

基于“Topic”的语义信息组织技术研究 ——以DITA和Topic Maps为中心*

□ 李颖 张秀梅 / 中国科学技术信息研究所 北京 100038

摘要: 基于“Topic”的语义信息组织与图书馆学的经典Subject理论密不可分,可将其视为较为抽象化的Subject理论在语义Web环境下的一种“演化”和实用化的主题技术。文章首先解读Subject和Topic的概念,并概述两种基于Topic的国际标准技术,即Topic Maps和DITA;其次,简要描述Topic Maps的技术和应用,并重点介绍在图书馆界鲜为人知的DITA技术及其应用;再次,对基于“Topic”的语义信息组织方法予以总结。

关键词: Topic, Subject, 主题, DITA, 主题图, 语义信息组织

DOI: 10.3772/j.issn.1673-2286.2012.04.002

1 关于主题

Topic和Subject都是主题。在图书馆情报领域,没有人不知道Subject词表、Subject编目和索引……最近,人们开始慢慢熟悉Topic Maps(主题图)、Topic模型等术语。而DITA为何?它与Topic有何种连带关系?恐怕目前图书馆界知道的人还是凤毛麟角。或许构建本体、发布关联数据等大潮,让我们无暇再顾及其他?不论国内图书馆是否关注Topic,作为较为抽象的概念化Subject理论,作为一种语义Web环境下的“演化”及实用化的主题技术,基于Topic的语义知识组

织技术在国外已被广泛地应用于各行各业,并在实际应用系统中发挥着不可替代的作用。我们认为,Topic技术不仅是Subject理论的演化和继承,其重要意义是将图书馆经典理论中最为核心的概念体系扩展向网络世界,通过与网络世界其他技术的碰撞与融合,使得Subject概念更加丰富。本文通过分析介绍Topic Maps和DITA两种基于Topic的技术,将呈现给读者主题理论的一种“最佳实践”的应用技术。

Topic与Subject为近义词,与上百年的Subject方法的研究历史相比,Topic技术发展不过十几年的历程。所以,为了更好地理解Topic,

需要对Subject理论作一回顾与梳理。特别指出,从已发布的具有里程碑特性的IFLA最新版“主题规范控制数据功能需求的概念模型”(FRSAD)来看,我们必须对主题概念进行高度的认识和理解,从而构建较为理想的、具有知识化服务功能的数字图书馆系统。

2 “Subject”与“Topic”

2.1 关于Subject

Subject是图书馆文献信息理论的基石,有关Subject究竟为何?如何决定文献的主题?历经百年以上

* 基金支持:中国科学技术信息研究所“汉语科技词系统建设与应用工程”重点工作和国家科技支撑计划“面向外科技文献信息的超级科技词表和本体建设”子任务(编号:2011BAH10B01)支持。

的发展, 图书馆员在实践中应用主题理论的同时, 也从未终止过争论, 这不断地促进了主题方法论的完善。维基百科给出的解释如下^[1]:

Subject用于文献的分类和检索, 满足用户的信息需求。为使文献可被发现, 将Subject标签赋予文献。针对应当发现什么, 归根结底是认识论的问题。为此, 出于不同的观点, Subject有着不同的诠释。

■ Charles Ammi Cutter的观点

Cutter理论认为, 主题的稳定性和稳定性取决于某一社会过程, 主题的含义固化在一个名字或名称中。Cutter关于主题的早期观点, 主宰了20世纪, 也反映在ISO标准中。某种程度上, 主题在社会的形成过程中形成。可以看出, Cutter理论并不特别详细具体或清晰, 人们只得到了主题的社会特征的模糊构想。

■ S. R. Ranganathan的观点

Ranganathan在其冒号分类法中, 给出了一个明确的主题定义: 主题是ideas的组织化产物。Ranganathan的学生指出, 主题是ideas的组织化和体系化产物。除了强调主题的组合、组织化和系统化外, Ranganathan的主题概念强调实际的需求, 从吻合普通人能力和专业的角度来确定主题, 强调主题描述要满足精度和召回率的要求。

■ “面向内容”对“面向需求”的观点

“面向需求的标引是基于对用户需求的预测, 用户需求决定如何标引文献, 标引者关注的是用哪个叙词可使某一主题被发现”, “考虑所有可能的查询并决定哪一个是相关叙词”。而主题是在文档内容中(客观)还是在用户的心智(主观)中, 这是哲学范畴面向主题的问题, 没有定论。

