“进化型”结构化信息体系的构建
——基于DITA结构化信息体系的架构及实施技术路线

□ 杨公亮 栗晓磊 / 北京万方数据股份有限公司 北京 100038

摘要：文章首先介绍了DITA（达尔文信息类型化结构体系）的基本概念以及其在技术文档写作中应用的优势。其次，文章通过浅谈三种DITA主题——任务主题、概念主题与参引主题，讲述如何用DITA进行技术文档创作。再次，文章展开对DITA创作架构的介绍，阐述了如何通过DITA图和导航、链接、元数据、条件处理属性以及内容重用等功能进行文档的架构。最后文章讲述了如何将传统信息转化为DITA的实施技术路线，组建可检索的、可组织的、可重用的信息。

关键词：DITA，面向主题创作，信息架构，DITA图，元数据
DOI：10.3772/j.issn.1673—2286.2012.04.003

序言

二十一世纪信息化技术发展日新月异，数字化出版已逐渐成为内容出版的主流形式。如何对数字内容进行高效的组织与利用，也成为各领域所关注的热点。

技术文档作为重要的参考资料，在IT产品与服务的设计、研发、生产、使用及维护等各个阶段中都起着至关重要的作用。优秀的技术文档不仅可以帮助IT开发人员进行更高效的生产活动，还能指导用户快速掌握产品的使用和维护技巧。管理人员也可通过技术文档完成有序的项目管理工作。

对于知识密集型的企业来说，企业的内容资产数量庞大而且构成复杂，如何对这些内容进行有机组织与长期管理已经成为企业管理中亟待解决的重要问题之一。同时，对于许多国际化企业而言，如何在激烈的市场竞争中低价高效地将新产品在分布于全球各地的目标市场上快速完成本土化，也是企业发展壮大过程中必须要面对的核心问题之一[1]。IBM公司作为全球领先的信息产品与技术服务商，在上世纪末已经遇到了上述难题，并通过多年实践提出了达尔文信息类型化结构体系DITA（Darwin Information Typing Architecture）[2]。

1 认识DITA

1.1 DITA基本概念

DITA是一套基于XML的结构体系，面向主题，用于结构化创作与发布技术信息[3]。DITA通过在XML架构中定义出一套DTD或Schema的方法，为技术文档的数字化出版提供开放的标准文档格式，并在此基础之上为数字化内容从加工生产到发布应用提供一套完整的解决方案。

1.2 DITA在国内外发展情况

DITA发展至今已经过了11年历程[4]：

- 2001年3月：由IBM公司首次提出DITA的概念；
- 2004年4月：针对DITA的OASIS技术委员会成立；

- 2005年2月：SourceForge支持DITA开放工具集；
- 2005年6月：DITA 1.0版本获批成为OASIS标准；
- 2005年8月：DITA开放工具集1.0版本发布；
- 2006年3月：OASIS发起成立DITA在线社区http://dita.xml.org/；
- 2007年8月：DITA 1.1版本通过OASIS审核，其中包括Bookmap的应用；
- 2010年10月：DITA 1.2版本通过OASIS审核，增加了新的内容重用特性、行业定制以及增强的术语控制等内容。

1.3 DITA的特点

DITA的特点可以从两个方面概括：模块化和专门化[5]。

模块化：在DITA体系中，文档的内容被划分成多个模块，这样既可以实现内容的重复利用，也便于在不同模块间跳转。相对于传统的文档创作模式，使用DITA进行文档编创将更加灵活可控。

专门化：通过内容的专门化，DITA的功能可以得到扩展。专门化使得信息具有了面向对象的特性。文档编创与使用人员只需要使用他们需要的信息，开发人员也可以根据需求随时定制新的信息类型，并且这些新的信息类型可以保持与原有类型的一致性。这一特性大大地增强了技术文档的可读性，且降低了文档的更新维护难度。

