DDC与UDC对比分析——以工程学科为例*

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

摘要：为了更深入理解DDC与UDC的差异，设计出更优质的知识组织系统（Knowledge Organization System），文章采用领域专家分析和量化对比方法，对最新版本的DDC和UDC工程相关范畴类目进行对比分析。分析结果表明：①DDC与UDC的知识面覆盖重合度较高；② DDC范畴类目划分整体上较UDC更为细致；③DDC与UDC范畴类目设置存在较大差异；④DDC与UDC的差异主要是由两者知识描述角度的差异造成的。

关键词：知识组织系统；DDC；UDC
中图分类号：G254
DOI：10.3772/j.issn.1673—2286.2014.11.007

1 背景

《杜威十进图书分类法》（Dewey Decimal Classification，DDC[1]）由美国图书馆学家麦尔威•杜威于1876年发明，对世界图书馆分类学有相当大的影响，已翻译成西班牙文、中文、法文、挪威文、土耳其文、日文、僧伽罗文、葡萄牙文、泰文等出版，目前被全球超过130个国家的20000余个图书馆所使用[2]。

《国际十进分类法》（Universal Decimal Classification，UDC[3]），又称为通用十进制分类法，是世界上规模最大、用户最多、影响最广泛的文献资料分类法之一。自1899-1905年比利时学者奥特勒和拉封丹共同主编、出版UDC法文第一版以来，现已有20多种语言的各种详略版本。近百年来，UDC已被世界上几十个国家的10多万个图书馆和情报机构采用。UDC目前已成为名副其实的国际通用文献分类法[4]。

作为最权威最有影响力的两种知识组织系统，DDC和UDC不仅为知识服务工作者提供了有力工具，也为知识组织系统（KOS）开发人员提供了很好的借鉴参考工具。DDC与UDC之所以能够相互独立存在，在于两者存在差别。找出两者差别，对深入理解DDC、UDC以及知识组织系统开发有着重要的理论和实践指导意义。

2 原则与方法

2.1 范畴类目对比分析的原则

本文所谓范畴类目，是指知识组织系统中的一个结点，包括一个范畴号（class notation）以及范畴号对应的范畴标题（class caption）。DDC与UDC均主要由范畴类目组成。对DDC/UDC作对比分析，实际上是对其范畴类目体系作对比分析。DDC与UDC范畴类目体系均采用层级的知识结构，下级范畴类目是对其上级范畴类目的进一步细分。范畴号标识了范畴类目在整个知识组织系统中的层级位置。范畴标题则是对范畴类

* 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。

目的文字性描述。对范畴类目作对比分析时，不仅要考虑标题相似性，还要考虑范畴类目所处层级位置。

在对比分析DDC与UDC范畴类目时，本文遵循如下原则：

（1）范畴类目的层级位置和标题共同决定了范畴类目的含义。含义的相似程度决定了范畴类目的相似程度。

（2）范畴类目间的比较通常只在同级别内进行，即顶级范畴类目与顶级范畴类目对比，二级范畴类目与二级范畴类目对比，三级范畴类目与三级范畴类目对比。特殊情况除外。

需要说明的是，本文所称范畴类目级别如顶级、二级、三级等，不是DDC或UDC范畴类目体系中的绝对层级，而是相对级别，其中顶级范畴类目是指某个领域（如农业学科）知识的初次划分，二级范畴类目是指在顶级范畴类目基础上对某个领域知识的二次划分二次划分，依此类推。

根据上述原则，本文定义范畴类目相似的三个程度：

匹配：即两个范畴类目具有较高相似度，通常是指范畴类目含义近乎一致；

不匹配：即两个范畴类目具有较低相似度，通常是指范畴类目含义完全没有交叉重叠；

部分匹配：即两个范畴类目的相似度介于匹配与不匹配之间，通常是范畴类目含义有部分重叠。

另外，本文定义了术语“不匹配率”，用来衡量两个分类体系的差异程度。所谓不匹配率，是指不匹配的范畴类目数占范畴类目总数的比率。由于本文是对DDC与UDC的对比分析，因此，计算不匹配率时，不匹配范畴类目数是同一层级DDC的不匹配范畴类目数与UDC的不匹配范畴类目数之和，范畴类目总数是同一层级DDC与UDC的范畴类目数之和。

