







工业和信息化部“十四五”规划教材



“十二五”普通高等教育本科国家级规划教材

“十三五”江苏省高等学校重点教材

# 信息组织

## (第3版)

◆ 叶继元 编著

电子工业出版社

Publishing House of Electronics Industry

北京 · BEIJING

## 内 容 简 介

本书为工业和信息化部“十四五”规划教材和“十二五”普通高等教育本科国家级规划教材。

本书系统讲述了信息组织的基本知识、经典理论，详细介绍了信息的描述、著录、标引和排序等方法，注重引入与信息组织有关的最新研究成果，如信息构建理论、本体论、搜索引擎、本体描述语言、SKOS 语言、学科信息门户等。

本书共 9 章，内容包括：信息组织概述、信息组织的理论与方法基础、信息描述语言、信息著录法、信息标引法、信息排检法、信息组织成果与工具、语义网环境下的信息组织、非网络环境下的信息组织。每章正文之前有内容概述、本章重点和要掌握的要点，每章后附有本章小结和习题，以满足教学和自学的需要。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

### 图书在版编目（CIP）数据

信息组织 / 叶继元编著. —3 版. —北京：电子工业出版社，2023.6

ISBN 978-7-121-46051-7

I. ① 信… II. ① 叶… III. ① 信息组织—高等学校—教材 IV. ① G254

中国国家版本馆 CIP 数据核字（2023）第 136690 号

责任编辑：章海涛

特约编辑：李松明

印 刷：

装 订：

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编：100036

开 本：787×1092 1/16 印张：30 字数：760 千字

版 次：2010 年 2 月第 1 版

2023 年 6 月第 3 版

印 次：2023 年 6 月第 1 次印刷

定 价：74.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 [zlts@phei.com.cn](mailto:zlts@phei.com.cn)，盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

本书咨询联系方式：192910558（QQ 群）。

## 前　　言

时间过得很快，一晃 8 年的时间就过去了。本书第 2 版修订完成时间是 2014 年 5 月，正式出版时间是 2015 年 3 月。第 3 版的修订工作始于 2019 年 2 月，至今又是将近 4 年时间。现将修订情况简要说明如下。

第一，对第 2 版教材中出现的疏漏之处进行了勘误，对前后重复交叉之处做了结构性调整。例如，将第 1 章中有关信息描述与揭示、信息序化与存储的定义同后续章节中的定义进行了对照，从而确保表述一致性；结合新进展、新标准、新规范、新状态等对定义、实例、案例等做了修改，如在 3.3.2 节中增加了最新的《中国分类主题词表》（第 3 版）的内容。

《中国分类主题词表》（第 3 版）共 2 卷 8 分册，更新重点在于增改主题词、与《中图法》（第 5 版）类目对应，起到《汉语主题词表》第 3 次修订的作用；根据文献保障原则，新增了大量主题词，删除了一些错误或无文献保障的主题词；为满足自然语言检索需要，增加大量的非优选主题词（人口词、人口短语），并为大多优选词建立语义关系，增强了优选词及主题词组配式的检索效率，以及文献主题族性检索的作用。《中国分类主题词表》（第 3 版）的电子版、Web 版为实现机助标引提供了知识库和应用接口。在 6.2 节中以《中图法》（第 5 版）为例，对学科体系分类排检法中的分类号（排检标识）排序进行了说明；对涉及《现代汉语词典》等新版本、语义网规范的版本更新内容，均进行了重新编写。

第二，新增了一些信息组织领域的新进展，对某些概念进行重新定义和解释，如对 RDF 数据存储、SPARQL 查询语言、关联数据应用实例、知识图谱信息组织方法等新进展的介绍。又如，7.1.1 节“目录的类型”中对评论性书目的概念进行了重新定义。评论性书目（Critical Bibliography）是对著者、成书年代及版本插图、构成材料进行考证的书目。这类带有编者评述的书目，有的从内容上对文献进行分析、评论，有的从文献形式上进行考订、比较，都是为了指导阅读和帮助选阅文献。7.1.2 节“目录举要”中结合新近发展对 Ulrich（国际期刊指南）进行了重新描述。Ulrich 是全球范围内的期刊名录数据库，其印刷版创刊于 1932 年，早年也是美国鲍克公司的目录产品，目前以网络数据库的形式提供服务。Ulrich 提供了世界范围内超过 30 万种正式出版的期刊和非正式的连续出版物，既包括学术性的、商业性的期刊，又收录开放存取的出版物；既有经同行评议而选出的刊物，又有大众化普及性的杂志；既有收费刊物，又有免费刊物，还有一定数量的报纸、快报。Ulrich 多用于查找与期刊出版有关的各类问题，如期刊刊名的变更情况、期刊被文摘索引及数据库收录的情况等，近年来还链接了 Web of