2 使用DITA进行创作

2.1 基于主题（Topic）的技术文档创作

DITA的设计初衷是构建一种基于主题的技术文档创作模式，并以主题为基础来创作、组织和链接信息内容。

所谓主题，在技术信息中有时称为一个条目（article），它具有标题和内容。主题是独立的信息单元，一个高质量的主题仅包括一个讨论对象。每个主题既要足够长以保证有其独立存在的价值，又要保持精简（“简约”理念）以确保仅集中于讨论一个对象，而不延伸到别的问题。在多数与产品、服务或技术一起提供的技术文档中，主题应该从属于一组经过有序组织的主题集。这一主题集能够以HTML网页、在线帮助或者PDF手册等多种形式打包输出。

DITA的价值不仅体现在可以让读者更好地阅读技术文档，还体现在它也能为技术文档的编辑创作者提供极大的便利。对于信息的使用者来说，精心创作的基于主题的内容能够大大提升文档的可检索性、可导航性及可用性。而对于编创团队来说，基于主题的创作方式还能支持信息的重用与快速重组，以及便捷的文件管理与更灵活的链接[6]。

2.2 使用不同类型的主题

为了便于创建和发布按照类型和使用目的进行拆分的信息内容，DITA中提供了三种主题类型：概念主题、任务主题和参引主题：

- 任务主题：描述流程
- 概念主题：定义某事物是什么或者一个过程如何运作
- 参引主题：包括用户在完成某项任务时可能会需要的参考性信息

表1通过举例的方式，展示了三类主题在标题上的差异。后文将详细介绍三类主题的应用方式。

表1 不同类型主题的标题命名的实例

概念主题标题	任务主题标题	参引主题标题
用户角色	创建用户角色	支持的角色类型
高清电视	安装高清电视	电视机附件
意式咖啡	制作意式咖啡	意式咖啡配料
猫的行为	驯养猫	国内猫的品种
数据库	在企业级系统中配置数据库	数据库的类型
摄影	拍摄风景照	数码相机类型和功能

(1) 任务主题

DITA的任务主题具有良好的结构，可以帮助人们创作高质量的流程性信息。任务主题中包括很多在概念主题或参引主题中所不具备的XML元素，这些元素为内容提供了一个框架结构，如前提条件、步骤、示例和其他在创建流程中可能会用到的内容。

任务主题通常需要概念主题或参引主题提供支撑，但在通常情况下，用户会直奔任务主题，而仅在需要的时候才关注概念性或者参引资料。例如，在介绍如何训狗的用户指南中，需要创建任务主题“教简单的命令”，为了支持这个任务，还需要创建概念主题“家犬的行为”和参引主题“训练过程中狗的喂养和驯化”，这几个主题一起帮助用户实现其目标“训练狗使其按命令去行动”。

若需要向用户提供面向任务的信息以帮助他们完成现实任务，创作高质量的任务主题则非常重要。创作任务主题时应遵循以下的原则：

- 区分任务信息、概念性或参考性信息
- 一个主题描述一个任务
- 创建任务和子任务来组织较长的流程

(2) 概念主题

概念主题通常解释和定义概念，一般包括用户在使用产品或开始任务前需要了解的背景信息。创造有用的面向任务的信息不仅仅需要好的任务主题，还需要好的概念主题。为了实现这一目标，读者需要清晰、直接的流程描述，以及可扩展的概念性信息。

创作概念主题时遵循以下原则：

- 一个主题描述一个概念
- 仅当问题无法在别的地方准确描述时才创建概念主题
- 将任务信息和概念性信息分开

(3) 参引主题

在创作面向任务的技术内容时，通常会从用户关注的任务主题开始编创，然后再编写那些为了让用户理解想法以完成任务的概念主题。但是，仅仅依靠任务主题和概念主题还不够，编创技术文档的过程中经常需要用到参引信息来进一步支持这些任务。

参引主题是一系列实际因素的集合。这些实际因素可以是一个零件列表或零件描述、命令、应用编程接口（API）、书名、车型、动物物种、自行车轮胎尺寸或其他任何可以组织为一个表格或列表的对象。

创作参引主题的时候，要遵循以下准则：

- 每个主题只描述一种类型的参引材料。
- 逻辑化组织参引信息。
- 保持参引信息的一致性。

2.3 善用简短描述 (Short Descriptions)