■ 主题知识的观点

Rowley和Hartley认为, 为实现一致性标引, 必须考虑主题结构和文档对特定领域知识发展的贡献特性。不能不遗憾地说, 如何基于主题的概念模式构建知识系统的研究并不深入。反过来讲, 图书馆的主题理论需要在知识信息时代下, 进行扩展应用和完善。

■ ISO观点

ISO在其Topic Maps标准(ISO 13250)中有关Subject的定义, 强调其“断言”(asserted)特性, 以利于主题图标准在概念的闭环中有效地发挥作用。然而, 在一个较为宽泛的语境中, 这一定义效果不佳, 因为它不包含如何确定某一文档或语段的Subject标识或符号的规范。这样一来, 不同的Subject分析方法会意味着不同的结果, 哪一结果可以反映真正的主题? 对同一文献, 不同的人会有不

同的观点。如何能对Subject进行理论性的理解将有助于决定主题分析的原则。

■ IFLA的观点^[2,3]

在2010年6月IFLA审定的FRSAD (Functional Requirements for Subject Authority Data – A Conceptual Model, 主题规范控制数据功能需求的概念模型) 版本中, 曾蕾(Marcia Zeng)教授给出了主题的最新诠释。FRSAD描述了人类认识世界的基本框架, 已经超越了传统的图书馆主题理论的边界, 它是对客观知识结构的可编码和处理的描述, 是一种认识论模型, 是一种本体论。有人称FRSAD是书目世界的完整模型, 与FRBR相关, 见图1。

如图2所示, FRSAD的核心是Thema-Nomen模型, 这个模型假设任何作品(work)都具有主题, 主题被称之为Thema, 任何

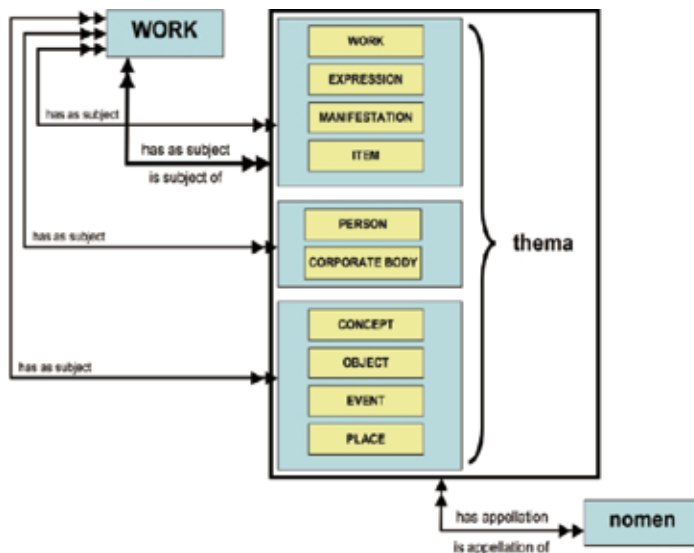


图1 FRSAD与FRBR的关系(出典: Marcia Zeng)

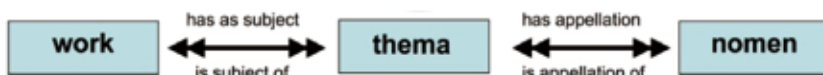


图2 FRSAD的概念模型

Thema都可以表达成Nomen。对应于作品(Work), Thema也是抽象的, 而Nomen是Thema的客观表达, 即用各种语言、符号表达Thema。FRSAD涵盖thesauri、classification schemes、subject heading systems、taxonomies、ontologies、pick lists以及terms、notation、category labels、identifiers等知识体系。由于Thema的存在, 图书馆才可能用不同的标引系统来描述同一种文献, 实现不同标引系统之间相互映射。模型的主要内容是阐述Thema和Nomen的属性和相互关系, 界定各种关系是模型的关键。