Science 系统的期刊引用报告 (JCR) 提供的期刊影响因子的数据；反过来，JCR 也链接了 Ulrich 数据库关于某刊的具体信息。

第三，增加了新数据库和大量应用实践相关内容，如 7.2.3 节中增加了 CSSCI 最新来源期刊、扩展版期刊的种数，以及有关 Scopus 的介绍。Scopus 是著名学术出版社爱思唯尔 (Elsevier) 继其全文期刊数据库 Science Direct 之后推出的以期刊文献为主的大型文摘索引数据库。Scopus 覆盖了来自 150 多个国家、超过 40 多种语言、5000 多家国际出版商的约 21900 种来源期刊，其中大多数为同行评议期刊，还包括一定数量的图书、丛书、会议论文、专利文献等，且每日更新。Scopus 大量应用实践相关内容反映在 8.2.4 节 (RDF 数据的存储)、8.2.5 节 (RDF 查询语言 SPARQL)、8.5.5 节 (关联数据应用实例)、8.6 节 (基于知识图谱的信息组织方法) 等中。

第四，修改了某些内容。结合本科生的学习特点，删除了难度较大的部分，如术语注册的内容 (比较深，属于研究层面) 等。原书第 2 版中有些内容直接翻译自英文资料，文字拗口，不易理解，故重新写过，力求深入浅出。又如，本书修改了所有 RDF 代码并进行了验证，以保证所有代码运行正确，新增了一些实例代码，以便于理解。

本书第 3 版的修订组织和统稿由叶继元负责。感谢以下人员为本书自第 1 版以来付出的工作，包括：**刘磊**、冯英华、陈文娟、刘丹、王晓艳、薛春香、谭华军、岳泉、罗琳、徐美凤、郭春侠、储节旺、华薇娜、欧石燕、王军、黄如花、徐青梅、刘凤仪、陶蕊、戎迎盈、杨鑫淼、孙毓、廖林望、高蕴梅、马梦恬、解荣、杨斐、成永娟、周星宇、郝建华、王雅戈等。

本书在 2016 年被评为“十二五”普通高等教育本科国家级规划教材，2017 年被评为“十三五”江苏省高等学校重点教材，2021 年被评为工业和信息化部“十四五”规划教材。特向各位评审专家表示衷心的谢意。南京大学教务处/本科生院和电子工业出版社的相关人员在本书出版中给予了大力帮助，在此表示衷心感谢。对本书引用论著的作者、编者、教材使用者、提出修改建议者也表示衷心的谢意。

