

# 印刷数字资产库文件编目与编码的研究

杨华, 刘真

(南京林业大学 江苏省制浆造纸科学与技术重点实验室, 南京 210037)

**摘要:** 为实现印刷数字资产库中文件的高效管理和快速检索, 根据印刷数字资产库对文件管理及检索的纵横需求, 建立了一套三级编目方法及相应的编码规则。经过对分类检索与逐一检索对应的检索量进行比较, 证实了三级编目能够有效提高文件的管理检索效率。

**关键词:** 数字资产库; 编目; 编码

**中图分类号:** TS801.8 **文献标识码:** A **文章编号:** 1001-3563(2011)01-0101-03

## Study of Catalogue and Coding in Print Digital Assets Database

YANG Hua, LIU Zhen

(Jiangsu Provincial Key Lab of Pulp and Paper Science and Technology, Nanjing Forestry University, Nanjing 210037, China)

**Abstract:** In order to realize the effective document management and searching, the vertical and horizontal function requirement of print digital assets database was analyzed. A three levels cataloguing method and coding rule were created. The assortative searching was compared with the one-by-one searching and it was proved that the three levels catalogue can increase the document management and searching efficiency.

**Key words:** digital assets database; catalogue; coding

由于印刷数字资产库中的数字文件量极为庞大, 如何能够让数字资产库的使用者方便有效地管理海量数据以及快速迅捷地进行文件存取, 成为一个印刷数字资产库能否帮助企业提高效益的关键。而对印刷企业数字文件进行合理的分类编目, 并构建相应编码规则, 是实现文件高效管理检索的基础<sup>[1]</sup>。因此, 笔者根据数字资产库对文件存储管理及检索的需求, 对编目与编码进行功能性分析, 并对印刷数字文件进行三级编目, 建立相应的文件编码规则与分类检索的方法。

### 1 印刷数字文件的一般功能结构

在印刷数字资产库的构建过程中, 既需要理清印刷文件的分类从属关系, 利于印刷资产库概念模型的构建, 又要满足印刷文件之间的相关性调用需求, 符合工作人员的调用习惯和实际生产组织方式<sup>[2]</sup>。因此, 编目方法与编码规则必须能够确定印刷文件的上

级所属类别, 即纵向关系; 同时, 还应该考虑各个印刷文件之间存在的生产或者管理上的共存关系, 即横向关系。

在纵向关系方面, 将印刷数字文件纵横关系(见图1)中黑色实心箭头所指各个节点取出就得到纵向

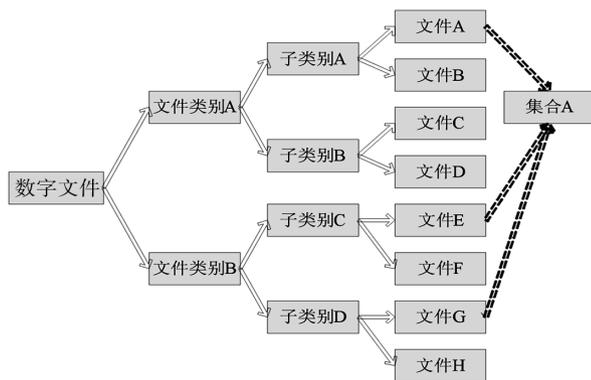


图1 印刷数字文件纵横关系

Fig. 1 Vertical and horizontal relationship of digital documents

图, 见图2。由图2可以看出, 纵向关系是一个多级包

收稿日期: 2010-10-07

作者简介: 杨华(1985-), 男, 河南人, 南京林业大学硕士生, 主攻印刷数字资产库、印刷品质量控制。



图 2 印刷数字文件纵向图

Fig. 2 Vertical relationship of digital documents

含关系,即下级节点是上级节点的子类别,并且文件需要经过多次分类才能最终确立其所属类别。以印刷过程中常用到的 PDF 文件为例,它是拼大版过程所用文件格式之一,因此是拼大版这一类别的子类。而拼大版又是生产过程的子类,因此就有了“生产文件”到“拼大版过程”到“PDF 文件”这一多级纵向关系,使文件所属类别清晰明了,利于数据库的构建。