2.2 关于Topic

在Topic Maps和DITA技术中, 对其所用的Topic有明确的定义。作为一个单独的词汇, Topic为何? 有些人将Topic视作Subject的同义词, 还有一些人将Topic视作Subject的要素。出于不同的环境, Topic的定义也是形形色色。这里首先归纳一下语言学上Topic最通用的几种定义^[4]:

a)它是子句中的短语, 用以理解该子句其他部分论及什么。

b)它是语篇中的短语, 用以理解该语篇的其他部分论及什么。

c)位于子句特殊位置(通常在子句的右边或左边), 主题通常会出现在该位置。

尽管在通常的英语句子中, Topic与Subject视为等同, 但是, 从上述Topic定义中不难看出, Topic的原始含义更接近于中文的话题和论题, 没有像Subject那种规范控制。由一组相关的Topic构成的Topic图与本体类似, 但又不像本体那样过

于抽象和复杂。所以, Topic具有很大的灵活性。

从曾蕾老师对Subject模型的描述来看, Subject是一个非常复杂的体系, 是概念的范畴, 具有很强的指导意义。而具体到实施环节, 没有可以完全复制的技术路线, 相关的应用还处于探索之中。在知识信息时代, 为继承Subject理论, 将其用于语义信息组织, 同时, 为了避免Subject的选择是否权威和规范这类无休止的争议。用Topic来取代Subject, 无疑是一种最佳的实践方案。Topic既可以单纯作为一个话题处理, 也可以向复杂的Subject概念模型和本体的方向发展, 灵活性强。这也是本研究与Topic相关技术的初衷: 面向快速构建应用型语义化信息系统的目标。

以下将从语义信息组织的视角, 集中讨论两种与Topic有关的标准技术, 即Topic Maps和DITA, 说明Topic的应用场景, 与读者分享一种更实用化的、轻量级语义信息组织手段。

3 关于“Topic Maps”

3.1 Topic Maps技术^[5]

主题图原创于90年代的末期, 用于表现图书的索引结构, 以使不同来源的索引可以合并。不久, 主题图的开发人员就意识到, 通过小小的泛化就可以建立一个“元层模式”, 满足更加广泛的应用。由此, 诞生了ISO的标准。主题图的国际标准为ISO/IEC 13250, 用于知识的表现和交换, 强调信息可发现性。主题图在很多方面类似于概念图(concept maps)和思维导图(mind maps), 然而只有它被标

准化。

主题图采用3种要素来表现信息:

■ **主题(topics)**: 表现任何概念, 包括人、国家、组织以及软件模块、个别文档和事件。

■ **关联(associations)**: 表现主题之间的超图(Hypergraph)关系。

■ **信息源指引(occurrences)**: 表示与某一特定主题相关的信息资源。

主题图也是语义技术的一种形态, 在W3C的语义Web标准RDF/OWL/SPARQL家族和主题图标准的家族之间的互操作问题上, 人们已经做了很多工作。主题图的语义表现力等同于RDF, 但主要不同是: 1) 主题图提供了高层次的语义抽象, 即提供了“主题、关联和信息源指引”的模型, 与此同时, RDF仅仅提供了通过一种关系所关联的两个主题的某一模板; 2) 主题图在节点之间允许N元(n-ary)关系(即超图hypergraphs), 而RDF仅限于三元组。

主题图可以表示知识概念间的相互联系, 也可以定位某一知识概念所对应的资源位置。所以它可以称作主题知识描述和知识导航技术, 是一种十分灵活的知识模型。

3.2 Topic Maps应用

主题图的应用领域涵盖提供共享词汇、语义检索、信息/知识的结构化和体系化、信息/知识管理、内容管理、信息抽取、Web门户、制造管理、信息公开、学习支持等。

其开发步骤包括: 决定应用的问题域和范围、本体构建和主题图构建。如下以“研究会”为应用案

例,介绍主题图的构建过程:

1) **决定应用的问题域和范围**。本案例应用的问题域是地域研究会的相关主题和信息。范围为最近3年间的相关活动。

2) **本体构建**。主题中的本体,列举问题域中存在的词(主题&概念)以及之间的关系,对它们进行范畴归类、构建类。本案例主题图中的本体,描述了以下的Topic类型和关系类型。

Topic类型: 组织、场所、人、发表、事件、研究领域、信息技术、地域等7类。

关系类型: 位于、所属、主办、发表、构成、分类、利用信息技术、以地域为对象等8类。

有关地域及研究领域等,可以描述包含关系。有关人,也可以定义师生关系、父子关系、共同研究者关系等等。

3) 主题图构建

由以上的案例看出,主题图的构建弹性很大,可简可繁。可小规模启动,并进行合并。由于主题图的案例介绍较多,笔者在以前也发表过相关的综述研究和应用开发,本文不再赘述。

4 关于“DITA”

