAT or script) to create the basic files
- Fine tune (manually) as needed

EPUB publishing tools

- Adobe TCS (RoboHelp, FrameMaker)

- ComponentOne Doc-To-Help

- EC Software Help & Manual

- MadCap Flare

- Webworks ePublisher

Publishing tool issues

- Internal EPUB file structure - simpler makes it easy to make manual modifications

- Use of "fake" lists vs. real HTML lists

- Use of @style attributes on elements vs. CSS classes

- Limited metadata

- Control of "chapter" breaks
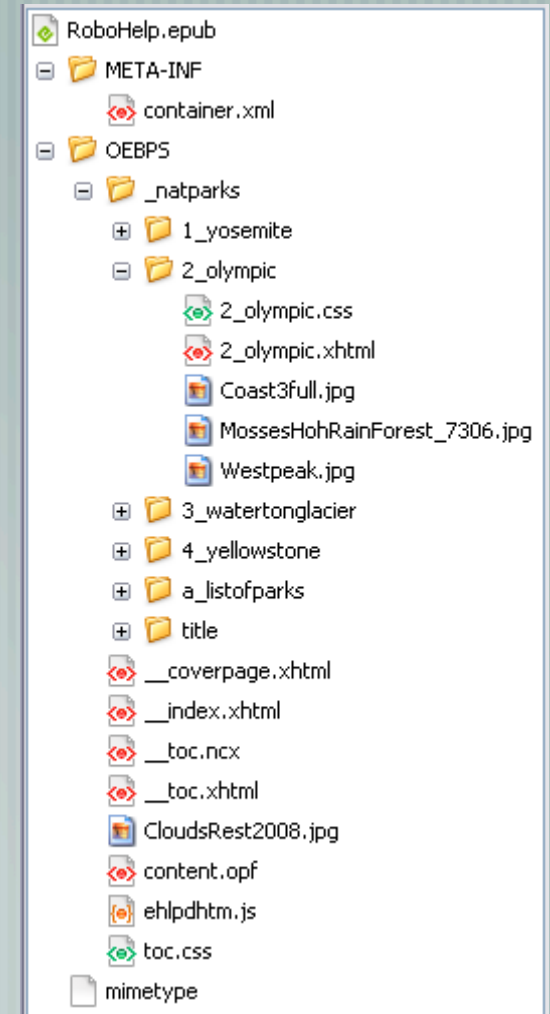
- "EPUB 3" tools really don't offer more than EPUB 2

Adobe TCS

EPUB 3 (pretty good)

Import from FrameMaker, Word, DITA, and Help sources

+ TOC, Index, Glossary, reasonable metadata, cover page

– Awkward file structure, "fake" lists, hard-coded @style attributes

RoboHelp.epub
- META-INF
  - container.xml
- OEBPS
  - _natparks
    - 1_yosemite
    - 2_olympic
      - 2_olympic.css
      - 2_olympic.xhtml
      - Coast3full.jpg
      - MossesHohRainForest_7306.jpg
      - Westpeak.jpg
    - 3_watertonglacier
    - 4_yellowstone
    - a_listofparks
    - title
    - __coverpage.xhtml
    - __index.xhtml
    - __toc.ncx
    - __toc.xhtml
    - CloudsRest2008.jpg
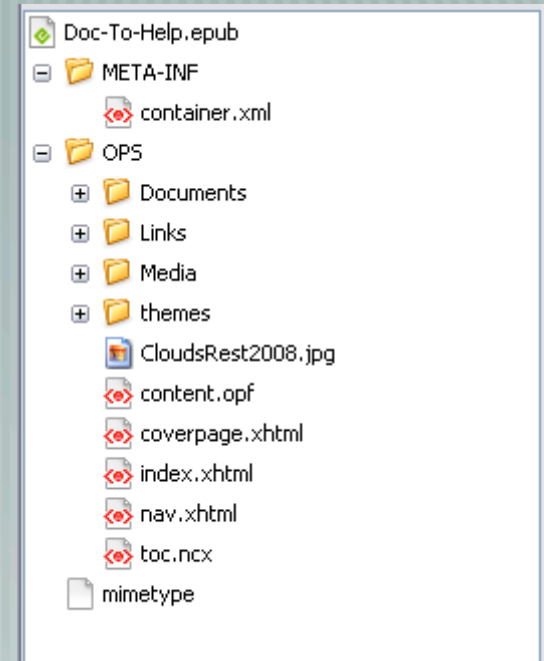   - content.opf
    - ehlpdhtm.js
    - toc.css
  - mimetype