信息资源丰富，信息技术发展迅速，信息组织内容面广，变化亦多。由于时间、水平限制，书中难免存在不当之处，敬请读者指正。

本书为教师提供相关教学资料，有需要者，请登录 <http://www.hxedu.com.cn>，注册之后进行下载。

作 者  
2023 年 4 月

# 目 录

<b>第 1 章 信息组织概述 .....</b>	<b>1</b>
1.1 信息组织的界定 .....	2
1.1.1 信息组织的概念 .....	2
1.1.2 信息组织的原理 .....	5
1.1.3 信息组织的内容 .....	6
1.2 信息组织的类型 .....	7
1.2.1 按信息的表现形式划分 .....	7
1.2.2 按信息的加工程度划分 .....	8
1.2.3 按信息的传播载体划分 .....	9
1.2.4 按信息的认识层次划分 .....	11
1.2.5 按信息的存在环境划分 .....	13
1.3 信息组织的作用 .....	14
1.4 信息组织的发展 .....	15
1.4.1 古代的信息组织 .....	15
1.4.2 近代的信息组织 .....	19
1.4.3 现代的信息组织 .....	22
1.4.4 信息组织的未来发展趋势 .....	31
本章小结 .....	37
习题 1 .....	37
<b>第 2 章 信息组织的理论与方法基础 .....</b>	<b>38</b>
2.1 信息组织的理论基础 .....	39
2.1.1 有序化理论 .....	39
2.1.2 信息构建理论 .....	42
2.1.3 知识论 .....	46
2.1.4 本体论 .....	49
2.1.5 分形理论 .....	51
2.2 信息组织的方法基础 .....	52
2.2.1 语言学 .....	52
2.2.2 逻辑学 .....	55
2.2.3 知识分类学 .....	59
本章小结 .....	70
习题 2 .....	71
<b>第 3 章 信息描述语言 .....</b>	<b>72</b>

3.1	信息描述语言概述 .....	73
3.1.1	规范语言 .....	73
3.1.2	自然语言 .....	80
3.2	分类语言 .....	86
3.2.1	分类法的原理 .....	86
3.2.2	分类法的编制 .....	91
3.2.3	国内外常用分类法介绍 .....	107
3.3	主题语言 .....	124
3.3.1	主题法概述 .....	124
3.3.2	国内外常用主题词表介绍 .....	132
	本章小结 .....	149
	习题 3 .....	149
	<b>第 4 章 信息著录法 .....</b>	<b>150</b>
4.1	传统著录法 .....	151
4.1.1	传统著录法概述 .....	151
4.1.2	文献信息著录规则 .....	154
4.2	机读目录著录法 .....	166
4.2.1	MARC 在全球的发展概述 .....	166
4.2.2	MARC 记录基本格式 .....	168
4.2.3	MARC 著录的优缺点 .....	173
4.3	元数据著录法 .....	175
4.3.1	元数据简介 .....	175
4.3.2	都柏林核心元数据集 .....	177
4.3.3	其他元数据 .....	179
	本章小结 .....	184
	习题 4 .....	185
	<b>第 5 章 信息标引法 .....</b>	<b>186</b>
5.1	分类标引 .....	187
5.1.1	分类标引要求 .....	187
5.1.2	分类标引方法 .....	188
5.1.3	分类标引规则 .....	192
5.2	主题标引 .....	198
5.2.1	主题标引概述 .....	198
5.2.2	主题标引方法 .....	200
5.2.3	主题标引规则 .....	205
5.2.4	主题标引与分类标引的比较 .....	208
5.2.5	关键词标引 .....	209
5.3	自动分类与自动标引 .....	212
5.3.1	自动分类概述 .....	212

5.3.2 自动标引概述 .....	217
5.3.3 西文信息自动标引技术 .....	219
5.3.4 汉语信息自动标引技术 .....	222
5.3.5 信息自动标引有待研究的问题 .....	227
本章小结 .....	229
习题 5 .....	230
<b>第 6 章 信息排检法 .....</b>	<b>231</b>
6.1 字序法 .....	232
6.1.1 汉字音序排检法 .....	232
6.1.2 汉字形序排检法 .....	236
6.1.3 外文字顺法 .....	241
6.2 类序法 .....	243
6.2.1 学科体系分类排检法 .....	243
6.2.2 事物性质分类排检法 .....	245
6.2.3 网络信息分类排检法 .....	246
6.2.4 主题词排检法 .....	247
6.2.5 网络信息关键词排检法 .....	248
6.3 时序法 .....	250
6.3.1 历法常识 .....	250
6.3.2 中国古代的时序法 .....	252
6.3.3 时序法的应用 .....	255
6.4 地序法及其他排检法 .....	256
6.4.1 地序法 .....	256
6.4.2 其他排检法 .....	257
6.4.3 计算机程序与动态信息排检法 .....	258
本章小结 .....	260
习题 6 .....	260
<b>第 7 章 信息组织成果与工具 .....</b>	<b>261</b>
7.1 目录 .....	262
7.1.1 目录的类型 .....	262
7.1.2 目录举要 .....	264
7.2 索引与文摘 .....	269
7.2.1 索引的概念 .....	269
7.2.2 索引的类型 .....	270
7.2.3 索引举要 .....	273
7.2.4 文摘 .....	277
7.3 全文数据库 .....	277
7.3.1 全文数据库概述 .....	277
7.3.2 全文数据库开发 .....	280