<shortdesc>元素可能是技术文档编创过程中最灵活但也是最具挑战性的元素，因为它不仅作为每一段主题的第一段文本，有时也会出现在链接中或搜索引擎结果中。在将现有文本转换成DITA后，初次编创DITA主题时，需要确保优先完成有效的简短描述编写工作。简短描述必须简明扼要地描述主题要点或目的。当用户开始一个新的主题时，通过阅读第一句文本，他们就可以迅速判断这些信息是否是他们所需要的。如果应用公共搜索引擎能够抓取到发布的在线HTML内容，搜索结果中会显示简短描述。能够使用户在发布的HTML内容中有效地找到所需信息是创建有效的技术文档过程中最终的一关。如果用户无法找到他们需要的信息，主题写得再好也无意义。

在编写简短描述时，要遵从以下准则，以保证简短描述表达简洁：

- 在每一主题中插入简短描述
- 不要描述表格、列表或数字
- 使用完整的、语法正确的句子
- 不要简单复述主题标题
- 不要使用自我参引的句式，如“这个主题描述……”

3 基于DITA的架构

在熟悉了上文使用DITA进行创作之后，就需要开始架构信息，需要组建可检索的、可组织的、可重用的信息。而为了达到这个目的，就需要了解基于DITA的架构，主要包含以下五个方面：DITA图与导航、链接、元数据、条件处理属性和内容重用。

3.1 DITA图与导航

关于DITA的架构，首先要知道的就是DITA图。依靠DITA图，用户可以跟踪信息路径。良构的DITA图可以帮助用户快速地找到所需信息，提供一条支持产品使用的任务流程，甚至可以给创作者提供一些捷径。使用DITA图可以包含主题到一个信息集里、可以定义信息架构、可以创建主题之间的关系，如图1所示。

DITA图文件的层级关系中定义了两个或者多个主题之间的关系。包含其他主题的被称为父主题，嵌入的

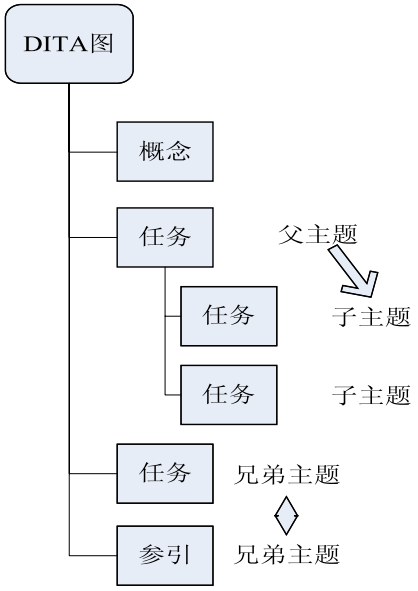

图1 DITA图的构成

主题被称为子主题，一同嵌入在同一层级的主题互称为兄弟主题。每一个参引主题可以参引一个主题、一个DITA图甚至一个非DITA文件。

有了DITA图之后，还需要进行信息建模。信息模型可以勾勒出哪些主题去哪，哪些主题是相互链接的，哪些主题需要什么样的链接关系。支持DITA的IBM Information Architecture Workbench（IBM信息架构平台）这样的建模工具可以去建立一个良构的主题层级关系。这些工具是可视化的辅助工具，可以组织主题到DITA图的层级中。

一般情况下，一个简单的DITA图代表一个主要的用户目标。但是，有很多时候，例如一个复杂庞大的帮助系统可能产生一个非常大的DITA图。所以，在一些DITA图中嵌套小的DITA图去建立子图，利用子图去细分内容进而更好地管理内容。

利用子图可以做到以下几点：

- 按照章节组织内容
- 更好地管理大型信息集
- 重用主题集
- 支持多人同时在同一个信息集上工作
- 剥离经常更新的内容，使其可以更容易被更新

有了信息的架构之后，信息架构师就可以分配一个或者多个主题给创作者去添加正确的内容以支持特定的产品、解决方案或者策略。DITA图允许直接参引非DITA内容，比如，信息中可能需要包含一个直接指向PDF或者外部网站的链接。

