

DITA

Darwin Information Typing Architecture

高昂

中国标准化研究院
高新技术与信息标准化研究所

2010-09

报告的主要内容

- DITA背景概述
- DITA设计原则
- DITA的主要元素
- 主题、映射与样式
- DITA与DocBook对比分析
- DITA工具和使用
- DITA编译过程演示

DITA含义

- DITA（达尔文信息分类体系结构，Darwin Information Typing Architecture）是一种面向主题文档类型定义（Document Type Definitions, DTD）系统，用于定义编写和交付内容信息的规则。DITA通过XML架构来描述、管理和发布内容信息，并用于内容信息的编写、生成和交付。
- DITA目前由结构化信息标准组织OASIS维护，DITA通过将文档组织为可重用的模块（Topics主题）并将模块进行组织来表示信息内容，DITA提供了主题的元数据注释，用户可以便捷的搜索、过滤和处理内容。
- **Darwin:** DITA utilizes principles of inheritance for specialization
- **Information Typing:** DITA was originally designed for technical information based on an information architecture of Concept, Task and Reference
- **Architecture:** DITA is a model for extension both of design and of processes

DITA是什么不是什么

- DITA is an architecture for creating topic-oriented, information-typed content that can be reused and single-sourced in a variety of ways
- It is also an architecture for creating new topic types and describing new information domains based on existing types and domains
- **DITA is not just another tool but an international standard to support structured authoring and reuse in any technical domain**

DITA核心设计原则

■ Topic orientation

- ❖ Discrete units of information covering a specific subject with a specific intent

■ Topic granularity

- ❖ Self-contained topics combine with other topics into information sets

■ Strong typing

- ❖ DTDs and schemas guarantee that DITA types follow identical information structures

■ Specialization

- ❖ Architecture for extending basic types to new types adapted for a particular use within an information set

■ Common base class

- ❖ Top-level "generic" base type provides "fallback" for all

DITA的Topic主题

- DITA 是**基于主题**的，因为主题是允许以不同单元**交付可重用上下文**而不影响作者效率的**最理想尺寸**。如果选择小一些的单元，作者需要核对所有上下文的单元，以确保信息流正确。如果选择大一些的单元，信息不容易重新合并到不同交付上下文所需的结构。
- 从Topic的粒度划分来看，Topic要大到能**自我容纳**，同时要小到特殊上下文需要的以任何高级结构的**有效重用**。

- <topic> 是单一**非嵌套主体**和任何数量的**嵌套主题**的容器。

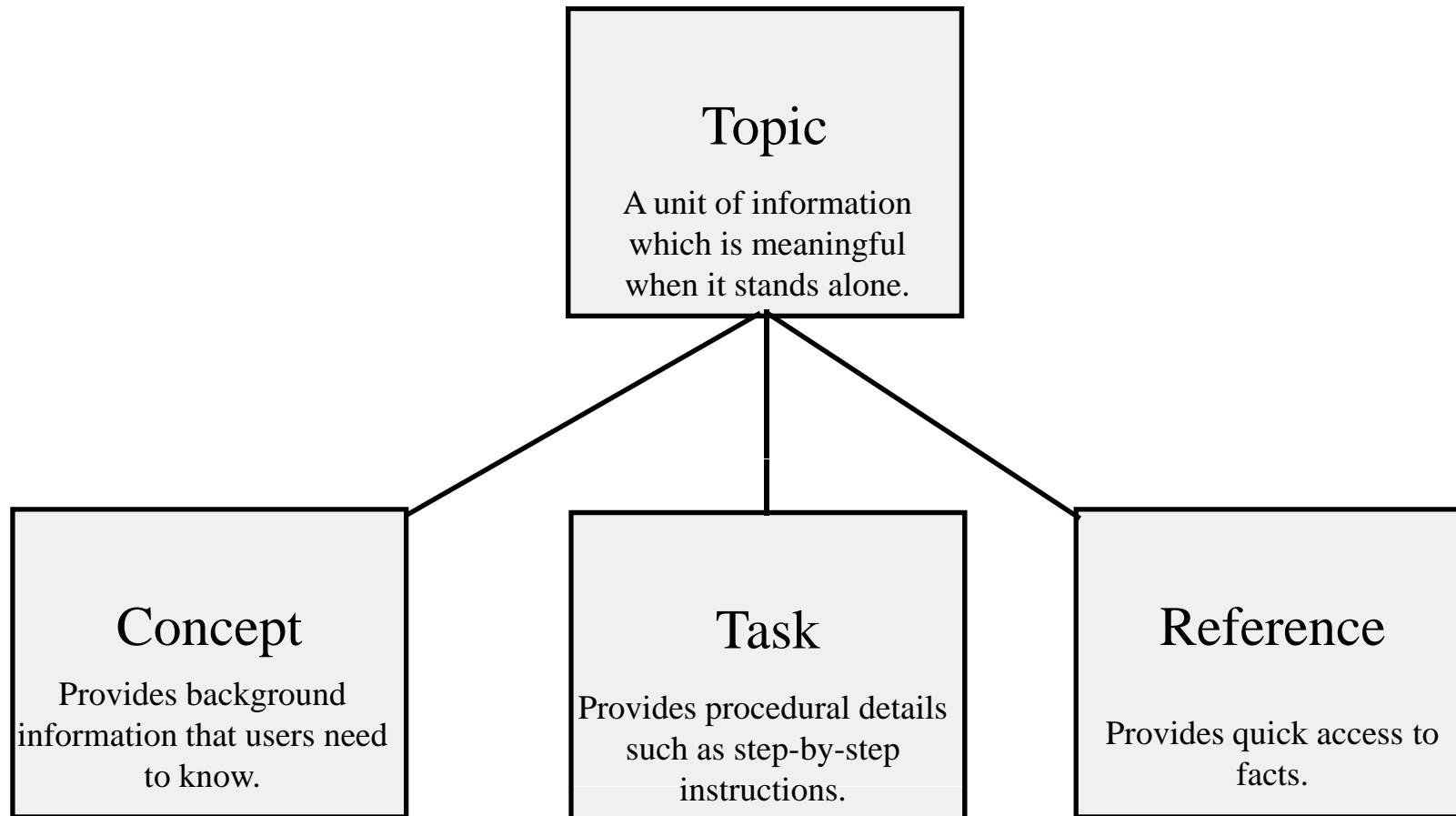
<title> 提供自我描述，与创作内容的标题一致。

<body> 是**段落级内容**和任何数量的**非嵌套部分**的容器。

主题可以增加一个前言，其中包括简短描述和其他可选的元数据。

DITA的Topic主题 (Cont.)