ComponentOne Doc-To-Help

EPUB 3 (decent)

Import from Word and Help sources

+ TOC, Index, Glossary, cover page, reasonable file structure

− "Fake" lists, limited metadata, hard-coded @style attributes, stylesheet problems

```
Doc-To-Help.epub
  META-INF
       container.xml
  OPS
       Documents
       Links
       Media
       themes
       CloudsRest2008.jpg
       content.opf
       coverpage.xhtml
       index.xhtml
       nav.xhtml
       toc.ncx
  mimetype
```
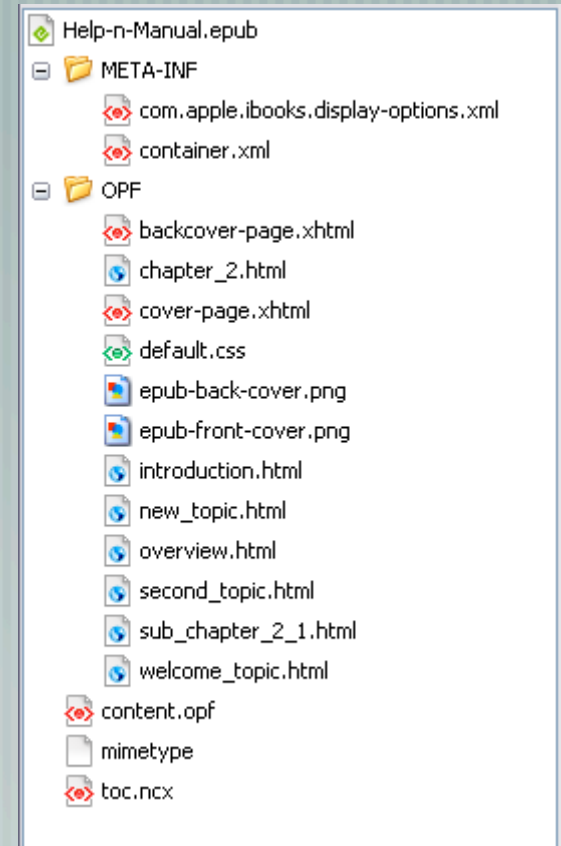
EC Software Help & Manual

EPUB 2

Import from Word and Help sources

+ TOC, good metadata, simple file structure, front (and back!) cover

− No Index support, "fake" lists (option for real lists), no chapter-break control

- Help-n-Manual.epub
  - META-INF
    - com.apple.ibooks.display-options.xml
    - container.xml
  - OPF
    - backcover-page.xhtml
    - chapter_2.html
    - cover-page.xhtml
    - default.css
    - epub-back-cover.png
    - epub-front-cover.png
    - introduction.html
    - new_topic.html
    - overview.html
    - second_topic.html
    - sub_chapter_2_1.html
    - welcome_topic.html
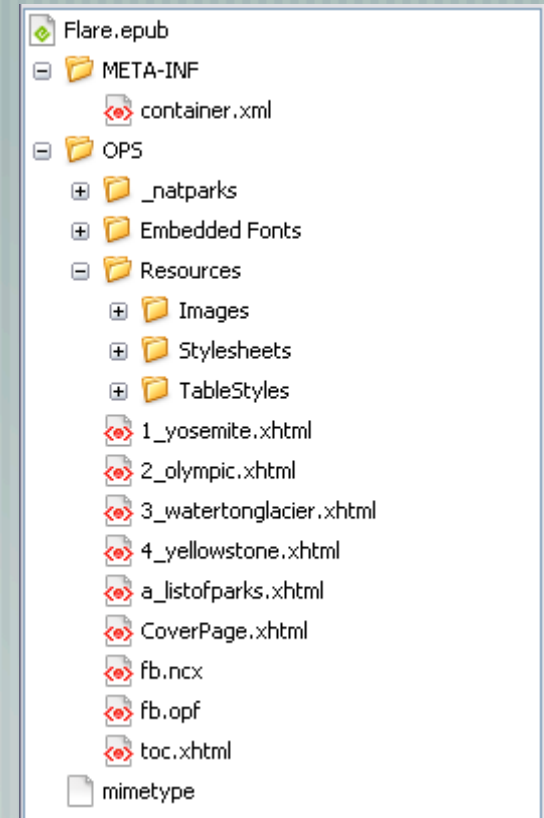 - content.opf
  - mimetype
  - toc.ncx