4.1 DITA技术^[6-9]

DITA(The Darwin Information Typing Architecture),全称为达尔文信息类型化结构体系,起源于“简约”(Minimalism)理念的资源创作技术。“简约”理论是人机界面领域的杰出贡献,其核心思想是“对于某一需求,与其提供读不完的信息,不如提供‘简约’的信息”。DITA是基于简约的一种

实用化、强有力的信息分类结构体系,是基于Topic的资源创作体系,也是当今构建信息的最强大的工具箱。它强调达尔文的进化论思想,即信息分类结构体系要具有“专业化”和“继承性”,可理解为信息的主题化和重用,也就是DITA允许新定义的元素能够继承已有元素的行为和属性,DITA内容、设计和过

程都可重用。主题是DITA中最小的信息单元,由标题和文本组成,足以回答读者的一个问题,又不会将其精力分散到无关信息上。见图3,简单地说,DITA体系是将资源主题化、构建关联主题的主题图,并以多种形态进行发布的机制。

DITA的历史

DITA是IBM公司的原创发明,

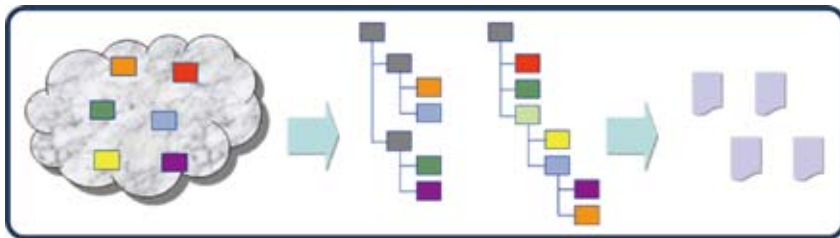


图3 DITA的概念图:资源主题化、DITA图构建、多形态发布

并贡献为OASIS标准。它是一个基于XML的体系结构,用于创建和发布技术信息。这种结构包含了XML架构的思想,包括模块化的信息结构、内容重用和专门化的机制等。在过去的几十年里,DITA取得了不断的发展。尽管目前主要仍应用于技术出版物、操作手册和联机帮助,但作为基于主题的语义信息组织与发布,开始在相关领域发挥着不可替代的作用。

DITA的7个特征:

- 1) 面向主题:信息单元涵盖一个具有特定含义的特定主题;
- 2) 主题粒度:可被独立检索的独立、自包含的单元;
- 3) 主题集合:基于可用的主题词池组合成可交付物;
- 4) 强大的类型化:具有各类信息自定义语义的机制,比如含步骤的主题;
- 5) 专门化:通过定义某一特定案例的新类型进行扩展;

6) 类型层级:类型可以层级细分;

7) 重用:DITA内容、设计和过程都可重用。

DITA元数据:将内容的价值杠杆化

参见图4,DITA支持一个丰富的元数据集,用以过滤主题列表和主题内容,满足个性化。

4.2 DITA应用^[10-12]

DITA被认为是当今构建信息的最强大的工具箱。DITA工具包括将信息内容按主题制作和编辑,具有构建元数据和主题关联图以及内容重用等功能。用户可以按照主题关联图获取必要的信息。通过实施DITA,机构可以从他们的技术文档中获取比以往任何时候更有价值的信息。已经采用DITA技术的软件行业有IBM、Oracle、EMC、Adobe,高科技企业Intel、