7.3.3 全文数据库举要 .....	282
7.4 数据库整合与导航 .....	286
7.4.1 数据库整合 .....	286
7.4.2 数据库导航 .....	291
7.4.3 学科信息门户 .....	295
7.5 搜索引擎 .....	299
7.5.1 搜索引擎概述 .....	299
7.5.2 搜索引擎工作机制 .....	302
7.5.3 搜索引擎举要 .....	303
7.5.4 搜索引擎的发展方向 .....	304
本章小结 .....	306
习题 7 .....	306
<b>第 8 章 语义网环境下的信息组织 .....</b>	<b>307</b>
8.1 语义网概述 .....	308
8.2 语义网信息描述和表示 .....	309
8.2.1 RDF 简介 .....	309
8.2.2 RDF 序列化表示格式 .....	311
8.2.3 RDF 评价 .....	312
8.2.4 RDF 数据的存储 .....	313
8.2.5 RDF 查询语言 SPARQL .....	313
8.3 语义网信息组织模式 .....	316
8.3.1 本体简介 .....	316
8.3.2 本体的类型 .....	317
8.3.3 本体的功能 .....	317
8.3.4 本体与传统受控词表的比较 .....	317
8.3.5 本体描述语言 .....	318
8.3.6 本体的构建 .....	324
8.3.7 基于本体的信息组织实例 .....	326
8.4 基于语义网的知识组织系统 .....	332
8.4.1 SKOS 语言简介 .....	332
8.4.2 SKOS 语言应用实例 .....	336
8.5 关联数据 .....	341
8.5.1 关联数据简介 .....	341
8.5.2 关联数据中资源的命名及访问机制 .....	342
8.5.3 关联数据发布方法 .....	344
8.5.4 关联数据的访问 .....	345
8.5.5 关联数据应用实例 .....	348
8.6 基于知识图谱的信息组织方法 .....	350
8.6.1 知识图谱概述 .....	350

8.6.2 图书馆学相关知识图谱 .....	350
本章小结 .....	352
习题 8 .....	352
<b>第 9 章 非网络环境下的信息组织 .....</b>	<b>353</b>
9.1 印刷型文献的信息组织评价 .....	354
9.1.1 词典信息组织的评价和实例 .....	355
9.1.2 百科全书信息组织的实例与评价 .....	358
9.2 多媒体文献的信息组织评价 .....	365
9.2.1 光盘数据库信息组织的特点 .....	365
9.2.2 光盘数据库信息组织的评价指标 .....	366
9.2.3 光盘数据库信息组织的实例 .....	368
9.3 网站的信息组织评价 .....	371
9.3.1 网站信息资源组织的特点 .....	372
9.3.2 网站信息资源组织的评价标准 .....	372
9.3.3 Web of Science 信息资源组织评析 .....	376
9.3.4 光盘数据库与网络版数据库信息组织的比较与实例 .....	380
9.4 搜索引擎的信息组织评价 .....	382
9.4.1 搜索引擎的评价标准 .....	382
9.4.2 著名综合搜索引擎谷歌的评析 .....	387
9.5 学科信息门户的信息组织评价 .....	390
9.5.1 学科信息门户资源组织的特点 .....	390
9.5.2 学科信息门户信息组织的评价标准 .....	391
9.5.3 学科信息门户实例 IPL2 评价 .....	394
9.6 Web 2.0 环境下的信息组织评价 .....	396
9.6.1 Web 2.0 环境下的信息组织的特点 .....	396
9.6.2 Web 2.0 环境下的信息组织的评价标准 .....	397
9.7 数字图书馆的信息组织评价 .....	407
9.7.1 数字图书馆及其信息组织的特点 .....	407
9.7.2 数字图书馆信息组织的评价指标 .....	408
9.7.3 “美国记忆”数字图书馆项目评价 .....	412
本章小结 .....	415
习题 9 .....	415
<b>附录 A 内容索引 .....</b>	<b>416</b>
A.1 主题索引 .....	417
A.2 人名索引 .....	438
A.3 文献索引 .....	443
<b>参考文献 .....</b>	<b>456</b>



