

“英文超级科技词表”农业综合范畴体系构建初探*

孙巍, 张学福, 潘淑春, 苏晓路, 刘家益
(中国农业科学院农业信息研究所, 北京 100081)

摘要: 英文超级科技词表是一个包含三层结构的知识组织体系, 范畴体系为其顶层结构, 起到类聚概念、表达词表宏观知识结构的作用。文章以农业领域为例, 探索构建“英文超级科技词表”农业综合范畴体系的思路。首先, 分别对所遴选的主干、辅助范畴体系进行结构特征比较分析; 其次, 基于辅助范畴表对农业领域范畴体系进行初步构建; 最后, 分别将DDC、LCC与所构建的农业专业范畴体系进行比对分析, 进而为农业综合范畴体系构建提出指导性建议。

关键词: 范畴体系构建; DDC; LCC; 农业范畴体系; 类目映射

中图分类号: G254

DOI: 10.3772/j.issn.1673—2286.2014.12.002

1 引言

“英文超级科技词表”(以下简称“英表”)是“十二五”国家科技支撑计划项目“面向外文科技文献信息的知识组织体系建设与应用示范”的主要研制内容之一, 该项目由国家科技图书文献中心(NSTL)牵头, 由理、工、农、医四大领域相关机构专家分工协作共同完成。

“英表”并非传统意义上的叙词表, 从逻辑结构上讲, 它是一个具有三层结构的科技知识组织体系, 自下而上依次为基础词库、概念网络、范畴体系。基础词库层是将不同来源异构的词汇集, 按一定规范进行描述, 并采用统一格式进行存储而形成的词汇元数据仓储; 对基础词库层中的词汇进行同义词归并, 形成以概念为单位的同义词群, 进而构成孤立无序的概念网络; 范畴体系为概念提供分类框架, 以此对无序概念进行分类类聚, 在一定程度上弥补概念网络在宏观知识结构表达上的不足^[1]。

本文正是为了满足课题中“英表”范畴体系构建的迫切需求, 在遵循理、工、农、医多领域机构协作共建“英表”范畴体系的总体原则基础上, 以农业领域为例, 分析与探索自主构建农业综合范畴体系的思路, 为“英表”农业综合范畴体系自主构建工作提供理论依据与实践指导。

2 “英表”农业综合范畴体系的功能定位

“英表”农业综合范畴体系自主构建是“英表”范畴体系构建工作的重要组成部分之一, 对NSTL农业文献资源的有效组织、NSTL农业英文文献信息服务能力的提升具有重要研究意义。

从功能上讲, “英表”农业综合范畴体系应具备从学科主题角度均衡合理分类与汇聚农业领域概念群、清晰完整地表达农业领域词表结构的功能; 承接农业概念与本体, 实现农业通用本体维(粒)度的有效控制, 便于建立本体与概念的映射关系; 充分考虑英语语

* 本研究得到国家“十二五”科技支撑计划“面向外文科技文献信息的知识组织体系建设与应用研究”(编号: 2011BAH10B00)资助。

言地区的政治、经济、文化背景,以及其主要服务对象国内用户对学科体系的理解,使所构建的范畴体系既能匹配英语语言地区人们对学科知识结构的理解,又符合国内用户的使用习惯。

3 分类体系及其结构特征分析

以主干分类体系为主,充分继承主干分类体系类目关系,以各专业领域分类体系为辅,对主干分类体系进行适当扩充与调整,是“英表”范畴体系协作共建的总体思路。因此,主干分类体系的适用性与专业辅助分类体系遴选是领域范畴体系自主构建工作优先考虑的问题。

3.1 主干分类体系及其结构特征分析

《杜威十进分类法》(Dewey Decimal Classification,以下简称DDC)^[2]、《美国国会图书馆分类法》(Library of Congress Classification,以下简称LCC)是英语地区影响很大的综合分类法,这两部分类法都有专设的维护机构持续进行维护和修订^[3],是选择主干分类体系的主要考虑对象。

(1) DDC

DDC于1976年首次出版电子版及印版,是一部通用分类法,系统性强,应用较广泛,目前已被全球超过135个国家的图书馆使用^[4],且被翻译逾30种语言版本。从其类号体制看,DDC是十进制分类体系,其各级类目基本按层累计方式编号,类目体系等级分明,可读性强,易于理解和使用;从其结构特征看,DDC类目分布较均衡,但其作为分类表,类间关系的表现力不如叙词表那样明显;从其更新与维护角度看,DDC设有专门的维护机构持续对其进行维护和修订,一直处于不断的更新与完善中,具有较好的可扩展性;从其应用范围上看,DDC除被广泛地应用于标注文献外,也能用来组织网际网络上的各种资源。