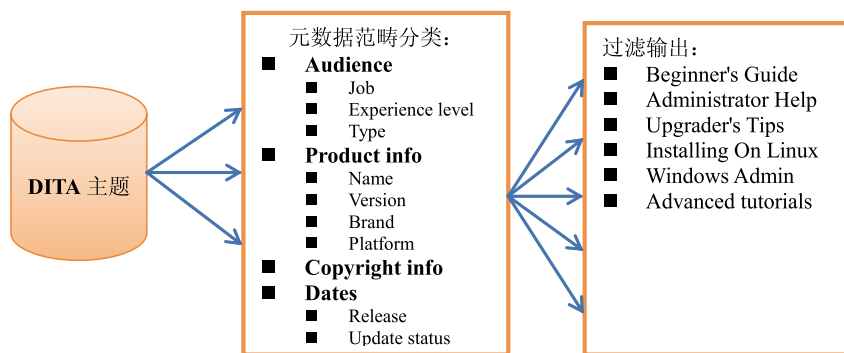


图4 DITA元数据

ATI/ADM、Philips，通讯行业的诺基亚，以及汽车关联企业、医疗器械行业的Siemens Medical和航空宇宙的波音公司。

DITA在不同领域的应用如下：

(1) 基于DITA技术文档创作：原创应用

DITA源于文档的创作和发布，是OASIS标准的XML数据模型，通过DITA开源发布的系统工具包，许多第三方机构的工具都支持该数据模型，包括Adobe FrameMaker、XMetaL、Arbortext、Quark XML Author、Oxygen XML Editor、easyDITA，以及SDL Xopus。

技术文档是软件工程的技术资

料，是软件产品的重要组成部分。技术文档要针对具有特定目的的特用户群呈现技术信息，其典型特征是任务导向。技术文档的创作代表了IT产业的发展。虽然国际技术文档行业的发展越来越成熟，文档创作技术与流程较为规范，而我国技术文档创作还没有受到足够的重视，DITA技术创作的名词知者甚少，恐怕文档工具的使用还多局限于MS Word。

DITA支持按信息类型分类的主题文档创作和发布，这些主题能够通过DITA图（maps）相互联系并整合发布为PDF版本、e-Learning课件、在线帮助等各种交付文档，见图5。

DITA创作流程关键步骤：分

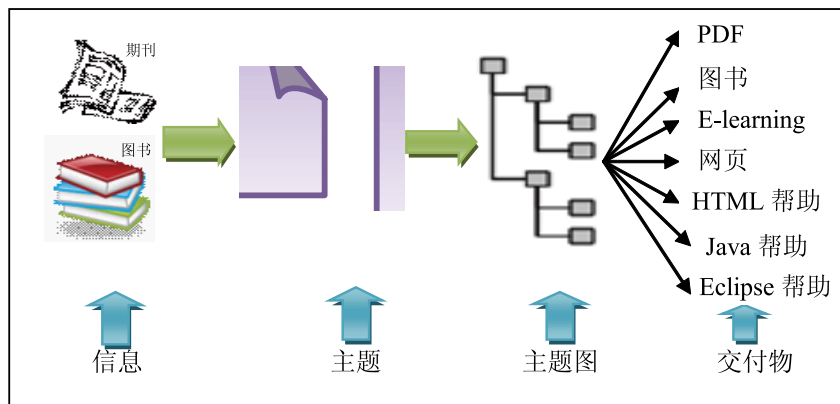


图5 基于DITA技术文档制作流程（概念图）

析文档中最小的可供维护更新的独立可重用单元，即主题；确立每个主题的类型，即任务主题、概念主题或参引主题等；识别元数据，以实现定制；依据用户的需求，定制化输出格式，实现单源文本的多元发布。

作为基于XML的信息结构，DITA继承了很多XML的优势，内容与格式分开，使文档工程师更加关注文档内容，提高文档质量。将文档归类为概念、任务和引用主题，使信息高度集中，能够确保用户更快地操作和完成任务。同时，用户可以根据公司性质、工作职责和任务目标等各种元素来检索信息。

基于主题创作与传统的面向书籍的文档写作方式相比，优势在于：

- 内容的可读性强。由于每个主题是独立的，信息也具有很强的可读性。读者无需按照顺序阅读各个章节，可依照主题按需选择。

- 文档的重写更易实现。文档能够很方便地在其他地方重复利用。

- 信息架构一致有效。不同的文档使用类似的信息结构，可更容易地设计新信息结构。

- 确保文档工程师所提供的信息为任务导向。技术文档写作的重要原则之一就是任务导向。而任务主题是DITA中最核心的主题。使用DITA可确保技术文档写作人员在写作过程中将主要的注意力集中于任务主题。

- 为用户提供可供预测的在线帮助结构。由于利用DITA发布的帮助文档的结构基本一致，无论文档工程师使用何种DITA工具进行发布，用户在第一次使用文档之后，对其后文档的结构都会比较熟悉。