# 第1章

## 信息组织概述

人类自古就在信息的海洋中生活。例如，原始人在森林中搜寻野果和野兽的信息，探悉各种猎物的信息。人类很早就懂得利用信息的一些性质来达到特定的目的。例如，结绳记事就利用了信息的可存储性。

在当代信息社会中，信息已成为人们熟知的流行词，日益受到人们的重视，同能源和物质一起，被称为人类社会生产与生活必不可少的三大资源。随着现代科学技术特别是信息技术的迅猛发展，社会已经逐渐从信息社会向知识社会迈进，信息资源的开发和利用日益受到人们的广泛关注。

目前，信息资源的特点是海量、类多、源广，特别是计算机网络的逐步普及、数字文本复制的便利和自由发表的实现，使得信息资源数量急剧增加，但是社会信息量的增长并不意味着用户获取的信息量的增长，恰恰相反，无序的信息资源不仅无助于信息资源的使用，反而会加剧信息增长与使用的矛盾。人们生动地称这种情况是“信息超载、知识缺乏”。越来越多的人认识到“原始信息本身并不能产生价值”，只有将其有效组织，按特定的需要集中和揭示，才能产生价值。要有效开发利用信息资源，必须采用相应的方法加以控制和处理，信息组织是关键措施之一。

通过本章学习，读者可以了解（或掌握）：

- ❖ 信息组织的概念
- ❖ 信息组织的原理
- ❖ 信息组织的类型
- ❖ 信息组织的作用
- ❖ 信息组织的发展

## 1.1 信息组织的界定

某种意义上，信息是推动人类社会发展的直接动力，这种动力在人类认识和改变的世界过程中起到至关重要的作用，已成为现代社会生产力的基本要素。为了更好地发挥信息的积极作用，高效的信息管理势在必行。在技术层面，信息管理的主要目的是通过手工、机械、智能（特别是计算机）方式，对信息进行收集、甄别、加工、处理、存储，使之序化，便于快速检索并提交至有特定需求的用户。对信息的加工、处理等序化工作则是信息组织的核心内容。因此，信息组织是信息管理的有机构成和重要环节。

现代社会信息的两个重要特征是庞杂性和分散性，人们的信息需求则呈现领域的高度选择性、内容的优质性、格式的易用性和时间的紧迫性等特点。两者之间的联系和矛盾的调解依赖于信息组织。信息组织是一个信息重构和增值的过程。在这个过程中，杂乱无章的原始信息变成一个有序的、优质的信息集成系统，一个相对“粗放”型的信息贫集将转化为一个“集约”型的信息富集，并为信息的进一步增值（如信息的分析研究）奠定基础。

### 1.1.1 信息组织的概念

“信息组织”（Information Organization）是随着社会信息化而出现的一个专业术语。人类的生存和发展离不开信息，早期的信息获取主要靠人与人之间的直接交往。有了文字后，人类在社会生产和生活中越来越多地依赖于文献信息。随着文献的大量出现和急剧增长，便有了文献的整理加工工作，被称为文献组织或情报组织，为人们便捷获取特定文献信息提供了可能。

随着现代科学技术的发展，人类社会逐渐从信息社会向知识社会迈进，信息资源的开发和利用日益受到广泛重视。目前，信息资源的特点是数量大、种类多、来源广、格式杂，特别是计算机网络的全球普及、数字信息复制的便利和自由发表的实现进一步显化这些特点。但是，社会信息量的增长与用户获取和使用的信息量的增长并不具备同步性，恰恰相反，无序的信息资源不仅无助于信息资源的使用，反而会为用户获取相关信息设置障碍，加剧信息增长与使用的矛盾。人们生动地称这种现象是“信息超载、知识缺乏”。越来越多的人认识到“原始信息本身并不能产生价值”，只有将其有效组织和整合，按特定的需要集中和揭示才能产生价值。从用户需求角度，多角度、多层次的信息组织将有利于满足用户多元化的信息需求。因此，要有效开发和利用信息资源，必须采用相应的方法对社会信息加以控制和处理，信息组织是其中的关键措施之一。