在横向关系方面,将图 1 中虚线箭头所指各个节点取出就得到横向图,见图 3。由图 3 看出,横向关系



图 3 印刷数字文件横向图

Fig. 3 Horizontal relationship of digital documents

是多个同类文件同属于一个共性集合的并列关系,即各个文件并列共存且拥有某一不同于各自上级纵向关系的属性。例如,假定“文件 A”是一个 PDF 文件,“文件 E”是一个 ICC 文件,“文件 G”是一个 TIFF 文件,它们各自的上层纵向关系不同,一个是拼大版,一个是色彩管理,一个是原稿文件,但是,它们都用于同一个活件生产。如果日后需要加印此次活件,那么工作人员就希望能够直接调用用于此次活件的所有文件,因此,需要建立各个文件之间的横向关系以满足实际生产管理的需要。

## 2 印刷数字文件的编目

编目是指在信息描述的基础上编制目录,对信息资源的形式和内容信息进行分析选择及记录,并按照一定的规则有序地组织起来<sup>[3]</sup>。对印刷企业中的数字文件的编目,就是对其纵向关系的确定。采取的编目方法见图 4。

图 4 所给出的三级编目方式,首先按照印刷企业运营习惯,将所有数字文件分为生产与管理两大类,此为一级编目。生产文件方面,按生产过程中的不同

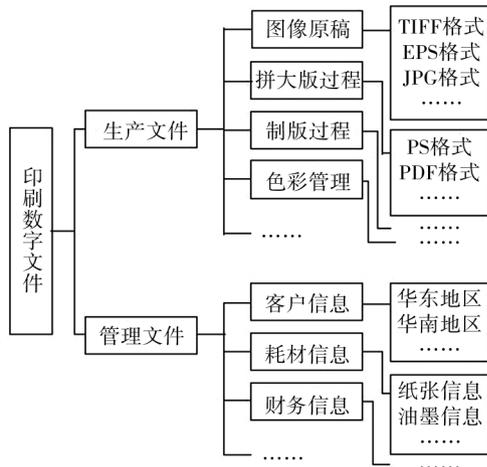


图 4 印刷数字文件编目方法

Fig. 4 Cataloguing method of print digital document

阶段进行二级编目,最后按照文件格式进行三级编目。管理文件方面,按企业管理的各个要素进行二级编目,最后根据各要素的构成要素进行三级编目。如此编目,符合印刷企业生产管理的基本模式,文件的从属类别也清晰明了。

## 3 印刷数字文件的编码

编码就是用预先规定的方法将文字、数字或其他对象编成数码的过程<sup>[4]</sup>。对于印刷数字文件来说,如果不进行编码就直接存入资产库中,那么能直接从数字文件中得到的信息就只有文件名、文件格式、创建时间、文件大小,而无法得知其在编目体系中的具体位置。这样就无法在印刷资产库中进行系统的、规范化的存储,令印刷资产库的文件管理杂乱无章。同时,也无法对文件进行分门别类的检索,降低资产库的工作效率。所以,编码规则的制定既要满足上文提到的编目方法,又要将文件的横向关系纳入规则制定所要考虑的重要因素。这样才能方便文件存储和管理,利于文件的快速检索。

建立的编码规则见图 5。以 1 位数来代表文件所属的第一级编目,2 位数代表其二级编目,3 位数代表其三级编目。这样,编码的前 6 位就可清楚地表示文件所属的编目,完成了纵向关系的构建。同时,每一级编目所对应位数的容量都超过其目录下子类的数量。以二级编目为例,其对应编码为 2 位数,容量为 00 到 99,而目前的印刷企业,无论是从生产工序来

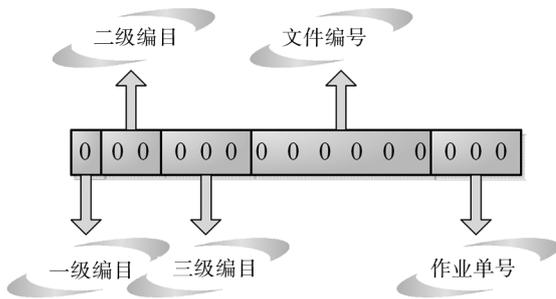


图5 印刷数字文件的编码规则

Fig. 5 The coding rule of print digital document

看,还是管理内容来看,都远远达不到100个。因此,这种编码规则使文件编目有良好的可扩展性,当有新的生产工序或新的管理内容产生时,不会因为编码容量的限制而去制定新的编码规则。