(2) LCC

LCC发展于十九世纪末二十世纪初,最初用于美国国会图书馆图书的整理、分类的分类体系,是专为具体图书馆编制的分类法,但是后来逐渐为美国国内许多图书馆所采用,成为一种普通的分类法。从其类号体制看,LCC是列举式分类体系,类目编号是一种完全的序数制,即两位字母加数字的编码方式,基本不能反映类目的从属等级关系,可读性较差^[5-6];从其编制体

制看,LCC各分册由各学科专家分别进行编制和修订,没有统一的编制规则,缺乏明确的分类理论指导,系统性较差^[7];从其结构特征看,其类目细分的程度详细不等,基本上是取决于国会图书馆的藏书数量和内容,类目分布均衡性较差;从其更新与维护角度看,采取周更新,分类体系在学科层面具有一定的扩展性。

3.2 辅助分类体系遴选与特征比较分析

3.2.1 辅助分类体系遴选

从收词规模、涉及领域、词表结构、概念分布等方面对搜集的38个农业相关词表进行了深入分析,层层筛选,最终选取了两大农业综合词表NAL^[8]和AGROVOC^[9]作为辅助分类体系。以何种方式方法,充分高效地利用辅助分类体系对主干体系进行有益补充与扩展,还有待于对这两大农业综合词表进行进一步的特征比较分析。

3.2.2 特征比较分析

采用多个领域范畴体系对主干范畴体系进行补充与扩展,操作方式大体分为两种。其一,选取一个主干范畴体系,整合多个领域范畴体系,将整合后的领域范畴体系与主干体系进行类目映射,完成主干范畴体系的补充与扩展;其二,选取一个主干范畴体系,多个领域范畴体系分别依次与主干范畴体系进行类目映射,逐步完成对主干范畴体系的补充与扩展。从词表特征角度讲,第一种方式对领域范畴体系的系统性要求相对较高,至少存在一个系统性较高的领域范畴体系,能够发挥领域主干范畴体系的作用。从领域综合范畴体系构建的工作量角度讲,第二种方式工作步骤相对复杂繁琐。

为了探索农业领域范畴体系构建的总体思路,从编制机构、用途、学科主题涵盖性、等级性、均衡性、词间关系表现力、可扩展性、可读性多方面对遴选的NAL和AGROVOC两大农业综合词表进行深入比较分析,分析结果如表1。从表中可以看出,相对于NAL,AGROVOC在等级性和均衡性上稍有逊色,NAL的系统性较强,可发挥领域主干范畴体系的作用;从学科主题涵盖性角度讲,两个表各有特色,互为补充。进而考虑采纳上述第一种构建方式,将两大专业词表进行整合作为农业综合范畴体系构建的辅助范畴体系。

表1 NAL与AGROVOC特征比较分析

特征	NAL	AGROVOC
编制机构	美国国家图书馆	国际粮农组织
用途	农业信息索引及提高农业信息的检索效率	标引或检索信息资源
学科主题涵盖性	学科覆盖面广	既有主题分类, 也有较粗的学科分类
等级性	NAL共有17大类, 分类等级最高达14级。从整体上看, 词表等级性较强	等级性较差
均衡性	从整体上看, 各领域的概念等级呈正态分布, 概念分布均衡性能较好	类目等级分布不均
词间关系表现力	主要包含三种词间关系类型: 层级关系(即属分关系) Hierarchical relationships、等同关系(即用代关系) Equivalence relationships、关联关系(即参照关系) Associative relationships	主要类间关系包括BT(广义词)、NT(狭义词)、RT(相关词)、UF(非叙词)
可扩展性	每年更新和修订	每3个月维护更新一次
可读性	通过“USE(用)”和“USED FOR(代)”来指定非叙词, 并使用斜体字, 以提醒用户。涉及词表中的其他词时, 运用范围注释Scope Notes来阐明词的含义与用法。同时还为词表提供了不同的定义, 如词的来源等	AGROVOC对一些术语进行了范畴注释, 对其概念含义进行了定义。对分类学的术语和地理学的术语进行了标注, 使得这些词的查找和下载非常容易