2.2 范畴类目对比分析的方法

遵照上述比较原则，本文设计了如下范畴类目对比分析的方法：

（1）顶级范畴类目匹配。选定某个领域，对比两个范畴类目体系的顶级范畴类目，形成DDC与UDC顶级范畴类目映射对（即两个可以匹配或部分匹配的顶级范畴类目，一个来自DDC，一个来自UDC），统计可以匹配、部分匹配和不匹配的DDC顶级范畴类目数和UDC顶级范畴类目数，计算不匹配率。

（2）二级范畴类目匹配。对每个顶级范畴类目映射对，对比它们下面的二级范畴类目，形成二级范畴类目映射对，统计匹配、部分匹配和不匹配的DDC二级范畴类目数和UDC二级范畴类目数，计算不匹配率。

（3）三级范畴类目匹配。对每一个二级范畴类目映射对，对比它们的三级范畴类目，统计匹配、部分匹配和不匹配的DDC三级范畴类目数和UDC三级范畴类目数，计算不匹配率。

3 数据预处理

数据源的选择。DDC与UDC均有不同语言的版本，而英文版应用最广泛。为统一语言，便于对比，本文选择最新的出版于2011年的第23版完整英文版DDC和最新的更新日期是2003年的英文版UDC[5]作为对比分析对象。为便于操作，本文对一些范畴标题进行了简单翻译，但是对比分析时仍以英文范畴标题为主要依据。不失一般性，本文选取了工程学科的前三级范畴类目作为对比分析对象。

层级校准。DDC与UDC的范畴类目体系结构具有一定差异。例如，存在着一些相匹配的范畴类目，在DDC中可能是顶级范畴类目，而在UDC中却是二级甚至是三级范畴类目。为了使对比分析更具可操作性，必须使两者中相匹配的范畴类目尽量处于相同的层级，但是同时还要考虑不破坏范畴类目体系本身的结构性，这就必须对一些范畴类目的层级进行调整。本文采用的方法是，由领域专家对DDC和UDC类目进行分析判断，把其中可作为工程学科顶级划分但不处于顶级的范畴类目提升为顶级范畴类目。被提升的范畴类目的所有下级范畴类目均顺次提升相同级数，以保持原有范畴类目的体系结构相对稳定。通过校准后，DDC有22个顶级工科范畴类目，UDC有30个工科顶级范畴类目。至此，DDC与UDC对工程学科知识的划分已基本处于相同层级，具有了较大的可操作性。

4 范畴类目对比分析

根据上文提出的对比分析原则和方法，经过数据预处理后，对DDC与UDC工程学科范畴类目进行对比。

虽然进行了层级校准，在对比分析过程中，仍然出现了极少量跨级匹配的情况。本文的处理方法是，在统计时，将该范畴类目的匹配结果分别记录在对应级别

的范畴类目匹配结果里。例如某DDC二级范畴类目与某UDC三级范畴类目部分匹配，则DDC二级范畴类目“部分匹配数”加一，同时，UDC三级范畴类目“部分匹配数”加一。

对比分析结束后，对各级范畴类目数量进行核查校验，发现各类型的范畴类目数量与范畴类目总数一致，证明数据统计无误。

4.1 顶级范畴类目对比分析

顶级范畴类目是对某个领域知识的初始划分，体现了知识组织系统在该领域的知识覆盖面。通过对比分析，一DDC与UDC在工科领域的顶级范畴类目的相似情况如表1。

表1 顶级范畴类目比较结果

	DDC	UDC	总数
匹配	5	5	10
部分匹配	13	17	30
不匹配	4	8	12
不匹配率	18.18%	26.67%	23.08%

从数量上看，DDC共有22个顶级范畴类目与工科直接相关，将工科知识分为22块。UDC共有顶级范畴类目30个，将工科知识分为30块。DDC一级范畴类目少于UDC一级范畴类目，这说明，在顶级划分上，UDC比DDC更细一些。

从相似程度上看，DDC和UDC共52个顶级范畴类目中，有5个DDC与UDC范畴类目匹配，得到5个范畴类目对；13个DDC范畴类目与17个UDC范畴类目部分匹配（有若干多对一的情况），得到13个范畴类目对；4个DDC范畴类目和8个UDC范畴类目不能在对方范畴类目体系中找到匹配或部分匹配范畴类目。不匹配范畴类目数共12个，占一级范畴类目总数23.07%，这表明，有超过76%的DDC/UDC顶级范畴类目是匹配或部分匹配的，DDC与UDC顶级范畴类目具有较大的相似性。

