

●王忠红 [新加坡]Abdus Sattar Chaudhry [新加坡]Khoo Christopher

# 用传统分类法、叙词表编制机构 Taxonomy 的实证研究

**摘要** Information Studies Taxonomy 的编制以新加坡南洋理工大学黄金辉传播与信息学院信息学系为实施机构,以杜威法和三部叙词表为结构库和词库,调查了杜威法和三部叙词表作为机构 Taxonomy 构建资源的充分性及其在主题分面编制中的作用,同时探讨了用传统分类法和叙词表编制机构 Taxonomy 类表的步骤与方法。该研究虽然在英文语境下进行,但所运用的思路与方法对于中文环境下机构 Taxonomy 的编制也是有借鉴意义的,其研究方法与研究结论对于在中文环境下用传统分类法、叙词表编制机构 Taxonomy 的研究也有推动作用。表 2。参考文献 7。

**关键词** 机构 Taxonomy Information Studies Taxonomy 实用分类系统 分类法 叙词表  
**分类号** G254.1

**ABSTRACT** Implemented in Wee Kim Wee School of Communication & Information, Nanyang Technological University, Information Studies Taxonomy uses DDC and three thesauri. In this paper, the authors analyze the roles of DDC and the three thesauri in the construction of the institutional taxonomy, and discuss the steps and methods for the construction of institutional taxonomy with traditional classifications and thesauri. Although the study is made in the English language situation, its methods can also be used in the compilation of institutional taxonomy in Chinese language situation. 2 tabs. 7 refs.

**KEY WORDS** Institutional taxonomy. Information studies taxonomy. Ontology. Classification. Thesaurus.

**CLASS NUMBER** G254.1

## 1 引言

Taxonomy 是由等级结构和主题词两个基本元素组成、具有多种用途的一种知识组织工具。它的主要功能是将资源对象类聚到一个特定的知识等级结构。Taxonomy 可用于处理源信息如自动分类、链接资源提供浏览等,也可用于优化检索提问和检索结果。目前 Taxonomy 主要应用于企业网络信息资源的组织与网站浏览、导航。

Taxonomy 在表现形式上与传统分类法、叙词表有相似性,它们常常被联系到一起。Taxonomy 与分类法、叙词表之间的区别主要体现在使用范围和职能方面,Taxonomy 常被应用于企业机构范围,主要用于支持浏览功能。Taxonomy 与传统分类法、叙词表之间的联系和区别表明了用分类法、叙词表编制 Taxonomy 的可能性和难度。用传统分类法、叙词表编制 Taxonomy 的理论可行性分析在《用传统分类法与叙词表编制 Taxonomy 的理论可行性研究》<sup>[1]</sup>一文中有阐述。

用传统分类法、叙词表编制 Taxonomy 的基本方法是将分类法作为等级结构依据,将叙词表作为子目词库。这种方法的好处是无需从零开始建词库,并且

可以继承分类法、叙词表本身的长处,既节省人力物力又有利于保证 Taxonomy 类表质量。编制 Taxonomy 的这种捷径已被若干个项目<sup>[2-4]</sup>采纳,但这些项目在 Taxonomy 编制过程中没有明确机构的环境特点,或者由于项目范围过小而未能全面涉及相关问题,因而严格来说还不属于机构 Taxonomy 范畴。也就是说,用传统分类法、叙词表编制机构 Taxonomy 的实证数据还比较缺乏。

本研究主要有三个目的:一是调查分类法、叙词表作为机构 Taxonomy 编制资源的充分性;二是调查分类法、叙词表在编制机构 Taxonomy 类表中的作用;三是试图通过该研究总结出用传统分类法、叙词表编制机构 Taxonomy 的有效步骤与方法。

## 2 用分类法、叙词表编制学术机构 Taxonomy 的实例研究

本研究用相关分类法、叙词表编制了一个学术机构 Taxonomy。信息资源管理教育为 Taxonomy 的学科领域,新加坡南洋理工大学传播与信息学院信息系为其实施机构。编制过程中选用的分类法资源为杜威十进制分类法(视窗杜威),叙词表包括 ASIS&T<sup>[5]</sup>、LISA 两部图书馆学情报学叙词表和 ERIC 教育叙词