4 农业领域范畴体系整合

相对于NAL, AGROVOC在等级性和均衡性上相对较差, 我们考虑对在其同一范畴体系下作相应类目补充。AGRIS^[10]数据库的资源是基于AGROVOC分类体系, 通过资源对分类体系进行调整所得, 主要用于AGRIS文献库资源的组织与检索。利用AGRIS范畴对AGROVOC进行补充, 在一定程度上弥补了AGROVOC学科分布较粗的缺陷。

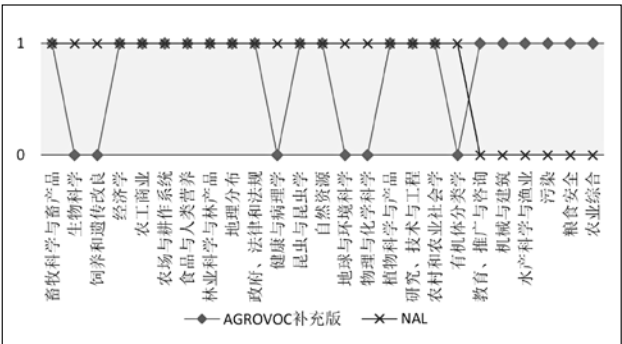


图1 AGROVOC补充版与NAL一级类目综合分析图

为了探索农业综合范畴体系构建的总体思路, 初步对AGROVOC(补充版)与NAL的一级大类进行了比对分析与整合。

从AGROVOC(补充版)与NAL的一级类目比较分析结果(如图1)可以看出, NAL的19个一级大类中, 与AGROVOC相吻合的有13个大类, 不吻合的类目包括生物科学、地球与环境科学、物理与化学科学、健康与病理学、有机体分类学、饲养与遗传改良。其中, 前五个类目部分需归入理、工、医科学中, 第六个类目属于农学类目。从另一角度看, FAO的19个一级大类中, 有六大类是NAL所缺失的, 即教育、推广与咨询, 机械与建筑, 水产科学与渔业, 污染, 粮食安全, 农业综合。其中, 污染类目可归为NAL的环境科学类目中, 其他类目可考虑作为NAL的补充类目。综合分析后初步生成了一个农业领域一级范畴(如表2)。

5 主干分类体系与农业辅助范畴体系类目比较分析

主干、辅助范畴体系的类目比对分析是为了进一步考察主干分类体系对农业领域范畴体系构建的适用性。本文重点从主干分类体系的专业领域涵盖性以及专业辅助分类体系对主干分类体系的补充性两方面, 分别将DDC、LCC的一级类目与农业辅助范畴体系(整合后的农业领域一级范畴)进行类目比对分析, 考察DDC、LCC的主干分类体系适用性。

表2 农业领域一级范畴

一级类	下级类来源表
农业-综合	AGRIS
畜牧科学与畜产品	NAL
生物科学	NAL
饲养和遗传改良	NAL
经济学	NAL
农工商业	NAL
农场与耕作系统	NAL
食品与人类营养	NAL
林业科学与林产品	NAL
地理分布	NAL
政府、法律和法规	NAL
健康与病理学	NAL
昆虫与昆虫学	NAL
自然资源	NAL
地球与环境科学	NAL
物理与化学科学	NAL
植物科学与产品	NAL
研究、技术与工程	NAL
农村和农业社会学	NAL
教育、推广与咨询	AGRIS、AGROVOC
机械与建筑	AGRIS
水产科学与渔业	AGRIS
粮食安全	AGROVOC

5.1 DDC与农业辅助范畴体系类目比较分析

DDC的分类体系并非学科分类体系,很难从学科的角度与农业辅助范畴体系相比较,其农业学科类目主要分布在二级类目“63农业技术”下,且主要是从农业技术角度列类,其他农业学科类目的分布较分散,因此,考察DDC对农业学科领域类目的涵盖性以及农业辅助范畴体系类目对DDC的补充性,仅能以农业辅助范畴体系为主,从字面上将DDC类目与农业辅助范畴体系类目进行一对一或者多对一比对,本文仅将上文构建的农业辅助范畴体系一级类目与DDC进行类目比对分析,初步考察DDC的主干范畴体系适用性。