- Topic是按单一题目范围组织的一段信息。从结构上看，它是一个后面跟随着文本和图象的主题，可选的按节组织方式。主题有许多不同类型，最通用的是概念、任务和引用。



DITA主题 (TOPIC) 示例

```
<task id="installstorage">
  <title>Installing a hard drive</title>
  <shortdesc>You open the box and insert the drive.</shortdesc>
  <prolog><metadata>
    <audience type="administrator"/>
    <keywords>
      <indexterm>hard drive</indexterm>
      <indexterm>disk drive</indexterm>
    </keywords>
  </metadata></prolog>
  <taskbody>
    <steps>
      <step><cmd>Unscrew the cover.</cmd>
        <stepresult>The drive bay is exposed.</stepresult>
      </step>
      <step><cmd>Insert the drive into the drive bay.</cmd>
        <info>If you feel resistance, try another angle.</info>
      </step>
    </steps>
  </taskbody>
  <related-links>
    <link href="formatstorage.dita"/>
    <link href="installmemory.dita"/>
  </related-links>
</task>
```

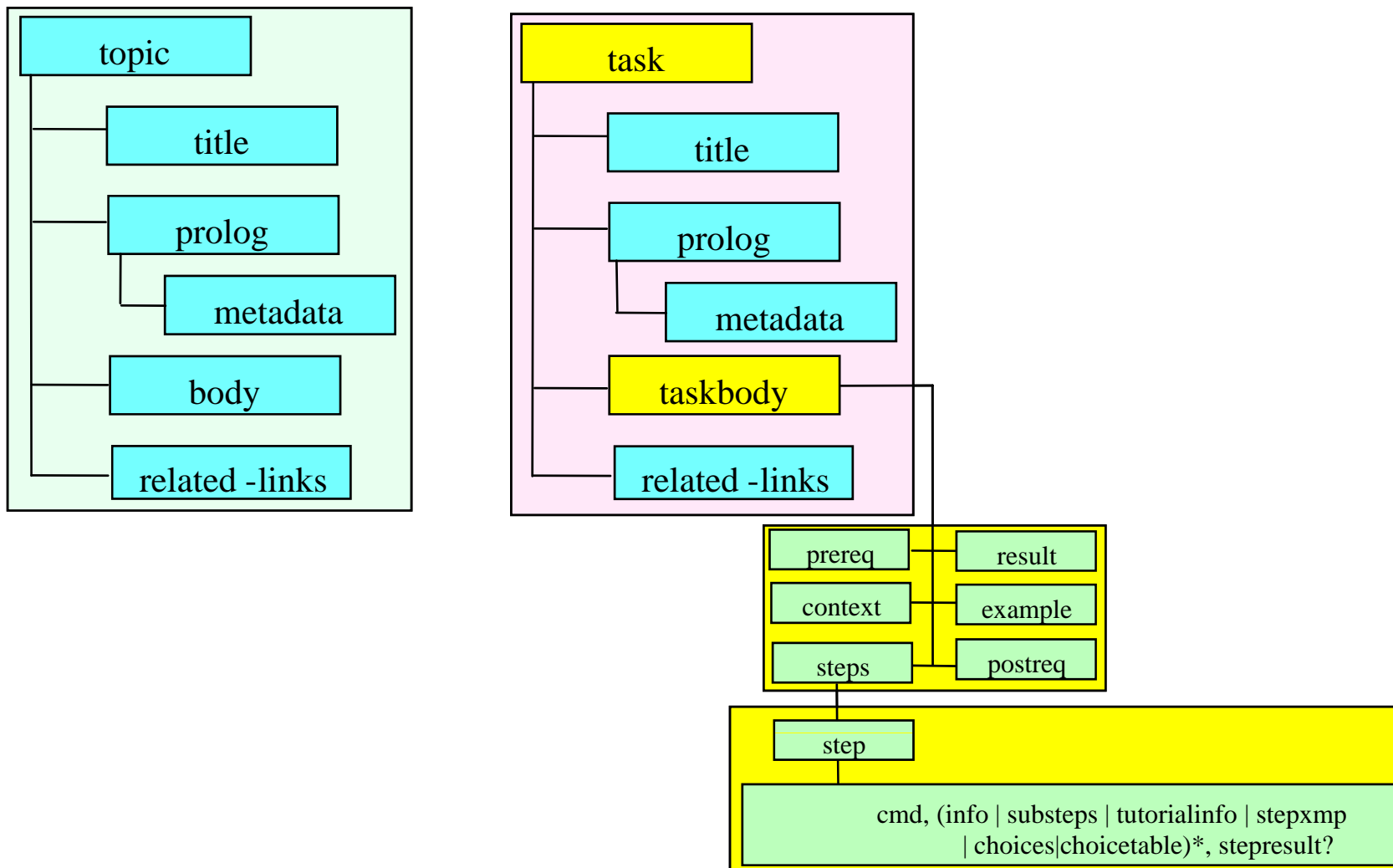
Identifier, title,
and shortdesc

Properties of
the topic

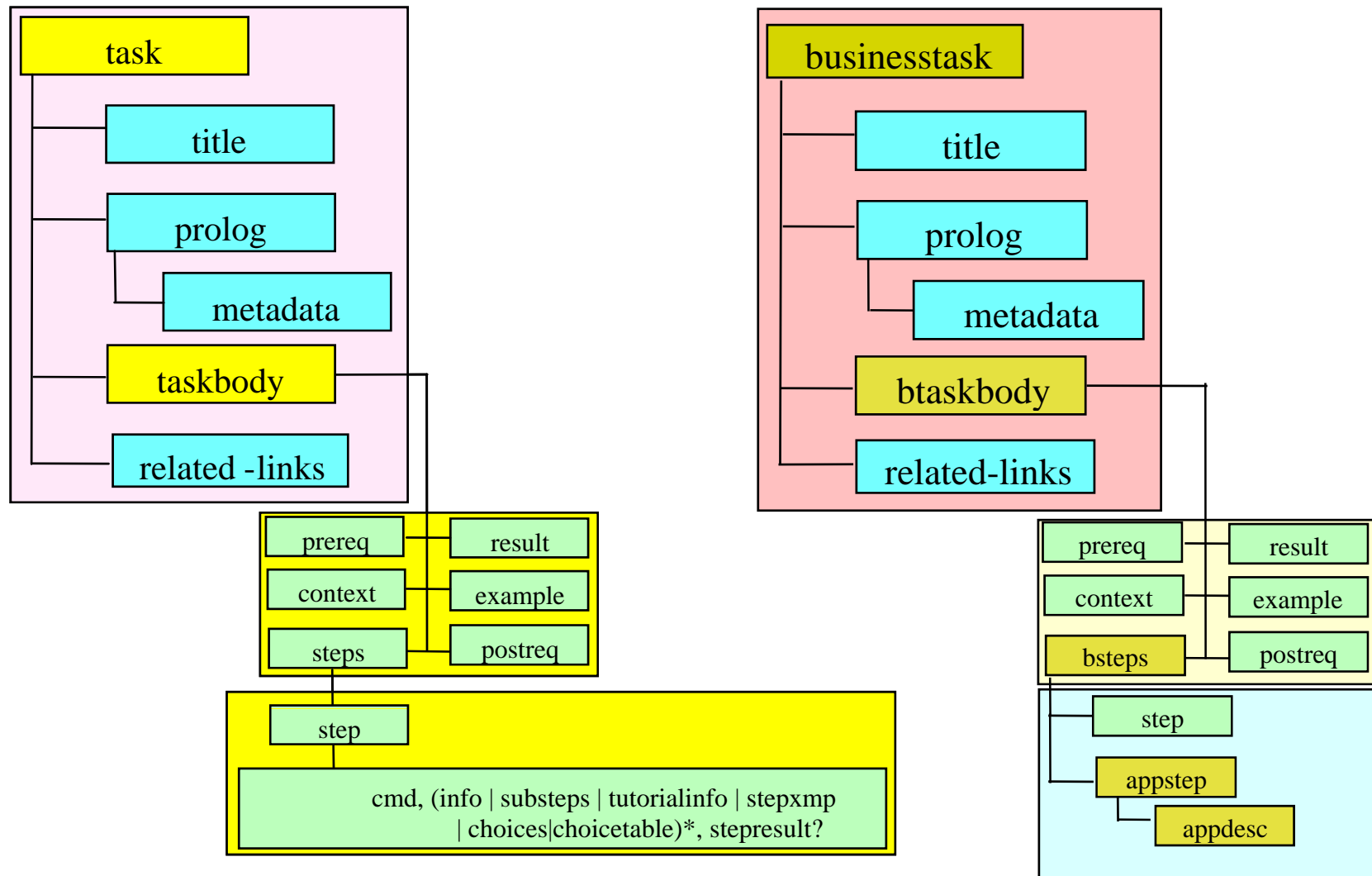