4.2 二级范畴类目对比分析

二级范畴类目是对顶级范畴类目的细分。通过对比分析，二级范畴类目相似情况如表2所示。需说明的是，在可以匹配的二级范畴类目中，DDC的单个范畴类目（范畴号为725-728）与5个UDC范畴类目（范畴号为721、725、726、727、728）的并集匹配，计数时，这5个UDC范畴类目计为1个匹配。因此在表2中，UDC二级范畴类目总数为109。

表2 二级范畴类目比较结果

	DDC	UDC	总数
匹配	20	20	40
部分匹配	41	41	83
不匹配	91	48	139
不匹配率	59.87%	42.48%	52.45%

从数量上看，DDC共有二级范畴类目152个，UDC共有二级范畴类目113个，DDC二级范畴类目比UDC二级范畴类目多出约34%，由此可知，在二级划分上，DDC更加细致。

在二级范畴类目中，DDC与UDC有20对范畴类目可以匹配，得到20个匹配范畴类目对；有41个DDC范畴类目可以与UDC二级范畴类目（少量三级范畴类目）部分匹配，有42个UDC二级范畴类目可与UDC二级范畴类目（或三级范畴类目）部分匹配；有91个DDC范畴类目和47个UDC范畴类目无法找到匹配或部分匹配。不匹配范畴类目总数为138个，占二级范畴类目总数52.45%。可见，DDC与UDC的差异在二级范畴类目中进一步扩大，不匹配率已超过一半。

4.3 三级范畴类目对比分析

三级范畴类目是对二级范畴类目的进一步细分。三级范畴类目的对比分析结果见表3。需说明的是，在匹配过程中，有一个DDC范畴类目（005.74）与两个UDC范畴类目（004.65、004.63）的并集相匹配，在计数时，UDC的两个范畴类目的共同计数为1，导致表中UDC范畴类目总数为214个。

从数量上看，DDC共有三级范畴类目561个，UDC共有三级范畴类目215个。DDC三级范畴类目已超出UDC三级范畴类目的两倍，可知，在深层级的知识划分上，DDC比UDC要细致很多。

表3 三级范畴类目比较结果

	DDC	UDC	总数
匹配	12	13	25
部分匹配	50	54	104
不匹配	499	147	646
不匹配率	88.95%	68.37%	83.25%

在三级范畴类目中，有12对三级范畴类目可以匹配，另有一个UDC三级范畴类目可与DDC二级范畴类目匹配；有50个DDC二级范畴类目可以与UDC二级范畴类目部分匹配（存在多对一情况和少量跨级匹配情况），有54个UDC二级范畴类目可以与DDC二级范畴类目部分匹配（存在多对一情况和少量跨级匹配情况）；有499个DDC三级范畴类目和147个UDC三级范畴类目无法找到匹配范畴类目。

从上述数据中可知，不匹配的三级范畴类目数为646，占三级范畴类目总数的83.25%，可见，到了三级范畴类目，DDC与UDC的差异已经非常明显，平均有占总数80%的范畴类目无法在对方的范畴类目体系中找到。

4.4 整体对比分析

如果不考虑层级，将前三级范畴类目作为一个整体进行对比分析，则结果如表4所示。

表4 范畴类目整体比较结果

不匹配范畴类目数	796
范畴类目总数	1093
不匹配率	0.72827081

DDC与UDC前三级工科范畴类目总数共为1093个，其中无法匹配的范畴类目数为796，占到总范畴类目数的72.83%。这表明DDC与UDC的范畴类目中，无法匹配的范畴类目数接近四分之三，还是比较大的。

5 结论与展望

5.1 结论

结合上述DDC与UDC范畴类目对比分析结果，将其从工程学科推广至一般情况，本文得到如下结论：

DDC与UDC的知识覆盖面重合度较高。分析结果表明，DDC与UDC项级工科范畴类目差异较小，不匹配率仅为23.07%。顶级范畴类目的较小差异反映了两者所覆盖知识面的较小差异。

DDC范畴类目类目划分较UDC更为细致。对工科范畴类目的分析结果表明，整体上，DDC前三级范畴类目总数为735，远多于UDC的358。局部上，虽然DDC的顶级范畴类目数量少于UDC，但随着层级加深，DDC范畴类目数量有大幅增加，而UDC相对增加较小，在二级层面，DDC范畴类目数量已超过UDC，到三级层面时，DDC有三级范畴类目561个，远超UDC的215个，两者差距随层级加深有进一步扩大趋势。

