



图书情报与信息管理实验教材

# 信息描述实验教程

An Experimental Instruction to Information Description

吴丹 等 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社



图书情报与信息管 理实验教材

# 信息描述实验教程

An Experimental Instruction to Information Description

编著 吴丹

参编 (按姓氏音序排列)

蔡卫萍 梁少博 刘孝平

罗博 史劲松 向雪



WUHAN UNIVERSITY PRESS

武汉大学出版社

## 图书在版编目(CIP)数据

信息描述实验教材/吴丹等编著. —武汉: 武汉大学出版社, 2016. 1  
图书情报与信息管理实验教材  
ISBN 978-7-307-16869-5

I. 信… II. 吴… III. 文献编目—教材 IV. G254.3

中国版本图书馆 CIP 数据核字(2015)第 222511 号

责任编辑:詹 蜜      责任校对:李孟潇      版式设计:韩闻锦

---

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 湖北民政印刷厂

开本: 720 × 1000 1/16      印张: 16.75      字数: 297 千字      插页: 1

版次: 2016 年 1 月第 1 版      2016 年 1 月第 1 次印刷

ISBN 978-7-307-16869-5      定价: 30.00 元

---

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。



## 图书情报与信息管理的实验教材

- |     |                 |
|-----|-----------------|
| 司 莉 | 信息组织实验教程        |
| 吴 丹 | 信息描述实验教程        |
| 肖秋会 | 档案信息组织与检索实验教程   |
| 许 洁 | 出版物排版制作软件实用教程   |
| 严炜炜 | 商务管理决策模型与技术试验教程 |
| 袁小群 | 数字出版物设计与制作      |
| 赵蓉英 | 信息计量分析工具理论与实践   |
| 吴志强 | 高级语言程序设计实验教程    |
| 陆颖隽 | 数据库原理与应用实验教程    |
| 黄如花 | 信息检索实验教程        |
| 王 林 | 信息系统分析与设计（课程设计） |
| 王 平 | 多媒体技术应用实验教程     |
| 姜婷婷 | 信息构建实验          |
| 孙永强 | 社会调查与统计分析实验     |
| 张 敏 | 电子供应链理论与实践      |

# 目录



<b>1 信息描述实验概述</b> .....	1
<b>2 中西文机读目录的比较</b> .....	3
2.1 实验目的与要求 .....	3
2.2 实验内容 .....	3
2.3 实验操作指导 .....	3
2.3.1 相关知识 .....	3
2.3.2 操作指导 .....	5
2.3.3 实验题目 .....	10
<b>3 CALIS Z39.50 客户端的使用</b> .....	11
3.1 实验目的与要求 .....	11
3.2 实验内容 .....	11
3.3 实验操作指导 .....	11
3.3.1 相关知识 .....	11
3.3.2 操作指导 .....	14
3.3.3 实验题目 .....	22
<b>4 MARC 的结构</b> .....	23
4.1 实验目的与要求 .....	23
4.2 实验内容 .....	23
4.3 实验操作指导 .....	23
4.3.1 相关知识 .....	23

4.3.2	操作指导	24
4.3.3	实验题目	27
<b>5</b>	<b>中文单行本图书原始编目</b>	<b>28</b>
5.1	实验目的与要求	28
5.2	实验内容	28
5.3	实验操作指导	28
5.3.1	相关知识	28
5.3.2	操作指导	30
5.3.3	实验题目	35
<b>6</b>	<b>西文单行本图书原始编目</b>	<b>36</b>
6.1	实验目的与要求	36
6.2	实验内容	36
6.3	实验操作指导	36
6.3.1	相关知识	36
6.3.2	操作指导	37
6.3.3	实验题目	41
<b>7</b>	<b>综合著录与分析著录</b>	<b>42</b>
7.1	实验目的与要求	42
7.2	实验内容	42
7.3	实验操作指导	42
7.3.1	相关知识	42
7.3.2	操作指导	44
7.3.3	实验题目	57
<b>8</b>	<b>规范记录的编制</b>	<b>58</b>
8.1	实验目的与要求	58
8.2	实验内容	58
8.3	实验操作指导	58
8.3.1	相关知识	58
8.3.2	操作指导	60