Type-specific
content body

Relationships
to other topics

从主题 (Topic) 到任务 (Task) 的细化

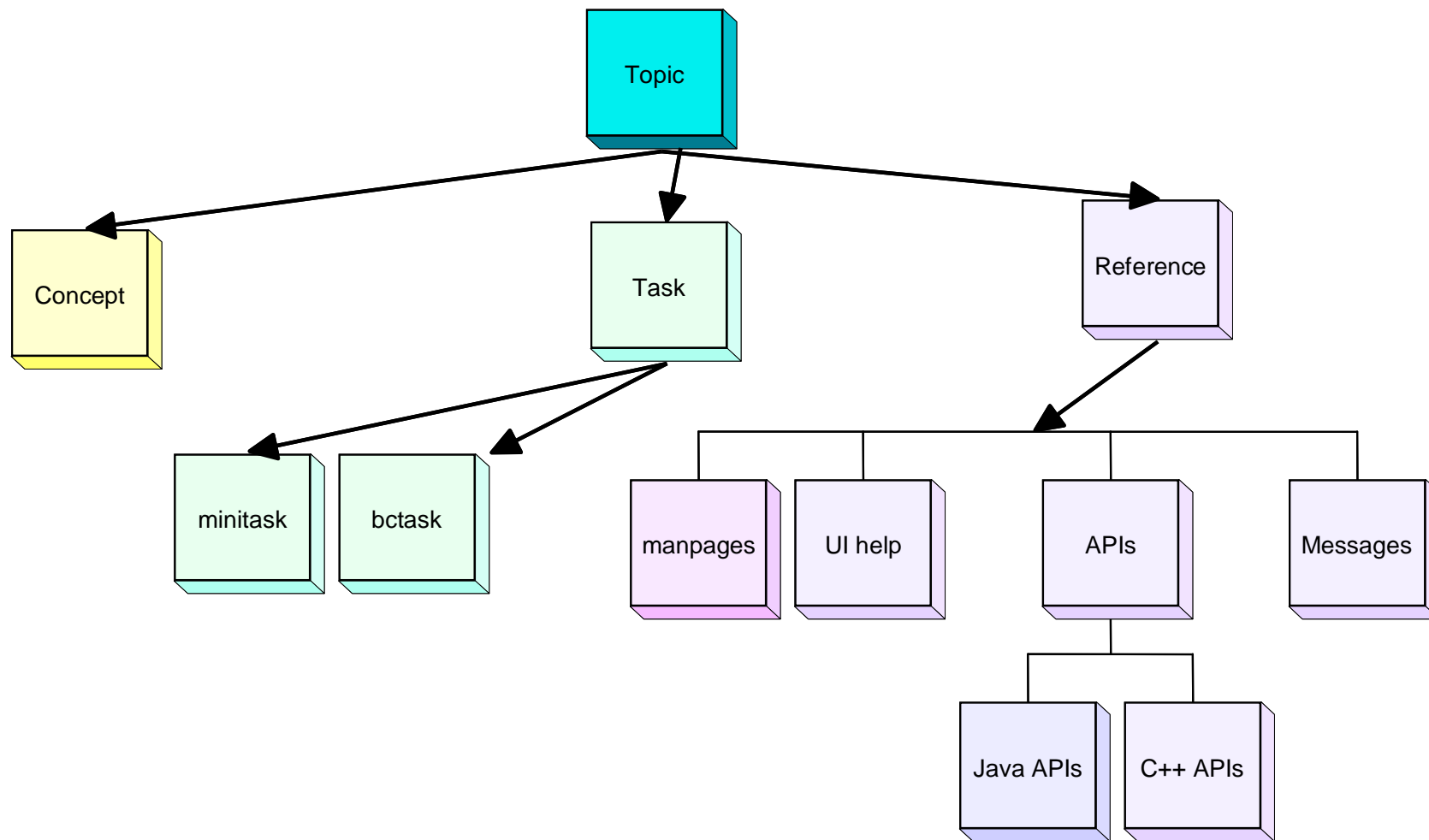


从任务(Task)到业务(Business Task)细化



主题的结构细化

- 以Topic为核心进行内容组织



映射（Map）示例

```
<map title="Tasks">
  <topichead navtitle="Installing"
    audience="admin">
    <topicmeta>
      <shortdesc>Install products before
        configuring or using them.</shortdesc>
    <topicmeta>
    <topicref href="installstorage.dita">
      <topicref href="unscrewcover.dita"/>
      <topicref href="insertdrive.dita"/>
      <topicref href="replacecover.dita"/>
    </topicref>
    <topicref href="installwebserver.dita">
      <topicref href="closeprograms.dita"/>
      <topicref href="runsetup.dita"/>
      <topicref href="restart.dita"/>
    </topicref>
    <topicref href="installdb.dita">
      <topicref href="closeprograms.dita"/>
      <topicref href="runsetup.dita"/>
      <topicref href="restart.dita"/>
    </topicref>
  </topichead>
  ...
</map>
```