(2) DITA在语义信息组织上的扩展应用

基于DITA的企业内容管理（ECM）架构。如图6所示。基于主题，内容仓储的粒度更加细化、重用和聚合内容更灵活，可定制化地发布各种格式的输出。

基于DITA和SCORM的e-Learning交付物。DITA是领先时代并以内容重用为特征的内容处理解决方案：DITA主题可重用和结构化，是语义内容组织的基础；DITA支持技术内容，因为它面向主题、基于任务。为此，针对e-Learning的交付物，可采用DITA进行目标再定向，DITA通过SCORM将任何面向主题的内容与learning交付物进行互连。通过表1的对比，可以设计出图7所示的基于DITA内容的新型learning处理模型。

DITA与S1000D。S1000D也是一个全面和完整的技术出版物国际规范，其丰富的内容主要用以支持装备技术出版物复杂的业务流程。通过表2中S1000D与DITA的对比，DITA可以弥补S1000D的不足。

(3) DITA与维基百科及混搭的融合

DITA是一个构架内容的开放标准，涉及内容、元数据及导航三个层面，在内容层，通过DITA的特化主题支持特定的信息类型，比如任务类型。核心内容与元数据分开，元数据是对内容的主题层和聚合的图（map）层进行管理。导航是通过DITA图来管理主题之间的关系，有内容列表、网站图及相关链接等导航要素。

目前流行的内容创作手段主要有维基和混搭，它们是企业的强大工具，利用维基可以快速、容易、

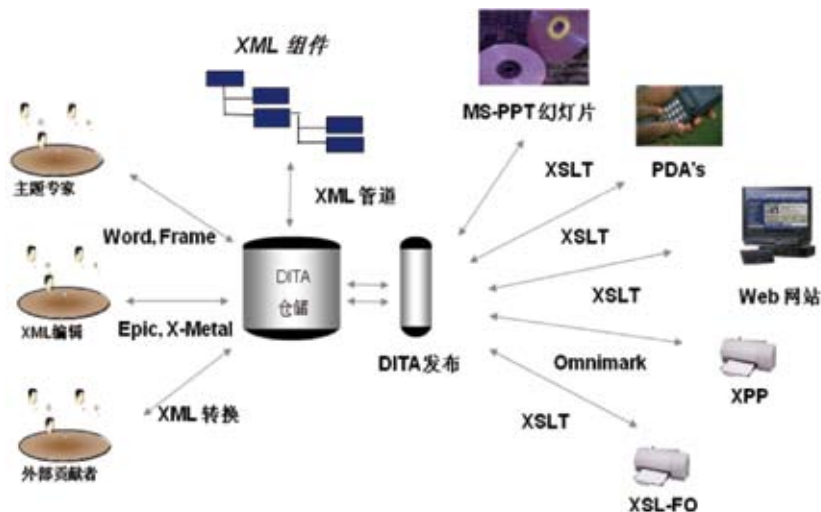


图6 基于DITA的企业内容管理架构

表1 SCORM与DITA对比

SCORM	DITA
SCORM将学习内容作为独立的、可重用信息单元进行存储，即Sharable Content Objects，简称SCOs（可共享内容对象）；	DITA内容作为基于主题信息单元进行存储；
SCORM依据确定的学习目标来聚合和排序SCOs成为较大的学习单元；	DITA图通过排序或主题关系，组织主题为交付物；
个别SCOs没有特化的语境，使得他们在多个交付物中可以重用。	DITA图可以通过目的和任务组织一组主题；
SCORM预先设定了一个特殊的内容模型，SCORM本身不预先设定某一特殊的内容格式。	DITA主题不依存语境，且可独立应用。
	DITA主题和图为基于SCORM的学习交付物，提供基本统一的XML内容模型。

表2 S1000D与DITA的对比

S1000D	DITA
S1000D打散（细分）信息为自包含的数据模块；	DITA打散信息为基于主题的单元，它可以被特化；可以小规模地启动；可以很快地设计新的信息类型，作为目前核心类型的特化，DITA可以模型化所有的S1000D信息类型，例如可以在不同的领域共享内容，每个领域可以具有信息的不同类型。例如，可以把军用设备指令与医药指引整合。
提供预先定义的一组信息类型来支持技术发布，它为民用和军用设备、海陆车辆或设备进行描述操作和参见；	使用DITA图来层级地结构化内容或关联主题。
资源内容的存储和检索采用通用资源数据库；	
采用出版模块将内容以多种格式输出。	