---

8.3.3 实验题目 .....	66
<b>9 图书馆编目系统的使用及中文套录 .....</b>	<b>68</b>
9.1 实验目的与要求 .....	68
9.2 实验内容 .....	68
9.3 实验操作指导 .....	68
9.3.1 相关知识 .....	68
9.3.2 操作指导 .....	70
9.3.3 实验题目 .....	87
<b>10 创建格式良好的 XML 文档 .....</b>	<b>88</b>
10.1 实验目的与要求 .....	88
10.2 实验内容 .....	88
10.3 实验操作指导 .....	88
10.3.1 相关知识 .....	88
10.3.2 操作指导 .....	91
10.3.3 实验题目 .....	94
<b>11 在 XML 文档中使用 DTD .....</b>	<b>96</b>
11.1 实验目的与要求 .....	96
11.2 实验内容 .....	96
11.3 实验操作指导 .....	98
11.3.1 相关知识 .....	98
11.3.2 操作指导 .....	98
11.3.3 实验题目 .....	105
<b>12 在 XML 文档中使用 Schema .....</b>	<b>106</b>
12.1 实验目的与要求 .....	106
12.2 实验内容 .....	106
12.3 实验操作指导 .....	108
12.3.1 相关知识 .....	108
12.3.2 操作指导 .....	108
12.3.3 实验题目 .....	111

---

<b>13 使用 XML 语言编写 RDF 文档</b> .....	114
13.1 实验目的与要求 .....	114
13.2 实验内容 .....	114
13.3 实验操作指导 .....	114
13.3.1 相关知识 .....	114
13.3.2 操作指导 .....	116
13.3.3 实验题目 .....	124
<b>14 使用 RDF 对 DC 元数据置标</b> .....	125
14.1 实验目的与要求 .....	125
14.2 实验内容 .....	125
14.3 实验操作指导 .....	125
14.3.1 相关知识 .....	125
14.3.2 操作指导 .....	132
14.3.3 实验题目 .....	141
<b>15 用 DC 元数据描述网页</b> .....	142
15.1 实验目的与要求 .....	142
15.2 实验内容 .....	142
15.3 实验操作指导 .....	142
15.3.1 相关知识 .....	142
15.3.2 操作指导 .....	143
15.3.3 实验题目 .....	146
<b>16 FRBR 在图书馆联机目录中的应用</b> .....	147
16.1 实验目的与要求 .....	147
16.2 实验内容 .....	147
16.3 实验操作指导 .....	147
16.3.1 相关知识 .....	147
16.3.2 操作指导 .....	149
16.3.3 实验题目 .....	152



<b>17 MARC 记录的 FRBR 化显示</b> .....	153
17.1 实验目的与要求 .....	153
17.2 实验内容 .....	153
17.3 实验操作指导 .....	153
17.3.1 相关知识 .....	153
17.3.2 操作指导 .....	157
17.3.3 实验题目 .....	160
<b>18 MARC 记录的 BIBFRAME 转换</b> .....	161
18.1 实验目的与要求 .....	161
18.2 实验内容 .....	161
18.3 实验操作指导 .....	161
18.3.1 相关知识 .....	161
18.3.2 操作指导 .....	168
18.3.3 实验题目 .....	170
<b>19 利用 GreenStone 软件构建个人数字图书馆</b> .....	171
19.1 实验目的与要求 .....	171
19.2 实验内容 .....	171
19.3 实验操作指导 .....	172
19.3.1 相关知识 .....	172
19.3.2 操作指导 .....	173
19.3.3 实验题目 .....	203
<b>附录 1 图书原始编目实验题目实例</b> .....	205
<b>附录 2 CNMARC 与 MARC21 字段一览表</b> .....	224
一、CALIS 联合目录 CNMARC 字段一览表 .....	224
二、MARC21 字段简表 .....	249
<b>参考文献</b> .....	259