表,其中 ASIS&T 叙词表为印刷版, LISA 和 ERIC 词表是通过南洋理工大学图书馆提供的数据库获取的电子版。该 Taxonomy 命名为 Information Studies Taxonomy。

## 2.1 用分类法、叙词表编制学术机构 Taxonomy 的步骤

Information Studies Taxonomy 的编制由三个步骤组成:第一步是调查 Taxonomy 实施机构的环境;第二步是设计、规划 Taxonomy 类表;第三步是编制 Taxonomy 类表。与以往的 Taxonomy 项目相比, Information Studies Taxonomy 的编制重视了实施机构的调研和根据实施机构规划、设计 Taxonomy 类表。其目的是使 Taxonomy 能结合实施机构的需求与特点,通过知识组织来真正促进机构的宗旨与运作。

实施机构的调研包括两个方面的工作:一是了解机构的任务目的、活动领域和成员情况,以帮助决定 Taxonomy 面向的任务和用户群体;二是了解知识资源在机构内的生产、流动、利用情况和机构的知识管理现状,以帮助决定 Taxonomy 的宗旨和角色。Information Studies Taxonomy 面向信息学系的教学科研任务,以教师、硕士生与博士生为服务的三个用户群体。由于目前缺乏一个有效的知识管理工具,系里生产的各种知识资源如教学大纲、教学讲义、学生的开题报告、学位论文以及师生的学术成果等分散在各处,无法进行集中组织与管理。系局域网上虽然建设了一个外部网络资源导航,但因久未更新已基本被淘汰。教师与学生不得不从学校教学管理系统、图书馆或万维网等其他渠道获取资源以进行教学科研工作,知识资源不能得到有效的积累与利用。Information Studies Taxonomy 的宗旨是整合、链接系内外各种知识资源,提供浏览与获取,以促进信息系的教学科研工作。建成后的 Taxonomy 将在系局域网上运行。

Taxonomy 类表的规划设计包括决定类表的学科覆盖范围、面向的内容和类表类型。根据信息学系的教学科研活动领域, Information Studies Taxonomy 覆盖信息资源管理、知识管理、信息系统与图书馆学情报学教育领域,同时研究方法与学术写作也在师生的信息需求范围之内。Taxonomy 面向的内容包括教学科研活动产生的各种知识资源和教学科研活动中需要的各种外部资源,涵盖文本、超文本和多媒体等多种载体形式。知识库的建设将以鼓励教师、学生上载自有资源的方式为主。

为有利于用户多途径浏览以及类表的控制与维

护,分面分类表被选为 Information Studies Taxonomy 的类表类型。Information Studies Taxonomy 由六个分面组成:课程资料、研究资料、教学资料、参考工具书、研究小组和主题分面。其中,课程资料专为学生修课用,教学资料针对老师教学准备用,研究小组是指按研究领域来类聚资源。

主题分面为类表中最大的一个面,它由选定的分类法和叙词表编制而成,包括基本大类、子类和子目三大部分。主题分面的编制包括四个步骤:一是确定基本大类;二是建立子目候选词;三是编制子类;最后是确定子目类名。

基本大类是自建的,由于杜威法以学科标准划分大类,不能提供帮助。基本大类的编制通过两道程序完成:第一道是调查用户信息兴趣主题,方法是从师生教学科研活动中涉及的各种知识资源如教学讲义、学术成果等中抽词,并将抽取的主题词进行汇集;第二道是参考有关资料如 IFLA 图书馆学情报学专业教育指南<sup>[6]</sup>,将代表用户信息兴趣的主题词划分为若干个合理的大类并给予合适的名称。主题分面由信息学科分支领域和相关领域、信息机构、信息与知识管理、馆藏管理与用户服务、信息与知识组织、信息查询与信息检索、信息技术、信息社会、信息产业、信息行业、图书馆学情报学教育、研究方法与学术写作等十二个基本大类组成。

子目候选词来自杜威法相关索引和三部叙词表。候选词的挑选经历两道程序,先是将用户信息兴趣主题词按领域列成表单作为挑选依据,然后根据表单从杜威法相关索引和叙词表中选词。选词不是一对一的过程,而是对用户信息兴趣的宏观把握。挑选后的子目候选词按领域和出处汇集,以准备编制子类。