然后,给每个三级编目一个容量为10万的6位数编码,也就是每一种文件格式或者每一种已经三级细化的管理信息都有10万的存储量,这个容量足以满足绝大多数印刷企业的实际需求。最后,以3位数来代表此文件属于当天第几个作业单,以此完成文件之间横向关系的构建。作业单即生产人员常说的“活件”,一个作业单包含了各类生产文件和管理文件。例如一个“活件”在投入生产前,就包含客户、销售人员、报价等管理文件,还包括生产页数、工艺种类等生产文件;在生产过程中,既包含各类生产文件,也包含生产时间等管理信息;生产结束后,主要包含所用的耗材数量、生产成本等管理信息<sup>[5]</sup>。由此可以看出,以作业单作为横向关系构建的集合体,能够包含绝大多数的生产管理文件,而且它符合印刷企业基本的运作模式,利于文件的横向检索。

在编码完成后,可以根据编目方法可以进行分类检索。而不进行编目与编码,则只能按照存储顺序逐一进行检索。为比较2种检索方法的检索量,令 $P$ 为检索量, $N$ 为资产库文件总量, $S_1$ 为一级编目数, $S_2$ 为二级编目数, $S_3$ 为三级编目数<sup>[6-7]</sup>,比较结果见表1。

表1 2种检索方法检索量的比较

Tab. 1 Comparison of the two searching methods

检索方法	检索量公式	检索量( $N=100\ 000$ , $S_1=2, S_2=10, S_3=5$ )
分类检索	$P = \frac{N}{2}$	$P = \frac{N}{2} = \frac{100\ 000}{2} = 50\ 000$
逐一检索	$P = \frac{N}{2S_1S_2S_3}$	$P = \frac{N}{2S_1S_2S_3} = \frac{100\ 000}{2 \times 2 \times 10 \times 5} = 500$

从表1看出,简单的逐一式检索方法的检索量是分类检索的100倍之多,可见分类检索能够大大降低文件的检索量,提高检索效率。同时,编码中的作业单部分能够提供横向检索功能,即找到一个文件后,可以根据作业单号和文件建立时间来找出同一作业单中的所有其他文件,这能够大大提高生产和管理效率。

## 4 结语

经过本文的论证与分析,得到如下成果:通过对印刷数字资产库构建过程中的需求性分析,指出对编目与编码的功能性要求;可采用三级编目方法对印刷数字文件进行分类;根据三级编目方法,推荐了一套相应的编码规则;比较和分析了分类检索与逐一检索方式2种检索方法。

在印刷企业的实际生产和管理中,用到的数字文件种类繁多,本编目方法可能无法涵盖其全部;但是,建立的编码规则在每一级编目上都留有空余,使得编目方法具有良好的可扩展性。

## 参考文献:

- [1] 宋培义. 电视媒体数字资产管理模式研究[D]. 北京: 北京交通大学, 2008.
- [2] 黄军勤, 张志刚. 论包装企业数字资产管理平台的构建[J]. 包装工程, 2007, 28(7): 96-98.
- [3] 王海文, 万晓霞. 当代印刷发展的核心技术[J]. 包装工程, 2007, 28(3): 172-175.
- [4] BOLTON Ruth. The Theoretical Underpinnings of Customer Asset Management: A Framework and Propositions for Future Research[J]. Journal of the Academy of Marketing Science, 2004, 38(12): 271-292.
- [5] 曾俊. 印刷企业数字资产管理数据模型研究及其系统设计[D]. 武汉: 武汉大学, 2005.
- [6] 冒晓东, 李炳芳. 一种动态模糊管理数字资产对象的方法[J]. 包装工程, 2008, 29(5): 148-150.
- [7] 叶青. 大型印刷包装企业ERP与供应链系统集成的研究[J]. 包装工程, 2006, 27(4): 204-206.