由表3和图2可以看出,DDC的二级分类中,与23个农业辅助范畴体系一级类目基本相关的类目有12个,参

与匹配的DDC类目有16个;DDC的三级类目中,与23个农业辅助范畴体系一级类目基本相关的类目有13个,参与匹配的DDC类目多达39个,说明DDC三级分类中,多数是以某个分支或者某几个分支的形式与农业辅助范畴体系的某个一级类目相对应,DDC的农业类目涵盖面相对较小。DDC四级分类中此种现象更为明显。

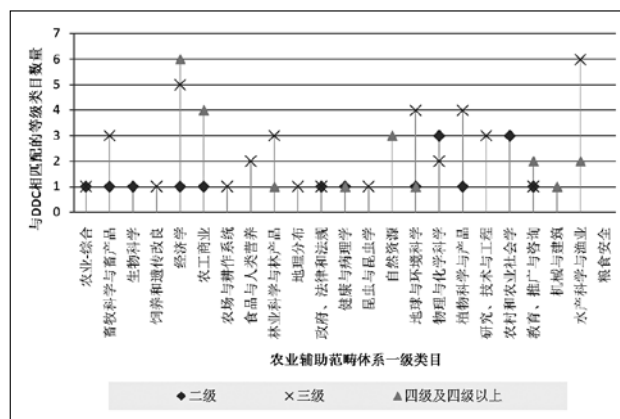


图2 农业辅助范畴体系（一级）与DDC的
类目比对分析图

进一步分析发现,即便从字面上将DDC与农业辅助范畴体系进行粗略比对,所对应的也只是类目的某个分支或者某个侧面,无法完全涵盖相对应的类目;DDC三级、四级类目中所匹配的类目相对具体些,但分类等级和维度与农业辅助范畴分类体系类目的等级差别较大,有些DDC三级、四级或者四级以上类目甚至是跨越了多个农业辅助范畴体系的一级类目。由此可见,DDC的农业领域范畴类目的涵盖性较差,利用DDC构建农业综合范畴体系将是一个费时、费力且不容易实现的工作。

5.2 LCC与农业辅助范畴体系类目比较分析

LCC是学科分类体系,可以从学科角度将其与农业辅助范畴体系类目进行比对。这里遴选了LCC一二级分类中的农学类目,对LCC一二级农学类目与农业辅助范畴体系一级类目进行类目匹配,匹配结果如表4。

为了考察LCC的农业领域涵盖性,以农业辅助范畴体系为主,将LCC的一级类目与其相匹配,匹配结果如图3,从图3可以看出,LCC的农业一二级类目涵盖了农业辅助范畴体系一级类目23个类目中的15大类,生物科学、食品与人类营养、健康与病理学、自然资源、地

球与环境科学、农村和农业社会学、机械与建筑、粮食安全8个类目没有涵盖其中, 涵盖率相对较高。相反, 为了考察农业辅助范畴体系对LCC的补充扩展性, 统计了LCC一二级农业各类目所匹配的农业辅助分类体系类

目数(如图4), 从图4中可以看出, 以LCC一级类目为主, 农业辅助范畴类目多数分布在LCC的S农业综合类目中, 其次是SF动物饲养类目中, 而没有与SK狩猎相匹配的类目。

表3 农业辅助范畴体系(一级)与DDC的类目比对表

农业综合范畴体系 (一级)	DDC		
	二级	三级	四级及四级以上
农业综合	630 农业与相关技术	111 本体	-
畜牧科学与畜产品	590 动物科学	598 鸟、599 哺乳动物、636 畜牧业	-
生物科学	570 生命科学与生物学	-	-
饲养和遗传改良	-	576 遗传与遗传改良	-
经济学	330 经济	331 劳物经济、332 财政经济、333 土地与能源经济、334 合作社、338 生产	332.6441 农产品贸易
农工商业	380 贸易、通信与运输	-	344.042 产品控制、363.19 产品安全、338.1 农业生产、338.10973 农工业
农场与耕作系统	-	633 大田作物	-
食品与人类营养	-	637 乳品加工及产品、641 食品与饮品	-
林业科学与林产品	-	634 果园、水果和林业、635 园艺、715 灌木种植	338.17498 林产品
地理分布	-	911 历史与地理	-
政府、法律和法规	340 法律	344 劳动、社交、教育与文化法规	-
健康与病理学	610 医药与健康	-	616.07 病理学(包括动植物和人)
昆虫与昆虫学	-	638 昆虫学	-
自然资源	-	-	333.7-333.9 自然资源与能源
地球与环境科学	550 地球科学	389 计量与标准化、551 地理、水文与气象、577 生态学、912 地图、绘制与设计	25.0655 地球科学-信息系统
物理与化学科学	530 物理、540 化学、660 化学工程	547 有机化学、572 生物化学	-
植物科学与产品	580 植物及植物学	582 特色植物与花卉、586 无籽植物、587 无籽导管植物、632 植物灾害与病虫害	-
研究、技术与工程	-	627 水利工程、631 技术设备与材料、664 食品技术	-
农村和农业社会学	300 社会学、710 景观美化与区划、640 农村与家政	-	-
教育、推广与咨询	370 教育	907 教育、科研与相关主题	352.74 知识传播与推广、352.743 调查与咨询机构
机械与建筑	-	-	631.3 农业工具、机器、器械、装备
水产科学与渔业	-	593 海洋与海岸无脊椎动物、594 软体无脊椎动物、595 节肢动物、596 脊椎动物、597 冷血脊椎动物、鱼、639 狩猎、渔业和维护	639.8 水产业、338.37 水产品
粮食安全	-	-	-