# 1

## 信息描述实验概述

信息描述 (Information Description) 是在网络时代中, 学术界对传统“文献编目”概念的一种继承和发展。在网络时代, 图书馆所搜集、整理的对象已经延伸到网络中的虚拟馆藏, 网络信息已经成为图书馆重要的信息资源。而“文献编目”无论从内涵还是从外延上都无法涵盖现有的图书馆编目对象和编目内容, 于是出现了“信息描述”的概念。信息描述, 是根据信息组织和检索的需要, 对信息资源的主题内容、形式特征、物质形态等进行分析、选择、记录的活动。其结果为描述记录 (即元数据记录), 作为信息资源的替代物组织检索系统。

近年来, 信息描述领域的研究、实践发展迅速, 都柏林核心元素集 (DC) 的广泛应用, 书目记录的功能需求 (FRBR) 的发布, 资源描述与检索 (RDA) 的发布与应用, BIBFRAME 项目的启动等, 新理念、新框架、新工具层出不穷。图书馆学信息描述课程需要跟上时代的步伐, 同时也需要相应的实验操作课程与信息描述理论课程相配合, 提升学生的信息描述基本技能, 并通过实践使学生深化对信息描述新理念、发展趋势的理解, 促进其相关领域的研究。

本教程的内容设计重视新颖性和实践性, 不仅将信息描述界最新的理念、工具等引入各个章节中, 而且为每个章节设置了可操作性强、挑战性较高, 并具有开放性的实验题目。每一章的内容包括三个方面: 实验目的与要求, 实验内容, 实验操作指导。实验目的与要求阐述了学生通过本章学习应该达到的目标; 实验内容是学生在本章应该掌握的具体知识; 实验操作指导主要介绍了该实验的背景知识、相关工具的使用指南, 以及具体的实验题目。整个教程共有 18 个实验, 实验内容分类及适用对象如表 1-1 所示。

表 1-1 本书实验内容分类及适用对象

分类	实验名称	适用对象
文献信息编目	中西文机读目录的比较	本科生
	CALIS Z39.50 客户端的使用	
	MARC 的结构	
	中文单行本图书原始编目	
	西文单行本图书原始编目	
	综合著录与分析著录	
	规范记录的编制	
	图书馆编目系统的使用及中文套录	
标记语言基础	创建格式良好的 XML 文档	本科生
	在 XML 文档中使用 DTD	
	在 XML 文档中使用 Schema	
	使用 XML 语言编写 RDF 文档	
DC 元数据	使用 RDF 对 DC 元数据置标	本科生、研究生
	用 DC 元数据描述网页	
信息描述新发展	FRBR 在图书馆联机目录中的应用	研究生
	MARC 记录的 FRBR 化显示	
	MARC 记录的 BIBFRAME 转换	
综合项目	利用 GreenStone 软件构建个人数字图书馆	研究生

## 2.1 实验目的与要求

- (1) 了解 CNMARC 的基本结构及在中文文献著录时的常用字段；
- (2) 了解 MARC21 的基本结构及在西文文献著录时的常用字段；
- (3) 了解 CNMARC 与 MARC21 之间的异同。

## 2.2 实验内容

- (1) 利用国内图书馆馆藏目录查询系统，查看一本中文图书的机读目录格式，分析 CNMARC 的基本结构及在中文图书著录时的常用字段；
- (2) 利用国外图书馆馆藏目录查询系统，查看一本外文图书的机读目录格式，分析 MARC21 的基本结构及在西文图书著录时的常用字段；
- (3) 以一部作品的中英文版本不同格式的机读目录为例，比较 CNMARC 与 MARC21，找出两者之间的映射关系。

## 2.3 实验操作指导

### 2.3.1 相关知识

机读目录 (Machine-Readable Cataloging) 是一种以代码形式和特定结构记录在计算机存储载体上，可为计算机识别和处理的目录。目前国内常用的机读目录格式有 CNMARC 和 MARC21，分别对中文文献和西文文献进行著录。