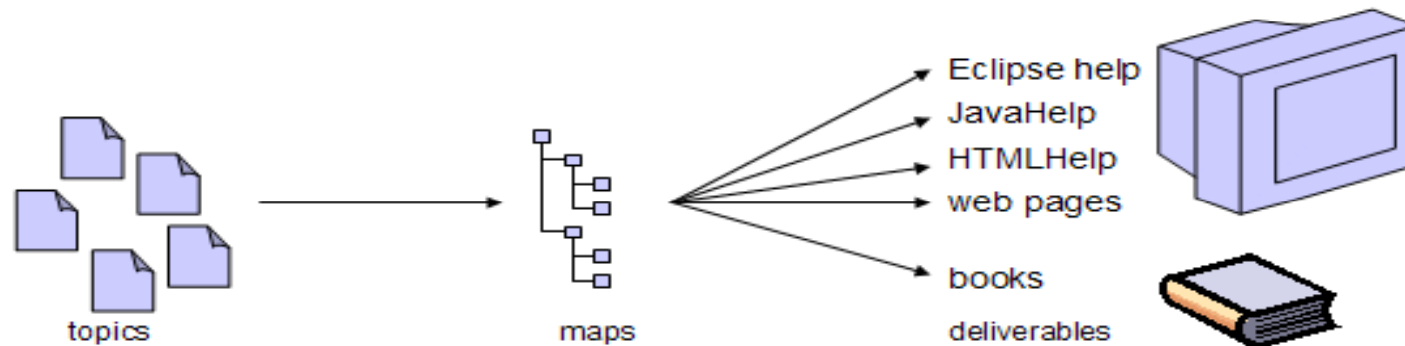
A heading doesn't have to have a topic

Title and properties can be assigned in the map

The map organizes a set of topics in a hierarchy

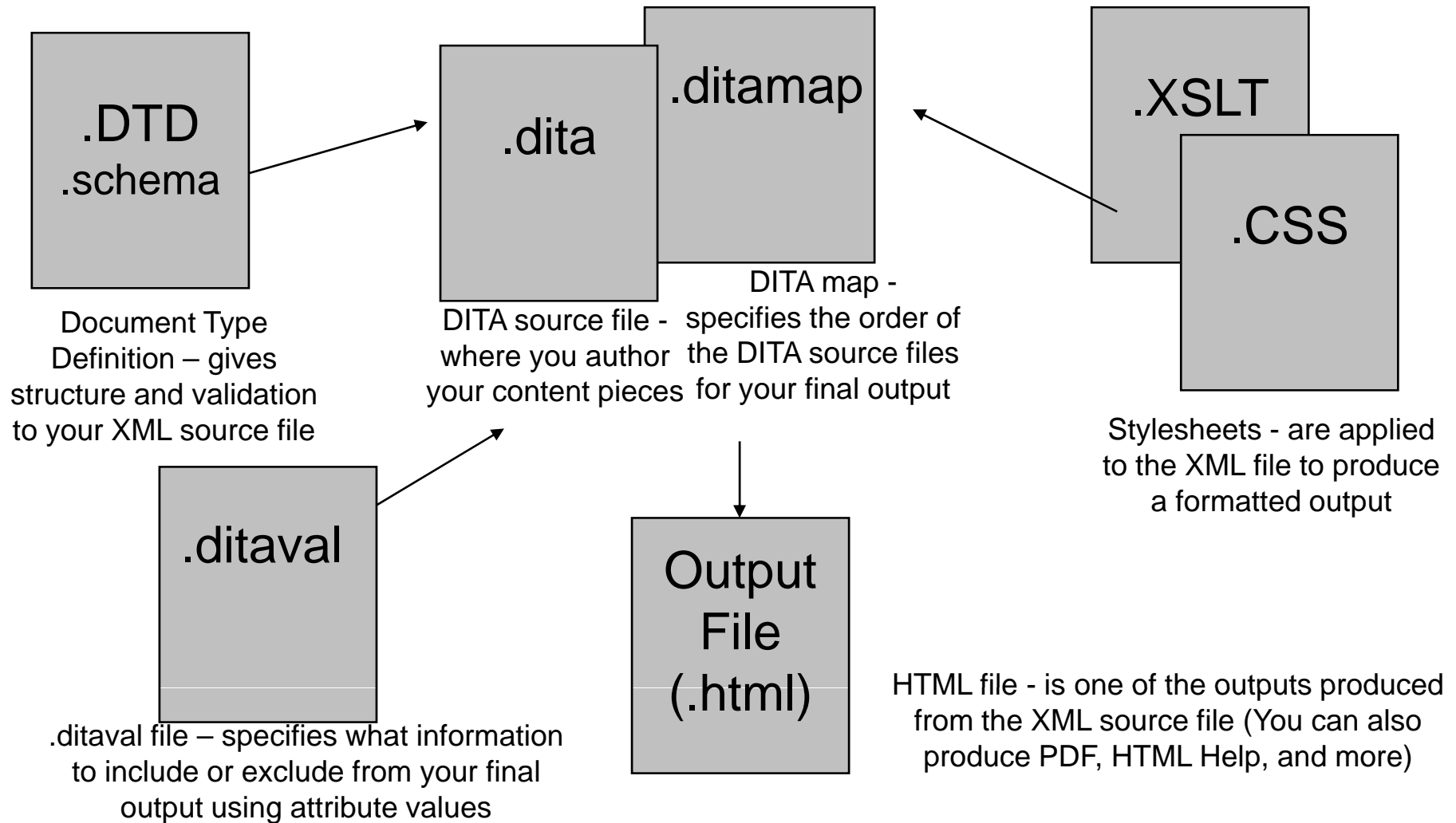
A topic can appear multiple times in the hierarchy

DITA映射方式



- A DITA map applies **context** to the **topics**
- Organizes **a set of topics** in a **hierarchy** and **sequence**
- Can **reuse** the **same topic** with different collections of topics
- Sets **properties** of the topic at a position within the hierarchy

DITA处理过程



DITA与DocBook标准特性对比

DocBook		DITA
DocBook is a document type definition (DTD) for writing technical books and articles, with the structure that such forms imply.	Definitions	DITA is both a DTD and a set of rules for writing online contextual documentation such as software help files.
DocBook is hierarchical by nature, and must be developed to allow for single-sourced content DocBook has a fixed element and attribute set.	Concepts	DITA separates content from context, allowing multiple architectures of information DITA is extensible, allowing the definition of information types
Book (section) oriented	Context	Topic oriented
Large set of elements and attributes targeted for technical documentation	Key Features	Topic based, separation of content and context, specialization
Simple. Can usually be downloaded and set up in a day.	Download and Installation	More complex. Usually requires more time and some technical expertise.
Regular XML file called an include file with all files included in the document	Document File	ditamap file with all files and each file's related links
PDF HTML HTMLHelp	Output	Unrestricted, but all need development
Well supported by the XML community	Community	Small but strong, growing community
With a little development, DocBook can do what DITA is attempting on a smaller scale	Complexity	DITA is impressively engineered, but can be too industrial-strength