DDC与UDC范畴类目设置存在较大差异。DDC与UDC虽然在知识覆盖面上有较高程度的重合，但具体的范畴类目设置差异很大。对工科范畴类目的分析结果表明，DDC与UDC的不匹配率平均为72.83%，其中顶级不匹配率为23.07%，二级范畴类目不匹配率为52.08%，三级范畴类目不匹配率为83.25%。DDC与UDC中有至少一半以上的工科范畴类目无法在对方范畴类目体系中找到匹配或部分匹配的范畴类目。

DDC与UDC的差异主要是由两者对知识描述角度的差异造成的。DDC与UDC知识覆盖面重合度较高。但是，为什么在二、三级范畴类目差异如此大？作者以经过校准后的二级工科范畴类目“交通运输”为例，分析了两者三级范畴类目不匹配的原因。经过层级校准后，DDC与UDC的交通运输及其下级范畴类目如表5、表6所示。

两者在二级范畴类目上完全匹配，但三级范畴类目却有较大差异。DDC先对交通运输的一般问题进行了划分，再分类型对交通运输进行划分，先总后分来描述交通运输相关知识；UDC则只分类型对交通运输进行了划分，分类描述交通运输相关知识。这两种不同描述角度，导致了“交通运输安全”和“交通导航”这两个不匹配三级范畴类目的产生。这种情况在一级范畴类目中也同样存在。这表明，DDC与UDC的差异，主要是由两者对知识的描述角度不同产生的。

5.2 展望

本文基于一套自定义的原则和方法，人工对DDC和UDC的范畴类目相似度进行了对比分析，得出了一些探索性结论，对于深入理解DDC、UDC和知识组织系统的设

表5 经过层级校准后的DDC交通运输及其下级范畴类目

transportation_ 交通运输	
	629.040289_Safety measures_交通运输安全
	629.045_Navigation_交通导航
	385_*Railroad transportation_铁路运输
	N_Water transportation_水运
	N_Air, Space transportation_航空航天运输
	388_*Ground transportation_地面运输

表6 经过层级校准后的UDC交通运输及其下级范畴类目

N transportation 交通运输	
	656.6 水运
	656.7 空运、空中交通
	656.1/.5 陆路运输

计是有帮助的。但本研究仍存在一些可改进之处，比如，对范畴类目相似度的判断，全由领域专家进行，没有判定范畴类目匹配程度的量化指标；层级校准方法虽具理论可行性，但在实践中可能存在一定误差；由工科范畴类目对比分析结论推广至一般性结论，可能存在一定误差等。

参考文献

[1] OCLC. Dewey Decimal System Home Page [EB/OL]. [2014-11-30]. http://www.oclc.org/dewey.en.html.

[2] Wikipedia. Dewey Decimal Classification [EB/OL].[2014-11-30]. http://en.wikipedia.org/wiki/Dewey_Decimal_Classification.

[3] UDC Consortium. Universal Decimal Classification Home Page [EB/OL]. [2014-11-30]. http://www.udcc.org/.

[4] Wikipedia. Universal Decimal Classification [EB/OL]. [2014-11-30]. http://en.wikipedia.org/wiki/Universal_Decimal_Classification.

[5] UDC Consortium. Overview of Last Reported Editions of the Universal Decimal Classification [EB/OL]. [2014-11-30]. http://www.udcc.org/files/UDCeditions_overview_2010July.pdf.

作者简介

刘家益，男，1986年生，中国农业科学院农业信息研究所研究实习员，研究方向：Web挖掘、知识抽取、本体等，E-mail：liujiayi@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：zhangxuefu@caas.cn。

DDC and UDC Comparative Analysis - A Case Study in Engineering

LIU JiaYi, ZHANG XueFu, PAN ShuChun, SUN Wei, HAO XinNing
(Agricultural Information Institute of Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: For a more in-depth understanding of the difference between DDC and UDC to design a better knowledge organization system, the latest version of the DDC and UDC Engineering disciplines-related categories were analyzed by the authors using the domain expert analysis and quantitative comparison method. The results show that: ①the knowledge coverage of DDC and UDC is high; ②DDC categories are divided more minutely than UDC; ③the category setting is quite different between DDC and UDC; ④the differences between the DDC and UDC categories are mainly caused by the angle of knowledge describing.

Keywords: Knowledge organization systems; DDC; UDC; Comparison; Analysis

（收稿日期：2014-11-20）