表4 农业辅助分类体系（一级）与LCC（农业一、二级）的类目比对表

LCC农业二级类目	农业辅助分类体系（一级）
S 农业综合	农业-综合、政府、法律和法规，地理分布，农工商业，经济学，物理与化学科学，研究、技术与工程，教育、推广与咨询
SB 作物栽培	农场与耕作系统、植物科学与产品
SD 林业	林业科学与林产品
SF 动物饲养	畜牧科学与畜产品、饲养和遗传改良、昆虫与昆虫学
SH 水产业、渔业和钓鱼业	水产科学与渔业
SK 狩猎	-
-	食品与人类营养
-	健康与病理学
-	粮食安全
-	农村和农业社会学
-	机械与建筑
-	生物科学
-	自然资源
-	地球与环境科学

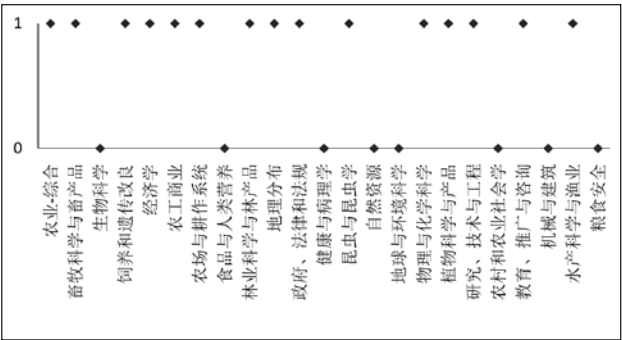


图3 LCC农业二级类目与农业辅助范畴体系比对应图

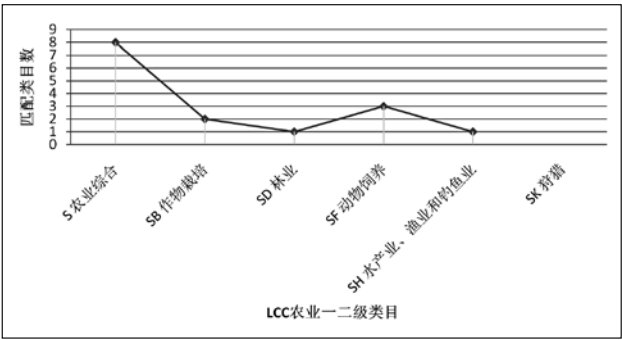


图4 LCC农业一二级类目与农业辅助范畴体系类目比对应统计图

5.3 分析结论

综上，与DDC相比，LCC的农业领域涵盖性能较好，农业辅助范畴体系对LCC的补充扩展性也优于DDC，尽管从整体原则上讲DDC和LCC均具备了充当主干范畴体系的特性，但针对农业学科领域，LCC更适合作为农业综合范畴体系构建的主干范畴体系。

6 结语与展望

本文以农业领域为例，通过对所遴选的主干、辅助范畴体系的结构特征比较分析，农业辅助范畴体系比

对整合，以及DDC、LCC与农业辅助范畴体系类目的比较分析，得出以下构建“英文超级科技词表”农业综合范畴体系的指导性建议。

（1）相对于DDC，LCC更适合作为“英表”农业范畴体系构建的主干范畴体系；

遴选主干范畴体系是“英表”范畴体系构建工作的首要考虑问题之一。遴选过程中需要权衡理、工、农、医、通用、综合各领域的实际构建需求，最终选取一个综合性分类体系作为“英表”范畴体系构建的主干范畴体系。构建初期需要各领域专家分别对主干范畴体系的适用性进行调研与分析。本文的综合分析结果表明，从DDC、LCC的自身特征角度讲，DDC、LCC