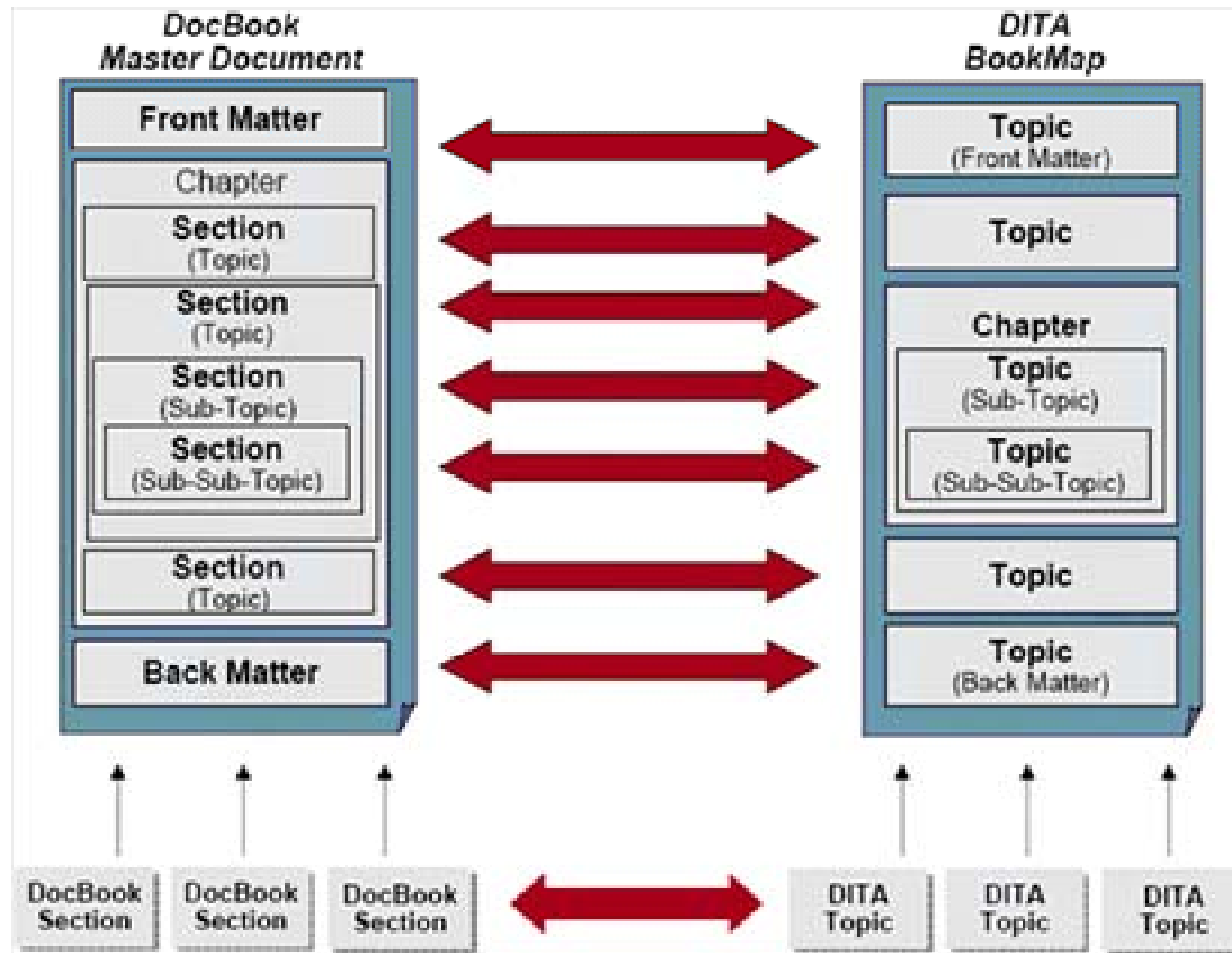
DITA与DocBook对比

- DITA和DocBook通过定义规范化的文档描述规则，来解决文档交付过程中遇到的问题。面向不同类型的交付出版物，DITA和DocBook各有见长，但在实际应用中也有自身的限制因素。
- DITA解决了出版物的结构化描述和内容重组问题，且支持多语言版本制作，适用于对格式有严格限定的技术手册类出版物。但DITA不能实现很完美的样式渲染，且对于内容与格式一体化的复杂出版物，DITA很难进行主题和界定与划分。所以使用DITA进行书籍出版的成本和难度较高。
- 相比较来说，DocBook适用于通用出版物，文档易于组织和排版。但DocBook内容以Section段落组织，不具备DITA的内容映射机制，无法做到类似Topic这样粒度的内容划分与重组。且对于内容需要频繁修改的文档排版，Docbook略显力不从心。