子类的编制通过四道程序完成。第一道是确定上位类目,以杜威法相关类目和叙词表范畴索引为参考依据。第二道是确定划分标准,以课程教学讲义中隐含的知识结构为依据,主要选用了三个划分标准:种属关系、方面和流程关系。第三道是确定下位类目,以上位类目为开端,从杜威法相关类目和叙词表提供的词间关系中挑选下位类目,然后根据采纳的划分标准最后确定下位类目。最后一道步骤是将下位类目匹配到符合用户视角的基本大类和上位类目,匹配以课程教学资料中隐含的知识结构作为参考依据,如“数字图书馆”可归入“图书馆”类也可归入“信息检索系统”类,相关课程教学讲义提示归入“信息检

索系统”类更为合适。在多个基本大类或上位类目可选择的情况下,选一个进行匹配,然后在其它类目建立参照。

子目名称的确定依据三条准则:首先,类目名称准确反映类目含义;其次,类名表现形式简单;最后,尽量使类名风格和形式保持一致。

该主题分面在初稿完成后,征求了 11 位教师与学生用户的意见,根据用户意见作了定稿。

## 2.2 分类法、叙词表在学术机构 Taxonomy 编制中的作用与缺陷

Information Studies Taxonomy 所确定的 12 个基本大类中,只有 5 个大类的类目体系参考了杜威法的相关类目。其原因有三:一是杜威法的基本大类不能全面覆盖主题分面领域,比如其基本大类 020“图书馆学情报学”没有反映“知识管理”领域的内容,再如“学术写作”领域也不在其 10 个基本大类的覆盖范围之内;二是作为一部综合性分类法,杜威法将某些有关领域集中到一个类目,而没有提供可供参考的类目等级结构,比如用来反映图书馆学情报学教育的类目 020.7、反映信息职业和研究方法论的类目 020.9 和 001.4 等;三是杜威法某些相关类目提供的等级结构不符合用户视角,比如按学科和人群划分的“信息检索系统”类目 025.04。因此,杜威法作为主题分面 12 个基本大类的结构资源库是不充分的。但是,杜威法的类目等级结构对于确定 Taxonomy 的上位类目仍然具有重要意义。上述 Taxonomy 实例编制过程所涉及的 5 个基本大类中,大部分一、二级类目都来自杜威法相关类目,比如“信息与知识组织”大类中的一级类目分别来自其 025.3 和 025.4 类目。杜威法的相关索引还提供了部分子目候选词。

叙词表在 Information Studies Taxonomy 编制过程中也发挥了重要的作用:首先,叙词表的范畴索引可以用来帮助确定上位类目,比如“信息社会”基本大类的五个一级类目来自于 ASIS&T 范畴索引;其次,叙词表提供的词间等级关系可以用来帮助确定下位类目,比如“人工智能(Artificial Intelligence)”子类的部分子目来自 ASIS&T 叙词表;第三,叙词表中的叙词则是重要的子目词库,主题分面中的大部分子目都来自所选用的三部叙词表;第四,叙词表提供的词间等同关系还可以用来帮助建立子目替代(Use For)关系。但是,所选的三部叙词表作为子目词库似乎也是不充分的,Information Studies Taxonomy 主题分面中还

有一些来自其它出处或来自用户的子目。这些子目可划分为三类:第一类是新概念、新名词,比如 collaborative tagging、mobile information retrieval systems、digital watermarking 等,ASIS&T 和 LISA 叙词表由于更新慢未能收录;第二类是为了准确表达概念而采用的复合词,比如 digital collection development 和 web resource cataloging 等,这一类复合词不在叙词表的收录范围之内;第三类是属于机构范围内特定使用的词,比如 knowledge management education、information system development methodologies、library and information science schools 等,它们一般不在通用叙词表的收录范围之内。

## 3 用分类法、叙词表编制机构 Taxonomy 的方法

### 3.1 资源库的选择

机构资源与分类法、叙词表资源都是编制 Information Studies Taxonomy 必不可少的。机构内部资源主要起两个作用:一是调查用户信息兴趣,作为确定基本大类和从词库中选择子目候选词的依据;二是用来构建分类法、叙词表未能覆盖的基本大类的类目体系,比如 Information Studies Taxonomy 中“信息与知识管理”大类就是以课程教学资料为构建资源的。另外,课程教学讲义可以用来帮助确定子类划分标准。来自同一行业的外部资源也有必要,比如 IFLA 的“图书馆学情报学专业教育指南”,在确定 Information Studies Taxonomy 的基本大类时就很有帮助。表 1 总结了各种资源在主题分面编制中的作用。