CNMARC 是北京图书馆依据 UNIMARC 编制完成的中国机读目录格式, 目前被广泛应用于国内图书馆的中文文献编目工作中, 已成为我国最为成熟的中文文献标准化数据处理格式。CNMARC 记录的结构主要由记录头标区、地址目次区、数据字段区和记录分隔符四部分组成, 其中记录头标区是由 24 个字符构成的定长数据区, 包括记录长度、记录状态、执行代码及指示符长度等 8 个数据元素; 地址目次区由若干目次项和字段分隔符构成, 每个目次项的固定字符长度为 12, 分别揭示数据字段区中对应字段的字段标识符、数据字段长度以及起始字符位置; 数据字段区设置有 10 个功能块, 分别为: 0××—标识块, 1××—编码信息块, 2××—著录信息块, 3××—附注块, 4××—款目连接块, 5××—相关题名块, 6××—主题分析块, 7××—知识责任块, 8××—国际使用块, 9××—国内使用块; 记录分隔符是位于每条机读记录的结尾, 用以区分不同记录的控制字符。

MARC21 是美国国会图书馆和加拿大国家图书馆在 USMARC 和 CANMARC 的基础上经过不断整合共同完成的一种新的机读目录格式, 能与 UKMARC 和 UNIMARC 等格式兼容, 被誉为 21 世纪通用的机读目录格式。MARC21 记录的结构主要由记录头标区、地址目次区和可变长字段区三部分组成, 其中记录头标区部分的固定字符长度为 24, 由记录长度、记录状态、记录类型和书目级别等 13 个固定数据元素组成; 地址目次区是记录中每个可变长控制字段和可变长数据字段位置的索引, 由若干个款目构成, 每个款目的固定长度为 12 个字符, 包括字段标识符、字段长度和字段起始位置; 可变长字段区设置有 10 个字段块, 分别为: 0××—控制信息、控制号和代码字段, 1××—主要款目字段, 2××—题名、版本和出版说明字段, 3××—载体形态等字段, 4××—丛编说明字段, 5××—附注字段, 6××—主题检索字段, 7××—名称等附加款目和连接款目字段, 8××—丛编附加款目及馆藏定位与检索等字段, 9××—本地使用字段。

CNMARC 与 MARC21 机读目录的主体结构原理大致相同, 只是在具体的字段及数据元素的设置与使用方面存在差异。在基本结构方面, CNMARC 和 MARC21 都由记录头标区、地址目次区和数据字段区构成, 且 CNMARC 中的数据字段区与 MARC21 中的可变长字段区基本上涵盖了主要的著录项目, 只是两者在具体的字段对应方面, 会有一对多、多对一以及多对多等多种关系。另外, MARC21 每条记录后均有记录分隔符, 只是未被视为主体的结构的一部分。

## 2.3.2 操作指导

### 2.3.2.1 CNMARC 的基本结构及在中文图书著录时的常用字段分析

利用国家图书馆的馆藏目录 (<http://www.nlc.gov.cn/>), 检索一本由编目精灵所著的中文图书《编目的未来》, 以机读格式显示, 浏览和分析该书的 CNMARC 书目记录。该书的 CNMARC 书目记录如图 2-1 所示:

```
LEADER 00000cam0#2200349###450#
001 ## 004560569
005 ## 20101025150525. 0
010 ## $ a 978-7-5013-4343-0 $ d CNY38. 00
100 ## $ a 20100603d2010 em y0chiy50 ea
101 0# $ a chi
102 ## $ a CN $ b 110000
105 ## $ a y z 000yy
106 ## $ a r
200 1# $ a 编目的未来 $ A bian mu di wei lai $ f 编目精灵著
210 ## $ a 北京 $ c 国家图书馆出版社 $ d 2010
215 ## $ a 10, 193 页 $ d 23cm
312 ## $ a 封面英文题名: The future of cataloging
330 ## $ a 本书介绍和评价图书馆标引和编目的新原则、新规则, 在编目过程中出现的新问题的解决方案, 处理原则和技巧。
510 1# $ a Future of cataloging $ z eng
606 0# $ a 文献编目
690 ## $ a G254. 3 $ v 4
701 #0 $ a 编目精灵 $ f (1964~) $ 4 著 $ 9 bian mu jing ling
090 ## $ a G254 $ b bmj
096 ## $ a G254 $ b bmj
```