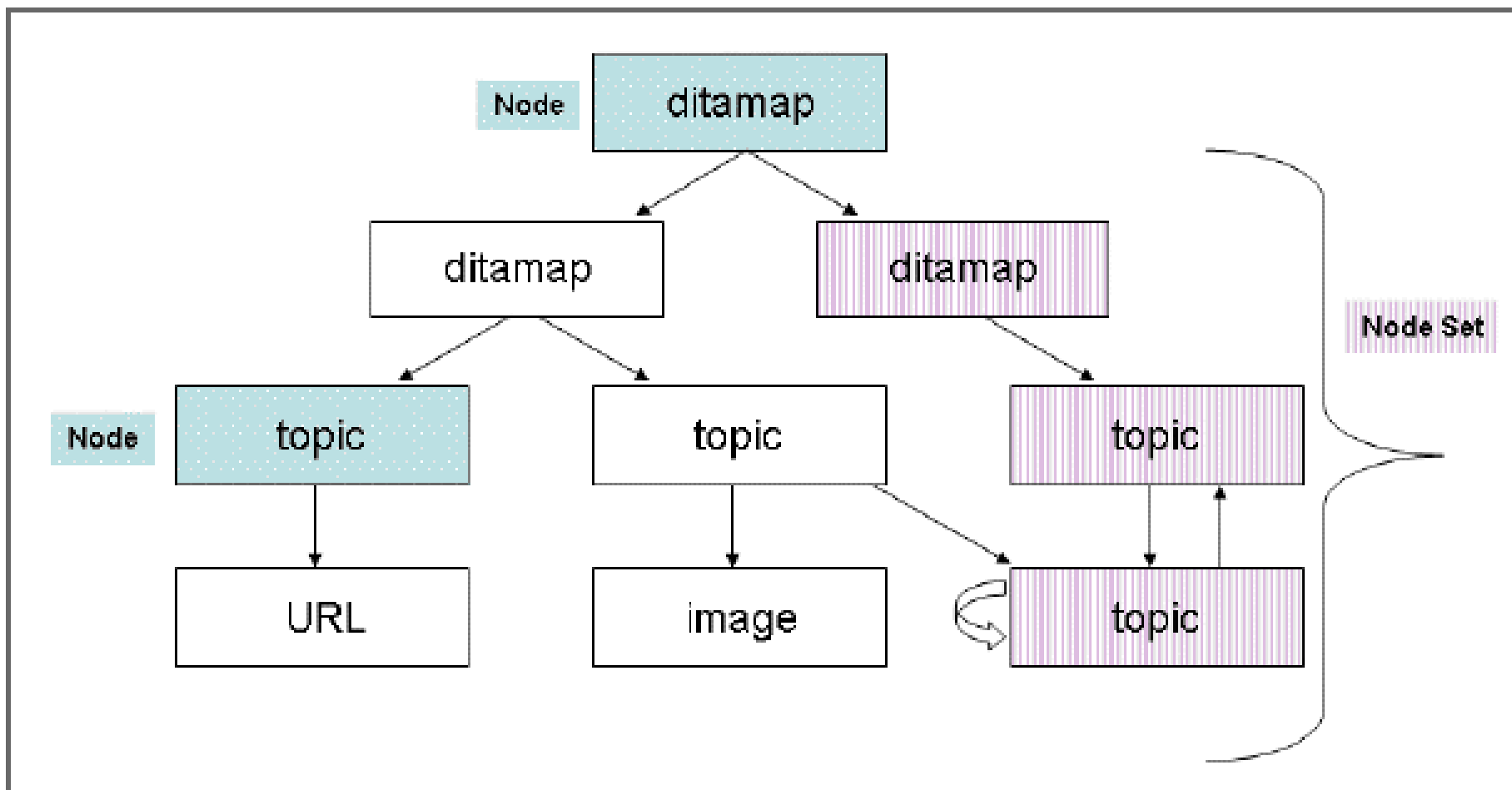
DITA与DocBook对比 (Cont.)

- DITA和DocBook专注于交付技术信息，但DITA侧重于交付主题，而DocBook侧重于交付书籍。DITA提供基于主题级粒度的信息分类，允许作者组织并描述特定信息领域。在生成多种文档格式的信息重用过程中，能够保持内容的高度一致性。在最终交付物的输出格式方面，DITA能够生成PDF、CHM、HTML等大部分的出版交付类型。DocBook常用的交付格式为PDF和HTML，其他输出格式需要借助相关的功能插件。
- 在学习应用方面，DITA包含众多语法和标签定义，在应用过程中的学习曲线较陡，在文档生成过程中一般需要使用Ant命令进行编译。而DocBook相对容易理解和使用，一般使用者从了解到上手使用DocBook制作文档仅用一天时间即可。希望了解更多DITA与DocBook的对比分析的朋友，可以进一步阅读DCL数据转换实验室包含更为详尽内容的文章。

Docbook与DITA结构对比
















标准 DITA 项目结构中的树形结构



DITA开放工具箱

- DITA开放工具箱是OASIS的DITA标准的DTD和Schema的Java实现，能够将DITA文档的Maps和Topics转换为最终的交付文档（HTML、PDF、RTF）。DITA开放工具箱包含Ant构建模板、DITA词汇表DTD定义，DITA词汇表的XML Schema定义以及Java文档处理类库等内容。对于DITA文档，源文件和映射文件中的标记项由DTD和Schema定义，借助XSLT和CSS文档，经处理渲染后得到最终交互文档。
- 在商业产品方面，提供软件产品生命周期管理的PTC公司开发的XML文档处理器Arbortext，支持DITA的编辑与转换工作。Syntext公司开发的开源的XML编辑器Serna Free支持DITA、Docbook、XHTML等XML文件类型，能够实现可视化的编辑工作。

DITA开放工具箱

	Contents or purpose
 DITA-OT	
 ant	– Templates for defining what to build
 css	– Styles for displaying XML source files in editors or browsers
 demo	– Experiments or demonstrations of DITA capabilities
 doc	– Documentation for DITA or the Open Toolkit
 dtd	– Document type definitions for the DITA vocabulary
 lib	– Java implementation of advanced processing
 out	– Formatted output (created during processing)
 resource	– Styles and other resources for XHTML and other output
 samples	– Example DITA content for exercising the processing
 schema	– XML Schema definitions for the DITA vocabulary
 temp	– Intermediate files (created during processing)
 xsl	– XSLT files for XHTML and PDF processing

DITA-OT的条件编译

在DITA-OT进行DITA文档编译时，可以让DITA-OT在编译时进行有条件的编译处理，针对不同类型的需求生成内容可定制文档。

<p audience="config"> 设置配置信息如下

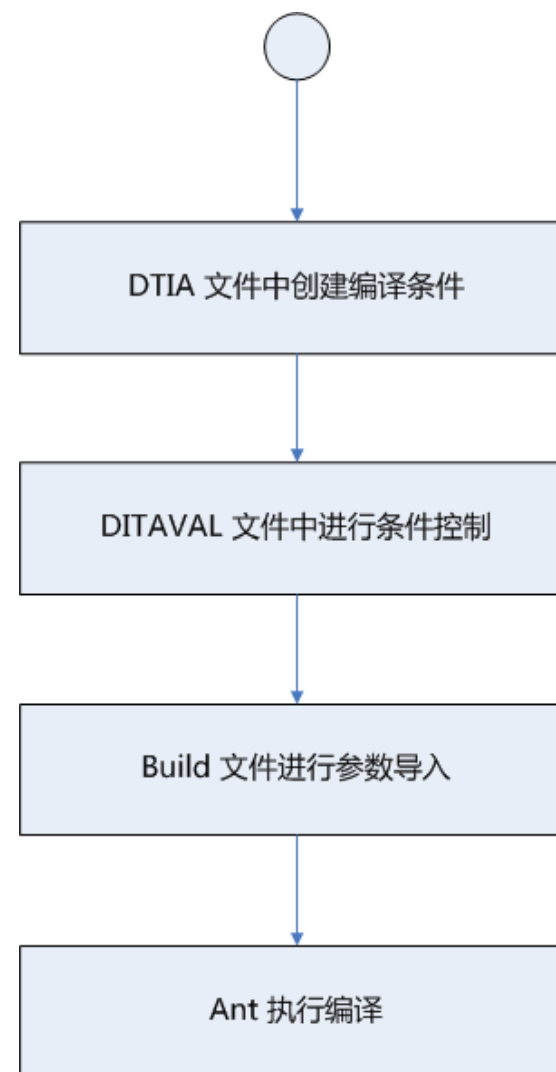
<li audience="admin programmer"> 程序管理员专用信息，这里的内容编译生成成为项目经理提供的文档。