机构内部资源的选择要尽量全面覆盖用户活动,以保证全面调查用户的信息兴趣主题和确定合理的基本大类。Information Studies Taxonomy 主题分面的编制收集了院系教学科研活动中涉及的各种资料,如教学大纲、教学讲义、项目申报报告、开题报告、学位论文、教师学术成果、学院年度研究报告以及院系网站内容等。

分类法与叙词表的选择除了考虑学科范围一致性以外,多个类表和词表的互相补充也是有益处的。一是可以帮助覆盖 Taxonomy 的主题范围,如上文所述,单个杜威法并不能覆盖所有的基本大类;二是多部叙词表可以提供多个子目候选词。在类表和词表的形式方面,印刷版词表易于使用,电子版词表更新快,反映新概念、新名词。

表1 各种资源在主题分面编制中的作用

资源类型		作用
杜威法	等级结构	确定上位类目
	相关索引	提供子目候选词
叙词表	范畴索引	确定上位类目
	词间等级关系	确定下位类目
	词间等同关系	建立子目替代关系
	叙词	提供子目候选词
机构资源	调查用户信息兴趣主题	
	确定基本大类	
	确定子类划分标准	
	编制分类法、叙词表覆盖范围以外的类目体系	
行业资源	确定基本大类	

3.2 编制方法

Information Studies Taxonomy 的编制步骤上文已有阐述,在正式开始编制 Taxonomy 类表以前,实施机构的调研和根据需求规划、设计类表至为关键,这两个步骤保证 Taxonomy 类表的准确定位和实用性。Information Studies Taxonomy 在这两个步骤进行了用户访谈和机构内知识资源的调查工作。主题分面的编制按基本大类、子类、子目依次进行,子类的编制按上位类目、划分标准、下位类目和匹配次序进行。在

子类编制以前,先建立子目候选词词库。

由于杜威法与三部叙词表作为构建资源库是不充分的,同时机构资源在主题分面的编制中也起着必不可少的作用,所以主题分面的编制集成利用了各种资源的作用。基本大类的确定前文已述,主要考虑基本大类之间的内在联系、逻辑性和用户的接受程度。子类中上位类目的确定方法是先从杜威法相关类目或叙词表范畴索引中挑选和汇聚概念,然后审阅挑选出来的类目或主题词能否涵盖所要收录的子目候选词,如不能,则从候选词中补充上位类目。比如“信息与知识组织”大类中的一级类目“资源描述(Resource Description)”就是从子目候选词中补充来的,用来包含元数据等下位类目,杜威法 020 类和所选三部叙词表都没有收录该概念。同理,下位类目也是从各个资源中挑选、汇集而成,然后根据采纳的划分标准最后确定。比如上文提到的“人工智能(artificial intelligence)”子类的子目是由杜威法 006.3 类目和 ASIS&T 叙词表的主题词汇集而成。划分标准的采纳考虑了用户视角,以教学讲义中的知识结构为参考,如前所述。主题分面采纳了三个划分标准,表 2 例举了采用这三个标准的子类。子类类目的划分不受数量限制,可同时采用多个划分标准,根据类目扩展的需要而定,也就是通常所说的“多重列类”。

表2 三个划分标准示例

种属关系	方面关系	流程关系
Libraries	Collection management	Expert systems
• Academic libraries	• Collection development	• Knowledge acquisition
• Government libraries	• Collection maintenance	• Knowledge representation
• National libraries	• Collection preservation	• Reasoning
• Public libraries	• Collection measurement	
• School libraries		
• Special libraries		
• Subject libraries		

下位类目的匹配尽量符合用户视角,同样以教学讲义中隐含的知识结构为参考。比如作为信息学系研究生培养专业的知识管理,在这里应更多地偏重于企业知识管理,而 ASIS&T 叙词表将其匹配到“信息科学”概念下,LISA 叙词表将其匹配到“知识”概念