不管是转换内容为DITA或者是新建一个文件，都需要决定要建立什么类型的DITA图：一个需要创建出多种输出格式的DITA图，或者对应每种输出格式的独立的DITA图。如果决定运用一个具有多种输出格式的DITA图，就需要决定是否禁止一些主题一同出现，或者是否不希望这些主题出现在目录中。

3.2 链接

有了DITA图之后，下一步就是要帮助用户找到那些信息，并且帮助用户从一个主题转向另一个主题。这一步指的就是精心设计的链接策略，它能创作有效的网站信息，并且改善内容的恢复和导航。

基于内容的主题只有链接到其他相关联的主题之后，才可以成为完整和有用的信息。通过正确的链接，用户能够很容易地顺着任务流找到相关主题，帮助他们完成目标。

链接分为以下四种：

（1）层级链接

层级链接指的是通过嵌套主题显示信息的导航。

（2）行内链接

行内链接指的是穿插在主题文本中的链接。

（3）关联链接

关联链接指的是穿插在关系表中的链接。

（4）集合类型链接

集合类型链接确定了嵌套主题之间的关系。

使用这些链接，可以将主题串联起来，形成一个信息条理清晰的网络。合理的链接对于老用户和新用户都有帮助：

- 新用户能通过链接获得更多信息。
- 老用户能跳过他们不需要的链接。

但是需要注意一点，链接太多也一样效果不好。一方面不要让用户搁浅在断尾的主题上，一方面要为用户提供足够的相关关联链接使其不至于迷失。必须在这两方面上找到正确的平衡点。重复链接很麻烦，链接太多会让人难以招架，链接到错误的主题更是无助于解决问题。

最后，决定一个链接策略，并且评估链接的有效性。有效的链接将提升信息的质量，并能够丰富信息的架构。

3.3 元数据

元数据用于描述DITA中所含主题和DITA图的信息。通过正确使用，元数据可以帮助编创人员产出针对特定用户、产品、版本等的内容，这些有针对性的内容能帮助用户找到需要的信息。

使用元数据是能同时帮助产品用户和编写人员提高内容可检索性的最佳途径之一。如果用户无法找到他们需要的信息，那么主题信息再精准、全面和严谨也无济于事。

良好的元数据策略可有助于创建的信息：

- 容易发现：元数据，如关键词和索引入口，可以帮助检索引擎和用户发现特定的内容
- 容易管理：内容管理工具可以使用元数据来查找和组织已经正确分类的主题和DITA图
- 面向特定读者：可以编写一组主题让它们包含（或者排除）基于版本、型号、操作系统、用户类型等属性的信息，以确保特定的用户只看到他们需要的信息

元数据也能实现诸如动态出版或分面检索等前沿的内容产出功能。例如，可以设置一套出版机制，这样用户就可以按照内容的类别来选择他们需要的信息。如按照软件产品对应的操作系统、服务器类型，或者硬件产品对应的型号、尺寸等。

元数据的类型

DITA拥有多种元数据类型，依据内容，可以使用一些或全部类型的元数据：

- 索引入口
- 条件处理属性
- 主题元数据
- DITA图元数据

3.4 条件处理属性

熟悉了元数据的基本情况之后，接下来需要了解的就是一种特殊的元数据——条件处理属性。通过运用条件处理属性，可以仅维护一个源文件集合，并创建输出的多种变体。那些不同的变体通过包含或者排除内容，决定什么样的信息出现在输出里。条件处理不仅有助于从一个单一主题创建不同的输出，还可以通过使用可视化帮助标注一些指定的内容。条件处理属性可以改善创作者发现指定内容的方式。创作者能够通过检索特征编号或者产品名字等关于主题的信息，迅速地在他们的内容管理或者版本控制系统中找到想要的主题。

通常情况下，使用条件处理属性可以做到以下几点：

- 在输出中包含或者排除一些内容
- 可视化标注一些内容进而改善可用性
- 改善检索和促进可检索性

在设置条件属性值之前，需要设计一个条件处理机制。这套机制需要为每一个条件处理属性指定一个值，以便能够创建指定的输出。基于组织的规模，可能需要为每一个产品区域或者产品集创建一个机制。一个计划恰当和记录清晰的条件处理机制能够帮助创建者避免一些条件的误用和条件值杂乱无序的蔓延。