<li audience="programmer"> 这里的内容编译生成成为程序设计人员提供的文档。

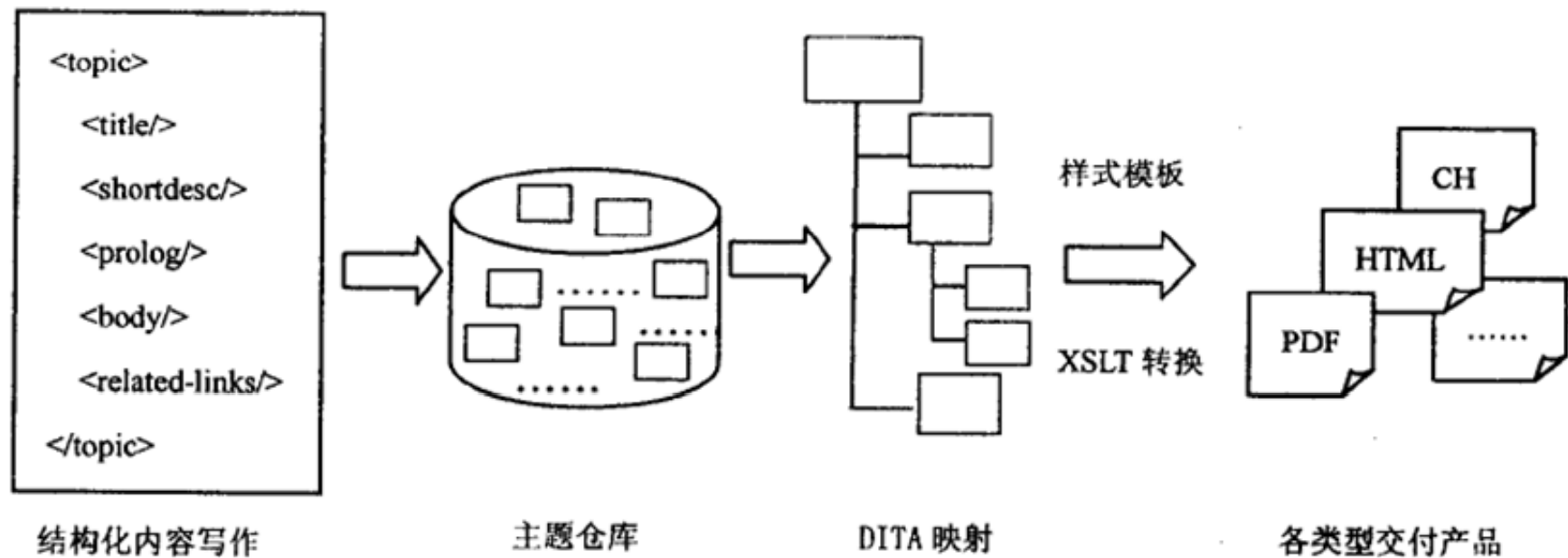
<li audience="programmer" platform="unix"> 这里的内容为在UNIX系统环境下工作的程序员生成。