MadCap Flare

EPUB 3 (getting better)

Import from FrameMaker, Word, DITA, and Help sources

+ TOC, Index, Glossary, cover page, reasonable file structure

– "fake" lists, limited metadata, hard-coded @style attributes, strange problems with TOC and index generation
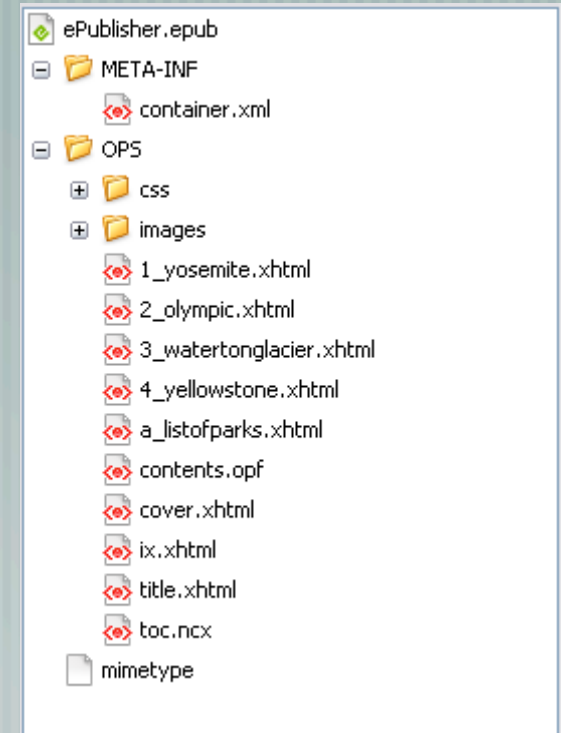
```
Flare.epub
META-INF
    container.xml
OPS
    _natparks
    Embedded Fonts
    Resources
        Images
        Stylesheets
        TableStyles
    1_yosemite.xhtml
    2_olympic.xhtml
    3_watertonglacier.xhtml
    4_yellowstone.xhtml
    a_listofparks.xhtml
    CoverPage.xhtml
    fb.ncx
    fb.opf
    toc.xhtml
mimetype
```

Webworks ePublisher

EPUB 2

Import from FrameMaker, Word, DITA

+ TOC, Index, Glossary, nice file structure, cover page

− Very limited metadata, "fake" lists, hard-coded @style attributes, requires considerable setup

More EPUB publishing tools

- DITA Open Toolkit + DITA for Publishers plugin - EPUB 2

- Docbook - EPUB 2 and 3, various options

- FrameMaker + ElmSoft EPubFm2 (FM 6-10) - EPUB 2

- Author-it + ePUB Publishing Extension - EPUB 2

- Word plugins? (YMMV)

- PDF to EPUB converters

EPUB tools and utilities

- EPUB editor
  - Oxygen XML editor - EPUB 3
  - BlueGriffon EPUB Edition - EPUB 3
  - Sigil - EPUB 2
- Calibre - multi-purpose tool
- epubcheck - validator
- kindlegen - converts EPUB to MOBI/KF8

DEMO!

- Open EPUB in Oxygen XML editor
  - Review general file structure and features
  - Show validation features
- Open EPUB in Sigil or BlueGriffon
- Show desktop readers

Resources

- EPUB 3 specification - idpf.org/epub/30
- EPUB 3 Fixed Layout Format specification - idpf.org/epub/fxl
- CSS 3 Module Info - www.css3.info/modules
- Liz Castro – www.pigsgourdsandwikis.com
- Ebook Ninjas Podcasts - ebookninjas.com
- Resources – www.epubtest.com/resources.php
- Follow #eprdctn on Twitter
- Scott Prentice <scott AT leximation.com> – www.leximation.com