图 2-1 《编目的未来》的 CNMARC 书目记录

浏览该书的机读目录, 可知整个书目记录主要由记录头标区和数据字段区组成, 另外两部分即地址目次区和记录分隔符由计算机自动生成, 不在馆藏系统的机内格式中予以反映。

记录头标位于书目记录的最顶端, 为不可重复的必备项目。分析本条记录

的头标区数据“00000cam0#2200349###450#”，其中“00000”为记录长度；“c”为记录状态，表示已经修改过的记录；“am0#”为执行代码(“a”为记录类型，表示这是一条文字资料印刷品的书目记录；“m”为书目级别，表示这是一条专著的书目记录；“0”为层次等级，表示这是一条无层次等级关系的记录；“#”为未定义)；第一个“2”为字段指示符长度；第二个“2”为子字段标识符长度；“00349”为数据基地址；“###”为记录附加定义；“450#”为地址目次区结构。

数据字段区包含功能块、字段、子字段和数据元素四个层次。数据字段区共设置有 10 个功能块，根据字段标识符的第一位数字，可判断该字段所代表的功能块，如本记录中的“001”字段标识符，最左边的数字“0”表示该字段属于标识块；“210”中的“2”则表示该字段属于著录信息块。数据字段区还采用了字段指示符和子字段标识符以提供和识别相关信息，如“200 1# \$ a 编目的未来 \$ Abian mu di wei lai \$ f 编目精灵著”这一著录信息块，“200”为字段标识符；“1#”为字段指示符，其中“1”表示题名作为检索点，“#”为未定义；“\$ a”、“\$ A”和“\$ f”则为子字段标识符，分别代表正题名、正题名汉语拼音和第一责任说明。

### 2.3.2.2 MARC21 的基本结构及在西文图书著录时的常用字段分析

利用美国国会图书馆的馆藏目录 (<http://www.loc.gov/>)，检索一本于 2013 年出版，题名为 *Health Information Seeking* 的外文书籍，查看其“MARC Tags”，浏览和分析该书的 MARC21 书目记录。该书的 MARC21 书目记录如图 2-2 所示。

浏览该条书目记录，可知 MARC21 书目记录主要包括记录头标区和可变长字段区，地址目次区由计算机自动生成，在馆藏系统中的机内格式中不予反映。

记录头标区位于书目记录的开头，为必备且不可重复的部分。分析本条记录的头标区数据“01905cam#a2200409#a#4500”，其中“01905”为记录长度；“c”为记录状态，表示已经修改过的记录；第一个“a”为记录类型，表示这是一条文字资料的书目记录；“m”为书目级别，表示这是一条图书的书目记录；“#”控制类型，表示无特定的控制类型；第二个“a”表示系统采用 UCS/Unicode 字符编码体系；第一个“2”为字段指示符长度；第二个“2”为子字段标识符长度；“00409”为数据基地址；“#”为编目等级，表示该记录为完全级；第三个“a”为著录格式，表示著录标准为 AACR2；“#”为记录连接要求，表示不要求连接相关记录；“4500”为地址目次区结构。

```

000 01905cam#a2200409#a#4500
001 17250192
005 20130102095722. 0
008 120406s2012 nyua b 001 0 eng
906 ## $ a 7 $ b cbc $ c orignew $ d 1 $ e ecip $ f 20 $ g y-gencatlg
925 0# $ a acquire $ b 1 shelf copy $ x policy default, Sel/rjc, 2012-09-24
955 ## $ w rd13 2012-04-17 $ d xh11 2012-06-21 to Dewey $ w rd13 2012-06-21
    $ a xn03 2012-09-20 1 copy rec'd., to CIP ver.
    $ f xh10 2013-01-02 CIP Ver. to CALM
010 ## $ a 2012014493
020 ## $ a 9781433118258 (hardcover; alk. paper)
020 ## $ a 9781433118241 (pbk.; alk. paper)
020 ## $ a 9781453908792 (e-book)
035 ## $ a (DNLM) 101581856
040 ## $ a DNLM/DLC $ c DLC $ d DLC
042 ## $ a pcc
050 00 $ a RA773. 6 $ b . J64 2012
060 10 $ a WA 590
082 00 $ a 613 $ 2 23
100 1# $ a Johnson, J. David.
245 10 $ a Health information seeking/ $ c J. David Johnson & Donald O. Case.
260 ## $ a New York; $ b Peter Lang, $ c c2012.
300 ## $ a xiv, 274p.; $ b ill.; $ c 24cm.
490 1# $ a Health communication, $ x 2153-1277; $ v v. 4
504 ## $ a includes bibliographical references (p. [228] -267) and index.
505 ## $ a Introduction to health information seeking-information fields and c
    arriers-Socio-psychological factors in health-information carrier: a
    focus on channel selection and useage-Model of information seeking-
    Outcomes of information seeking-Strategies for seekers ( and non seekers) -
    Strategies for health professionals-Summing up: information seeking in the information age.
650 #0 $ a Health $ x information services.
650 #0 $ a Information retrieval.
650 12 $ a Consumer Health information.
650 12 $ a Information Seeking Behavior.
650 22 $ a Health Communication.
700 1# $ a Case, Donald Owen.
830 #0 $ a Health communication (New York, N. Y.); $ v v. 4 $ x 2153-1277
985 ## $ a NLMCIP $ d 2012-04-18
985 ## $ a NLMCIP $ d 2012-04-11