<li platform="unix"> 这里的内容只显示给UNIX 平台。

</p>

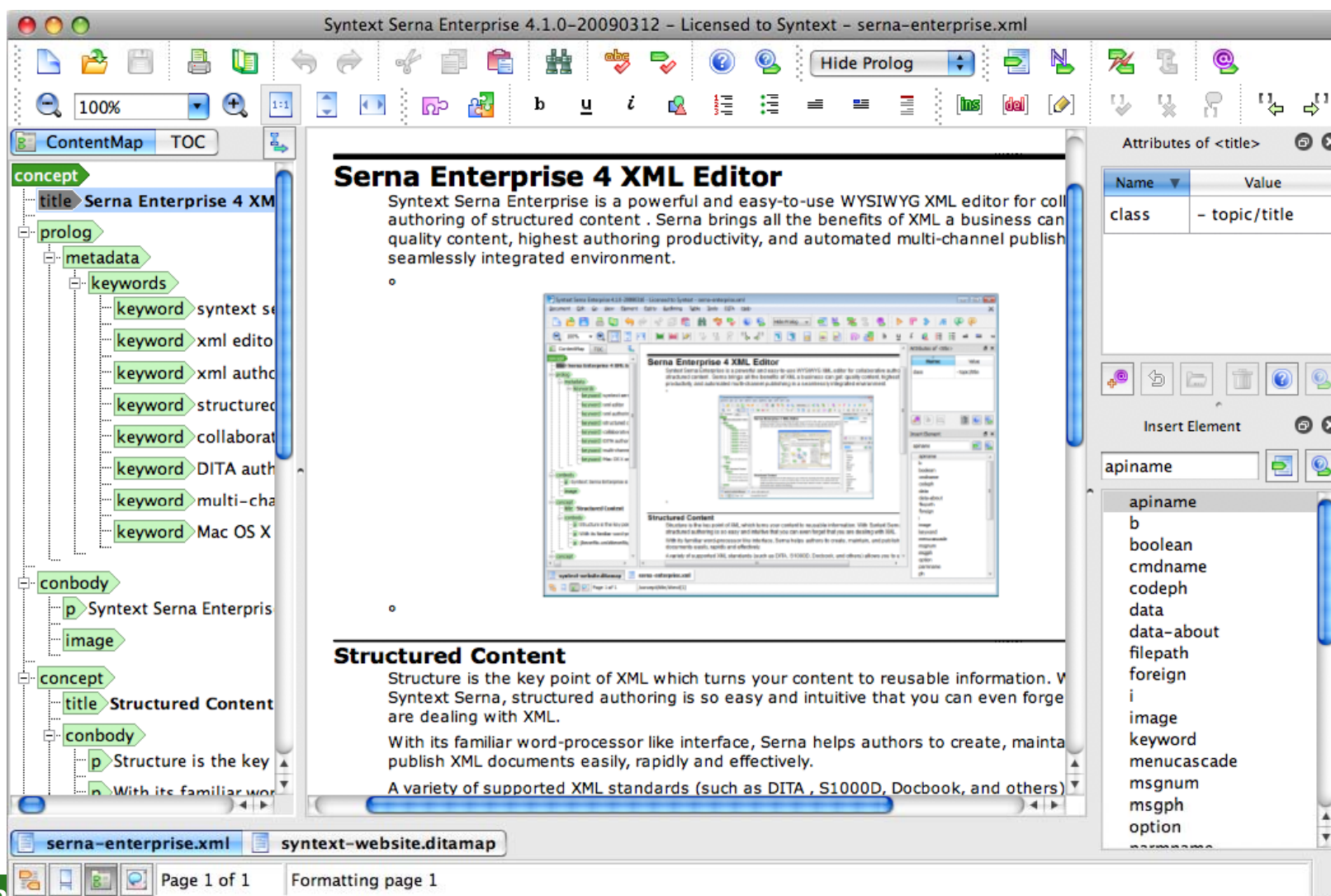


DITA的生产交付



DITA编辑器

- [Serna Free](#)是开源的、所见即所得的XML编辑器，支持DITA、Docbook、XHTML、等XML文件类型。



DITA编辑器

PTC Arbortext Solution for DITA

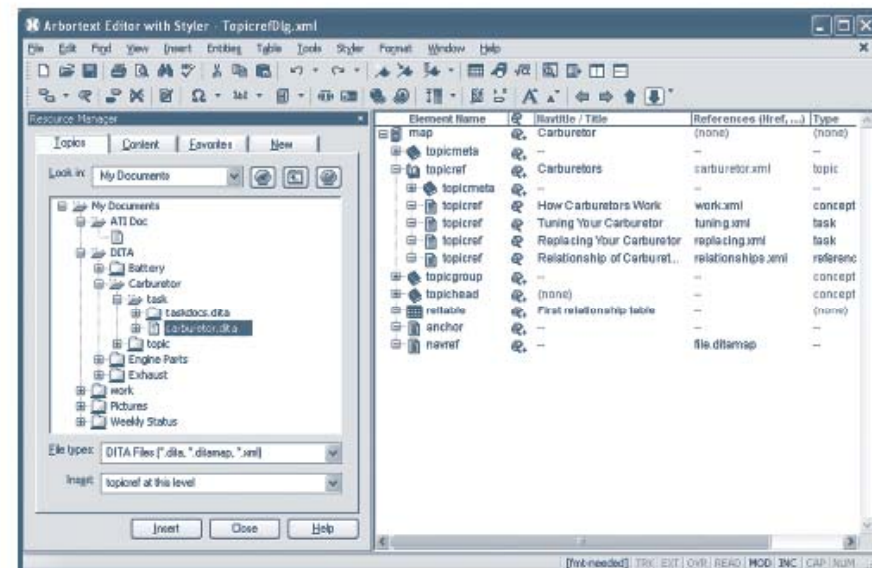
The Only Complete, Fully Integrated Solution for DITA



DITA is one of the most important innovations in XML publishing in recent memory. To achieve best practices in XML for publishing technical documentation, you should take a closer look at our complete DITA solution.

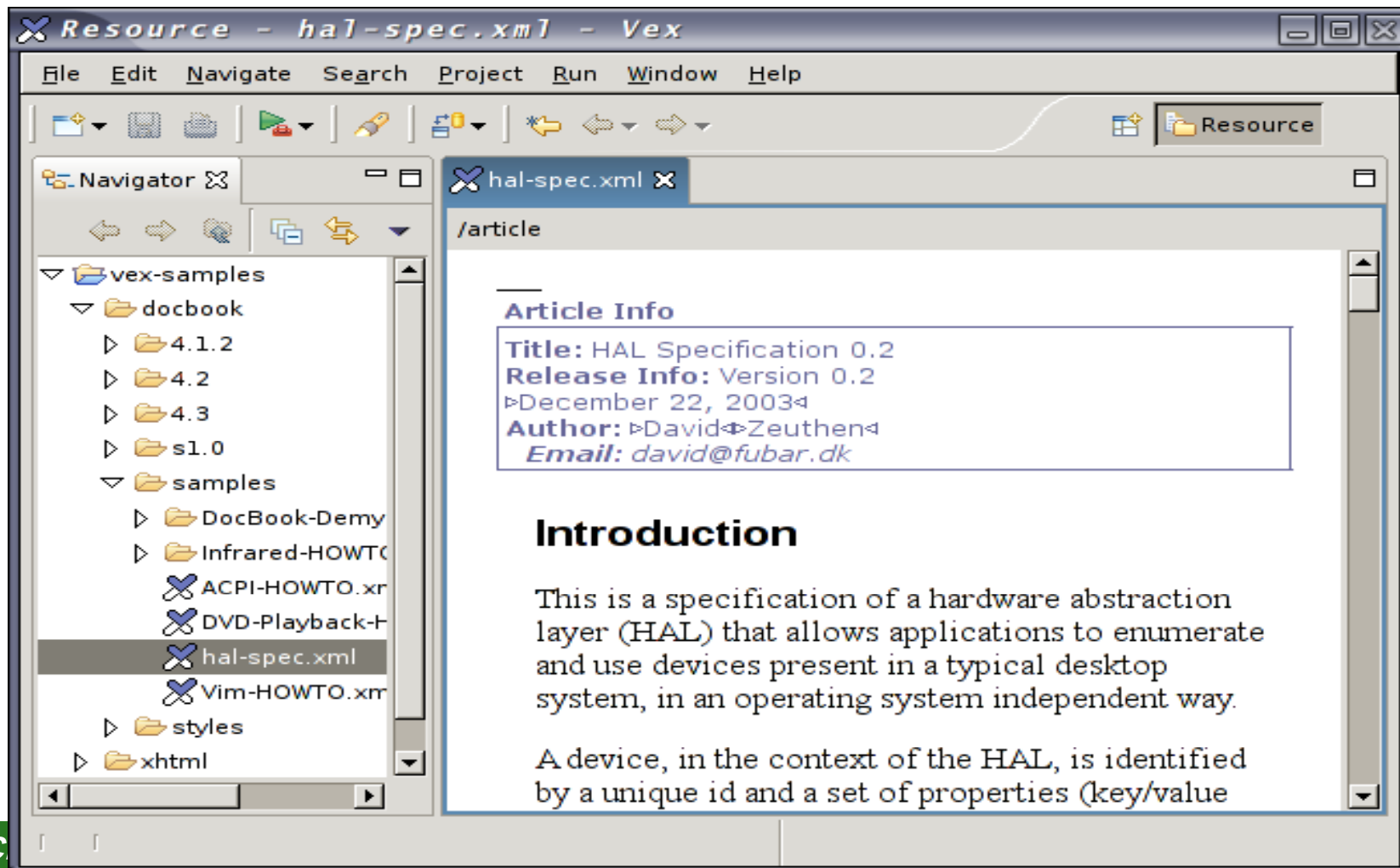
The promise of XML has always been to provide you with more productive authors, better quality information, and publishing processes that are not only faster but also less expensive. PTC Arbortext software has helped many companies achieve those benefits. But now DITA makes it easier than ever before.

What Makes DITA Different





- Vex is a Visual Editor for XML that hides the raw XML tags from the user. Vex uses standard DTD files to define document types and CSS stylesheets to define document layout. Vex contains definitions for DocBook and DITA.



DITA

Darwin Information Typing Architecture



Online community for the Darwin Information Typing Architecture OASIS Standard

Thank You !