下,显然都不符合面向用户的视角。参照用来解决下位类目可匹配到多处的矛盾。类目体系的扩展考虑了用户的浏览认知能力限度,类目大小,即横向类目与纵向等级数量以“10×4”为参照标准<sup>[7]</sup>。

子目类名依照上文所述的准确性、简单性和一致

## 中科院国家科学图书馆招聘高级人才

中国科学院国家科学图书馆为了提升整体服务和发展能力,加强人才队伍建设,现面向国内外诚招“百人计划”和“优秀人才计划”高级人才。“百人计划”主要研究方向为数字图书馆技术、战略情报分析,涉及数字化加工和呈现技术,信息检索、抽取、分析、可视化和海量信息存储技术,Web2.0与数字图书馆服务技术,以及科技政策与战略、规划研究,国际科技创新能力和竞争态势研究等。“优秀人才计划”主要研究方向为战略情报深度分析、数字图书馆关键应用、数字知识环境新兴服务等领域,主要涉及战略情报分析、数字知识环境建设、数字图书馆建设、数字信息资源长期保存、科技信息开放获取服务、知识产权保护、数字化学术出版与交流等。国家科学图书馆将为入选者提供优良的科研条件及待遇。详情请见国家科学图书馆主页(<http://www.las.ac.cn>)服务公告栏目。(中国科学院国家科学图书馆)

性的三条指导准则确定。首先,由于杜威法以学科立类,类目词带有专业特色,词汇形式又未经控制,需要对来自它的子目类名进行修改,比如用“互联网和内联网(Internet and Intranet)”取代“网络(Networks)”等。其次,来自叙词表的叙词为单元词,缺乏语境,在确定为子目类名前需根据类目语境进行修改,比如将“外包(Outsourcing)”改为“编目外包(Cataloging outsourcing)”等。第三,在没有合适的候选词的情况下,一部分中间类目名称需要自拟,如“信息系统类型(Types of information systems)”等。

### 4 结语

本研究虽然是在英文语境下进行,所运用的 Taxonomy 编制方法与思路对中文环境下机构 Taxonomy 的编制是有借鉴意义的,研究方法与研究结论对于中文环境下用传统分类法与叙词表编制机构 Taxonomy 课题的研究也能起推动作用。同时,本研究属于个案研究,它用一部综合性分类法和三部叙词表编制了一个学术机构 Taxonomy,所得出的研究结论只在一定范围内有效,如要进行广泛推广,还需进一步的扩展和更为充分的论证。此外,本研究还包括类表的用户评估,该项工作未在文中阐述。

### 参考文献:

- [1] 王忠红.用传统分类法与叙词表编制 Taxonomy 的理论可行性研究[J].大学图书馆学报,2007(4):36-40.
- [2] McGregor, B. Constructing a concise medical taxonomy [J]. Journal of the Medical Library Association, 2005, 93(1): 121-123.
- [3] Saeed, H. & Chaudhry, A. S. Using Dewey Decimal

Classification Scheme (DDC) for building taxonomies for knowledge organization [J]. Journal of Documentation, 2002, 58(5): 575-84.

- [4] Bertolucci, K. Happiness is taxonomy: four structures for snoopy[J]. Information outlook, 2003, 7(3): 36-44.
- [5] Redmond-Neal, A. & Hlava, M. M. K. Thesaurus of information science, technology, and librarianship[M]. 3rd ed. Medford, New Jersey: Information Today, 2005.
- [6] IFLANET, Guidelines for Professional Library/Information Educational Programs-2000 [EB/OL]. [2006-11-03]. <http://www.ifla.org/VII/s23/bulletin/guidelines.htm>.
- [7] Delphi Group. Information Intelligence: content classification and the enterprise taxonomy practice [R/OL]. [2005-10-10]. <http://www.delphigroup.com/research/whitepapers/20040601-taxonomy-WP.pdf>.

**王忠红** 新加坡南洋理工大学黄金辉传播与信息学院博士生。通讯地址: Division of Information Studies, Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore. 邮编 637718。

**Abdus Sattar Chaudhry** 新加坡南洋理工大学黄金辉传播与信息学院教授、系主任、博士生导师。通讯地址同上。

**Christopher S. G. Khoo** 新加坡南洋理工大学黄金辉传播与信息学院教授、副教授、博士生导师。通讯地址同上。

(收稿日期:2008-05-15)