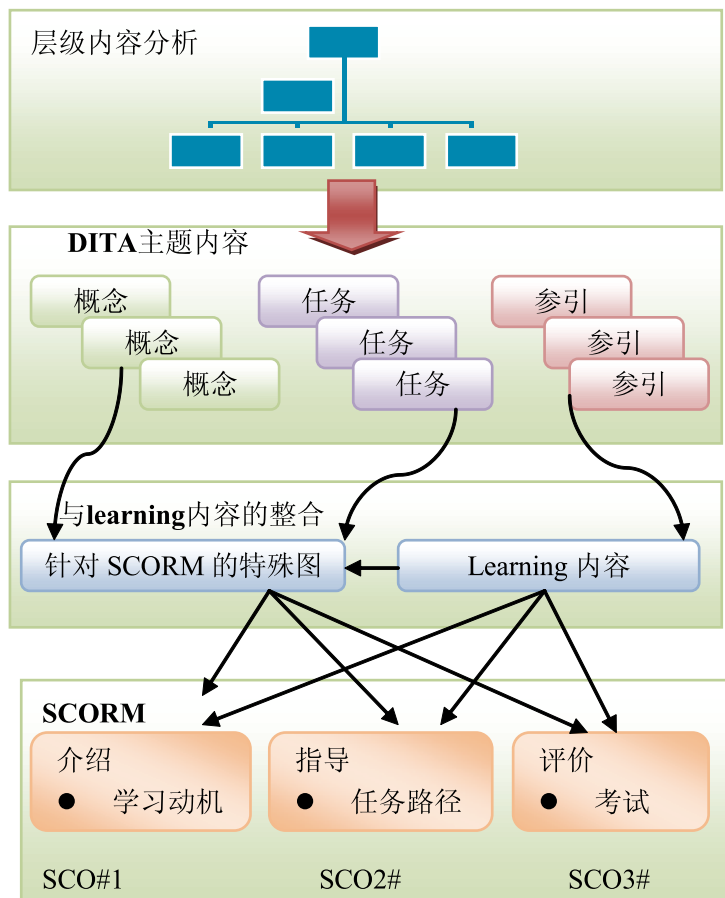


图7 基于DITA内容的learning处理模型

开放地对内容进行协同构建。混搭是快速、弹性开发的战术，是对可信内容和数据的杠杆性投资。维基与混搭可以做到容易协同、较快创新。

维基的问题是：内容非结构化，即使认为它有模板或隐含的语义，但无验证；内容非标准，将内容

从维基导出，即使在不同的维基之间，很困难；内容是混杂的，选择内容直接会导致断链。

混搭的问题是：内容资源非标准，每一个新资源都意味着新的构件或控件；混搭为非标准，不能与其他应用甚至其他混搭引擎共享混搭定义；混搭不堆栈（Mashups

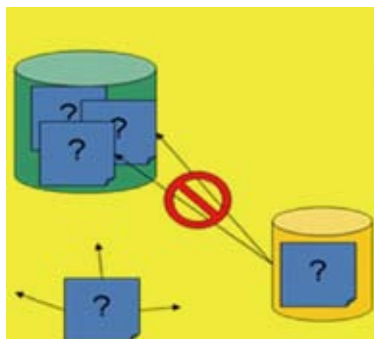


图8 维基的问题

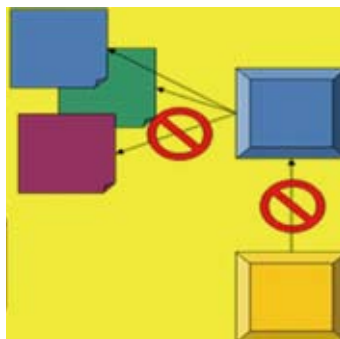


图9 混搭的问题

don't stack），每一个新的混搭都是一个非标准内容的新资源。

标准解决方案：XML，将内容用与应用分开；在应用间共享内容；DITA，强调重用标准内容资源；层叠集合标准，让集合重用集合；新内容类型和集合类型与既有应用协作。

应用场景：维基构建DITA，发布到维基；构建DITA，聚合到维基；构建DITA，移植到维基；构建维基，聚合到DITA；构建维基，移植到DITA；或一个native DITA维基。