均可以作为农业综合范畴体系构建的主干范畴体系, 只是工作量上存在差别, LCC相对于DDC, 工作量相对较小; 而从DDC、LCC的结构特征以及农学学科特征角度讲, LCC的学科性更强, 更适合作为主干范畴分类体系。

(2) 以整合的农业学科范畴体系的方式构建农业辅助范畴体系, 进而利用农业辅助范畴体系对主干范畴体系进行补充与扩展的方法更适合构建“英表”农业综合范畴体系。

本文提出了多个领域范畴体系对主干范畴体系进行补充与扩展的两种操作方式。通过对农业领域范畴体系的特征比较分析以及两种方式的操作条件分析得出, 所遴选的农业领域范畴体系满足第一种方式的操作条件, 且第一种方式的工作量相对较小, 构建的“英表”农业范畴体系系统性较强。因此, 先整合领域范畴体系, 再进行主辅范畴类目映射的操作方式是自主构建“英表”农业范畴体系的首选操作方式。

“英表”农业综合范畴体系的构建工作才刚刚开始, “英表”农业综合范畴体系构建的具体方法步骤、范畴体系互操作具体细则等是“英表”农业综合范畴体系构建工作的重点, 是需要深入研究的问题。

参考文献

- [1] 吴雯娜, 王星. 基于DDC的《英文超级科技词表》范畴体系构建研究: 以工程技术为例[J]. 图书情报工作, 2011, 55(22): 15-21.
- [2] WebDewey [EB/OL]. [2014-11-20]. <http://connexion.oclc.org>.
- [3] 马张华. 国外文献分类法修订维护的发展及对《中图法》的启示[J]. 国家图书馆学报, 2008(2): 40-44.
- [4] OCLC. Organize your materials with the world's most widely used library classification system [EB/OL]. [2014-11-29]. <http://www.oclc.org/dewey.en.html>.
- [5] 周宁. 信息组织[M]. 2版. 武汉: 武汉大学出版社, 2004: 52-60.
- [6] 曹树金, 罗春荣. 信息组织的分类法与主题法[M]. 北京: 北京图书馆出版社, 2000: 440-446.
- [7] HOLLEY R P. Classification in the USA [J]. International Classification, 1986, 13(2): 73-78.
- [8] USDA. NAL Agricultural Thesaurus [EB/OL]. [2014-11-29]. http://agclass.nal.usda.gov/dne/search_sc.shtml.
- [9] FAO. AGROVOC [OL]. [2014-11-29]. <http://aims.fao.org/agrovoc#VHuyRtlYByI>.
- [10] FAO. AGRIS category [EB/OL]. [2012-10-11]. <http://agris.fao.org/agris-search/index.do>.

作者简介

孙巍, 女, 1978年生, 中国农业科学院农业信息研究所副研究员, 研究方向: 农业知识组织与可视化分析, E-mail: sunwei@caas.cn.

张学福, 男, 1966年生, 中国农业科学院农业信息研究所研究员, 研究方向: 农业知识组织与可视化分析, 通讯作者, E-mail: zhangxf@caas.cn.

An Exploration of Construction on STEST

SUN Wei, ZHANG XueFu, PAN ShuChun, SU XiaoLu, LIU JiaYi

(Institute of Agricultural Information, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: “Science & Technology English Super-thesaurus” (STEST) is a knowledge organization system consisting of three-layer structure. There is a category system at the top layer to gather concepts and display the macrostructure of STEST. In the paper, ideas are explored for constructing the agricultural comprehensive category system. Firstly, we analyze comparatively the structure characteristics of the selected backbone category systems, and those of the auxiliary category systems. Secondly, based on the auxiliary category systems, an agricultural category system is constructed preliminarily. Finally, we compare separately DDC and LCC to the constructed agricultural category system by means of analysis, and then some guidance and advice to the construction of agricultural comprehensive system are proposed.

Keywords: Category system; DDC; LCC; Agricultural category system; Category mapping

(收稿日期: 2014-11-20)