在决定了条件处理机制之后，就可以把这个机制应用到主题和DITA图中去包含或者排除内容，用可视的线索标注指定内容和帮助创作者在内容管理系统中检索主题。也可以分配一些条件值去包含或者排除内容，比如在一个主题中包含或者排除指定的元素，在DITA图中包含或者排除整个主题或者在关系表中包含或者排除链接。

条件处理属性还可以用来标注内容，比如高亮显示整个主题、句子、列表项目、代码块、段落或者其他在输出中的元素，从而改善用户体验。

3.5 内容重用

在DITA里，重用就是内容创作好之后，在任何需要的地方再重新利用。

可能其他关于重用的术语比较常见，例如单一来源或是嵌入包含条目。不管这些术语如何，有效地在DITA中重用内容就意味着更好的质量，带来一致的输出和成本的降低。

(1) 重用的好处

重用有以下几方面的好处：

- 效率：重用可以提高创作者、编者和本地化团队的效率。
- 一致性和准确性：重用可以确保一致性和准确性，因为重用内容在所有位置都是一样的。
- 风险管理：重用可以通过把易变的内容变得易于更新，降低最新变化的有关风险。

(2) 重用的方法

- 通过使用内容参引来重用一些元素
- 重用主题
- 重用DITA图
- 重用来自于非DITA的内容

(3) 重用主题、DITA图、非DITA来源的内容

因为DITA是基于主题的结构，可以更容易地通过把它们插入多个DITA图中的方法来重用整个主题。当想做以下事时，重用主题很方便。除此之外，还能重用DITA图、重用非DITA来源的内容。

(4) 为重用的创作

从基于主题的创作最佳实践中可知，精心创作的主题在任何上下文中都有意义。创建有效的基于主题的信息为重用主题提供了基础。

关于重用主题，需遵循如下创作指导方针：

- 创作主题，从上下文中抽出这些主题时它们都有意义，并且假定大量信息将从上下文中获取。
- 因为用户可能只读一部分主题，所以要包括组织特征以使用户能够容易浏览内容，并且易于抽取出内容在别处重用。例如，通过在标题中使用<section>元素，帮助用户快速在概念和参引主题中发现相关信息。
- 不使用相关术语。例如，尽可能避免指代句子、段落和表格。不要使用诸如“上述表格描述了最新附件。”设想重用了一个表格，而这个表格来自另一个主题且表格放在表示有“上述”的句子后面，将不得不重新创作主题，移除这样的参引。
- 创作基于特征的结构化内容以使它能够被重用。例如，如果一个特别特征如生态过滤器被加到浓咖啡机的Deluxe模型上，创作内容没有参引Deluxe模型或任何其他特征。随后，如果Regular模型采用了生态过滤器特征，描述这些过滤器的主题就能被重用于上述两处信息中。

4 DITA的实施技术路线

DITA的种种特性让技术文档的创作和维护发生了天翻地覆的变化，但从传统的技术文档创作模式转换到DITA，有可能是一个巨大的工程。良好的实施技术路线可以避免诸多意外问题的发生，从而提高转换效率。

在完成DITA的转换后，为了确保内容的高质量，不仅需要编创内容本身，还要编排DITA标记或DITA代码。通过编辑DITA代码，可以让团队中的每个人都制作出一致的输出。而编创内容，可以使用户看到清晰、一致、完整、准确、可检索及其组织良好的信息。

4.1 将内容转换成DITA

精心设计的转换方案是成功的必要保障，遵循如下7个步骤，将使得转换工作更加有保障：

（1）进行实验性转换

（2）评估内容的状态

（3）对转换工程要计划、计划、再计划

（4）在转换之前要准备内容

（5）转换内容到DITA

（6）修复转换后的问题，重组内容，利用DITA功能的优势

（7）评价处理过程和工程项目，依据需要进行调整

4.2 DITA的编排与评审

转换成DITA之后，还需要对DITA进行编排和评审，不仅仅要编排文本，还要在主题图和DITA图中编排XML代码和标记。代码评审是指DITA的源文件被个人或者团队检查，以确定哪部分的标记需要更正或应用，代码评审能充分利用语义标注和基于主题的架构优势。