构建DITA，发布到维基：DITA包含资源；维基向外发布提供共同资源，例如：

DITA采用Wiki可做到：

- 用wiki的输入与输出
- 用DITA图跟踪wiki主题（条目）的变化

• 使能DITA的Web编辑器，但与wiki又不完全一致

• DITA到维基文本XSLT转化，反过来，编创者也可决定什么内容回到DITA资源。

5 基于“Topic”的语义信息组织方法的总结

本文从语义信息组织的视角，集中讨论两种基于Topic的标准技术：主题图和DITA，并重点介绍了还没有进入国内数字图书馆领域的DITA技术。作为两种语义信息组织技术，主题图重点是在元层构建主题图，更适合于知识导航，而DITA是从资源层的主题创建到多格式产品交付的系统化框架。主题图与语义Web家族的融合性越来越好，却没有RDF的三元组的限制，DITA对资源结构定义和内容重用方面有



图9 DITA与其他技术融合

的示范,为图书馆的知识服务与知识管理提供了基于图书馆经典理论的解决方案。

基于主题图的应用,实践已经不少,读者可以参考相关资料。对于想研讨DITA技术的读者,我们提出如下的建议:是否有长期的知识内容管理需要、但又想从小规模做起?是否有内容重用和聚合的需求?是否有多种格式输出的需求?是否有系统作业的需求?如果有这些需求,DITA技术和DITA与现有技术的融合,目前看来,是一种最佳的选择。

着不可替代的优势,可以弥补目前流行的各种内容知识组织和管理技
术的缺陷,也为图书馆主题概念的抽象模型的实施,提供了最佳实践

参考文献

- [1] Subject [OL]. [2012-02-25]. [http://en.wikipedia.org/wiki/Subject_\(documents\)](http://en.wikipedia.org/wiki/Subject_(documents)).
- [2] Subject [OL]. [2012-02-25]. <http://www.ifla.org/files/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>.
- [3] IFLA. Functional Requirements for Subject Authority Data (FRSAD) [OL]. [2012-02-25]. <http://www.ifla.org/en/node/1297>.
- [4] Topic [OL]. [2012-02-25]. [http://en.wikipedia.org/wiki/Topic_\(linguistics\)#Definitions](http://en.wikipedia.org/wiki/Topic_(linguistics)#Definitions).
- [5] Topicmaps [OL]. [2012-02-25]. <http://en.wikipedia.org/wiki/Topicmaps>.
- [6] Architecture [OL]. [2012-02-25]. http://en.wikipedia.org/wiki/Darwin_Information_Typing_Architecture.
- [7] DITA [OL]. [2012-02-25]. <http://dita.xml.org/>.
- [8] OASIS. Darwin Information Typing Architecture (DITA) Version 1.2 [OL]. [2012-02-25]. <http://docs.oasis-open.org/dita/v1.2/os/spec/DITA1.2-spec.html>.
- [9] DITA Best Practices [OL]. New Jersey: IBM Press, 2011.
- [10] Building an XML Publishing System With DITA [OL]. [2012-02-25]. <http://www.slideshare.net>.
- [11] 祝颖.基于DITA的技术文档写作研究[D].北京大学,2010.
- [12] Feedable, portable, mashable, DITable Michael Priestley [OL]. [2012-02-25]. <http://www.slideshare.net>.

作者简介

李颖, 信息系统专业博士。相关研究课题: 语义知识组织、基于主题的知识组织技术的应用等。E-mail: liying@istic.ac.cn
张秀梅, 副研究员, 《数字图书馆论坛》主编。E-mail: xiumei@wanfangdata.com.cn

Study of Semantic Information Organization Technologies Based on Topic – DITA and Topic Maps

LI Ying, Zhang Xiumei / Institute of Scientific and Technical Information of China, Beijing, 100038

Abstract: Semantic information organization based on Topic and classical Subject Theory of library science are in separable, which can be seen as an evolution and practical subject technology of conceptual Subject Theory in the Semantic Web environment. Firstly, this article interprets the concept of Subject and Topic, overviews two kinds of international standards technology based on topic, namely Topic Maps and DITA; then, it briefly describes Topic Maps technology and application, and focuses on DITA technology and application that has not been known in domestic libraries; finally, it wraps up semantic information organization based on topic.

Keywords: Topic, Subject, DITA, Topic maps, Semantic information organization

(收稿日期: 2012-03-15)