```

图 2-2 *Health Information Seeking* 的 MARC21 书目记录



可变长字段区设置有 00×~9××共 10 个字段块, 可根据字段标识符的第一位数字判断该字段所代表的字段块。其中书目记录中的 00×字段为可变长控制字段, 无字段指示符和子字段标识符, 如书目记录中的“001”、“005”和“008”字段, 为控制信息、控制号和代码字段; 01×~9××字段为可变长数据字段, 由指示符、子字段标识符和长度不定的数据元素组成, 如本记录中 260 字段的数据信息“260 ## \$ a New York: \$ b Peter Lang, \$ c 2012.”, 其中“260”为字段标识符, “##”为字段指示符, 此处第一个“#”表示无信息提供, 第二个“#”为未定义; “\$ a”、“\$ b”、“\$ c”、“:”、“,”等则为子字段标识符, 表示在编图书的出版地、出版者和出版日期。

### 2.3.2.3 CNMARC 与 MARC21 之间的比较

首先利用国家图书馆的馆藏目录查询系统检索王知津等学者的译作《现代信息检索》, 查看其 CNMARC 格式的书目记录, 然后在美国国会图书馆的馆藏目录查询系统中查找于 1999 年出版的外文图书 *Modern Information Retrieval*, 查看其 MARC21 格式的书目目录, 通过一部作品两种不同格式的机读目录, 找出 CNMARC 与 MARC21 之间的映射关系。两种格式的书目记录如图 2-3、图 2-4 所示。

以两条机读目录为例, 可以发现, CNMARC 与 MARC21 记录头标区的固定长度均为 24 个字符, 且大多数数据元素名称相同, 只是对字符位置的定义及数据元素的赋值稍有不同。如 CNMARC 将 08 字符位定义为层次等级代码, 09 字符未定义, 相对应的 MARC21 则将这两个字符位分别定义为控制类型和字符编码系统。而在数据元素赋值方面, 本条 CNMARC 书目记录中的记录状态“n”代表新记录, 其他的记录状态还包括“o”(有较高层次的记录)、“c”(修改过的记录)、“d”(删除的记录)等, 而 MARC21 中的记录状态“c”虽然也表示修改过的记录, 但其他记录状态中的“a”则表示有简编记录升级的记录, 这是 CNMARC 中没有的, 而 CNMARC 中的“o”状态也是 MARC21 中不存在的。

在数据字段区和可变长字段区方面, 根据具体的文献情况, CNMARC 和 MARC21 会选择相应的字段进行书目著录。在本例中, CNMARC 格式书目记录对 10 个功能块均有涉及, 但在 MARC21 格式的书目记录中, 10 个字段并未完全涉及, 以 4、8 和 9 开头的字段并没有相关的著录项目。此外, CNMARC 和 MARC21 对于各个字段著录的顺序有所不同, CNMARC 严格按照 0~9 的顺序进行著录, MARC21 却并非如此。关于 CNMARC 和 MARC21 之间的对应关