代码评审是验证DITA元素一个发现的过程，即使在DITA代码评审数月或者数年后，编创者依然能发现或者创造新的而且是在标记指引中没有提到的内容类型。

进行代码评审，要按照以下步骤：

（1）制定评审计划。

（2）可选：在代码评审前对内容的语法，标点和样式进行编排。

（3）评审DITA代码。

（4）举行会议，与创作者一起讨论评审结果。

（5）要求创作者按代码评审的讨论结果更改他们所有的主题。

4.3 内容编排

内容编排与DITA资源文档中的DITA标记的编排同等重要。除了执行代码评审，还可以对DITA资源信息作更常规的编排，确定语法问题：样式、标点符号、具体性、完整性、清晰度，及其他方面问题。而且用DITA编排，可以在同一时间内，指出DITA标识和内容中的问题。直接对DITA文档的内容进行编排具有以下的优势：

- 编排更全面
- 迭代编排
- 节省时间和资源
- 获得更多的编排实施

结论

DITA作为基于主题的体系结构，随着数字出版的发展，必将影响技术信息的创作。上文中，介绍了如何使用DITA进行创作，如何通过DITA架构信息，将内容转换为DITA并进行编排与评审，最终将内容更好地组织成为可检索的、可重用的技术信息，进而达到降低成本、增强技术信息的长期性，使得使用者在激烈的竞争中占得先机。

同时，通过DITA的全面介绍，读者可以直观地感悟出基于Topic的语义信息组织的一个侧面，对DITA在e-Learning、语义wiki等知识系统的扩展应用，作出适宜的选择。

鸣谢

在此感谢*DITA Best Practices: A Roadmap of Writing, Editing, and Architecting in DITA*一书的翻译团队，其中有张秀梅、李颖、程煜华、徐建武、刘立营、李飞、林云水、朱锐（排名不分先后），对此文撰写中一些指导方针的建议和帮助。同时，也祝愿早日看到中文版《DITA最佳实践指南——创作、编排和架构的技术路线》的发行。

参考文献

[1] HARRISON N. The Darwin Information Typing Architecture (DITA): Applications for Globalization [C]// IEEE International Professional Communication Conference Proceedings, 2005.
[2] IBM. Introduction to the Darwin information typing architecture [EB/OL]. [2005-09-28]. http://www.ibm.com/developerworks/xml/library/x-dita1/.
[3] The Rockley Group Inc. Preparing for DITA: what you need to know [R]. 2005:4-7.
[4] Organization for the advancement of structured information standards. Darwin information typing architecture [EB/OL]. [2010-12-01]. http://www.oasis-open.org/.
[5] Organization for the advancement of structured information standards. A short introduction to Darwin information typing architecture [EB/OL]. [2009-04-02]. http://dita.xml.org/sites/dita.xml.org/files/dita-introduction.ppt.
[6] BELLAMY L, CAREY M, SCHLOTFELDT J. DITA Best Practices: A Roadmap of Writing, Editing, and Architecting in DITA [M]. IBM Press, 2012.

作者简介

杨公亮（1983-），PMP项目管理咨询师，研究方向：信息组织与信息管理、医学信息服务构建。E-mail: yanggl@wanfangdata.com.cn
栗晓磊（1985-），研究方向：信息组织与信息管理。E-mail: lixiaolei@wanfangdata.com.cn

Evolved Structured Information System Architecting – DITA-Based Architecture and Writing Best Practice

Yang Gongliang, Li Xiaolei / WangfangData, Beijing, 100038

Abstract: DITA is topic-based, and it can improve the writing of technical information. Beginning with introducing three kinds of topics which are task topic, concept topic and reference topic, this article describes how to write in DITA. And later, it gives brief explanation of DITA maps and navigation, linking, metadata, conditional processing attributes and content reuse to give the user a picture of the DITA architecture, and then it tells how to convert content to DITA, and how to do code editing and content editing to create retrievable, organized and reusable technical information.

Keywords: DITA, Topic-oriented writing, Information architecture, DITA map, Metadata

（收稿日期：2012-03-15）