什么是信息组织？要正确了解信息组织的概念，首先必须了解数据、信息、信息体和知识的含义及其区别。所谓数据，是数字、字母和符号的集合，系客观事物与主观思维的具体表达，也可以说，数据是对声音、语言、体态、信息、符号、文字的一种表现形式的认识。所谓信息，广义是指客观事物存在、运动和变化的方式、特征规律及其表现形式，狭义是指用来消除随机不确定性的信息。需要明确的是，本书讨论的信息组织对象并不涵盖所有的信

息，不包括本体论信息和认识论信息，而是指可以组织的记录形式的信息、数据和知识，即信息体。在大数据时代，信息体包括的不仅是文本的数据、信息和知识，也包括非文本的数据、信息和知识（如音频、视频、图像、图形和网页）。尽管在现实生活中，数据和信息两个词经常替换使用，但严格地说，数据和信息是两个不同的概念。数据是按照一定规则排列组合记录信息的物理符号，是狭义信息和知识的基础。狭义信息是数据载荷的内容，同一信息的数据表现形式可以多种多样。信息与经验（知识准备）相结合则可转化为知识。知识是信息接收者通过对信息的提炼和推理而获得的正确结论。在一定语境下，信息与知识可以互换。在这个意义上，我们日常所说的许多信息系统和知识系统从本质上都是数据系统。

另外，信息组织与知识组织是既有联系又有区别的两个概念。信息组织是由文献组织发展而来的关于信息的组织与检索系统。文献是指以硬载体（如纸张等）为依托的知识或信息。但由于现代信息数量急剧膨胀，信息类型日趋复杂化（如脱离纸张的计算机数字信息），传统的文献组织工作已经无法对信息加以有效的控制和管理。随着信息技术的飞速发展、计算机和网络的广泛普及，一方面，人们开始将原始文献信息集合转化为有序、优质的计算机信息系统的数据库或文献信息资源的索引；另一方面，人们面临着如何对直接产生于计算机或来自网络的数字信息进行组织的问题。因此，以印刷型信息为主要对象的文献组织必然发展为以各种类型信息为对象的信息组织。知识组织是在文献分类法的基础上发展起来的，是关于知识组织和检索系统在现代网络信息环境下获取和利用知识的所有手段、技术和能力的总和，与文献组织和信息组织关系密切。信息组织对信息的有序排列是知识组织的原料基础，知识组织是信息组织发展到一定阶段的必然产物，比信息组织的内涵更深刻、丰富和明确，更能反映用户需求的实质，但其任务更艰巨。

目前，国内学者对于信息组织定义的表述各不相同，比较有代表性的有以下几种。

马张华认为，“信息组织，亦称为信息资源组织，是根据信息检索的需要，以文本及各种类型的信息源为对象，通过对其内容特征等的分析、选择、标引、处理，使其成为有序化集合的活动。”

马费成认为，“信息组织也称为信息整序，是利用一定的规则、方法和技术对信息的外部特征和内容特征进行揭示和描述，并按给定的参数和序列公式排列，使信息从无序集合转换为有序集合的过程。信息组织通过人工和机器干预，使信息有序增值，进而提供有效利用。”

柯平等认为，“信息组织是根据信息的内容特征和外部特征，采用一定的原则和方法，对信息进行加工处理，使之有序的过程。”

周宁认为，“信息组织是对信息资源对象进行收集、加工、整合、存储，使之有序化、系统化的过程。”

戴维民认为，“信息组织是为了方便人们检索、获取信息而将庞杂、无序的信息进行系统化和有序化的过程。广义上，信息组织的内容包括信息搜集与选择、信息分析与揭示、信息描述与加工、信息整理与存储。”

储节旺认为，“信息组织就是人们根据信息本身特点，运用适宜的工具和方法，依据一定的标准或规则，对其进行加工整理，排序组合，使之有序化、系统化、规范化、高级化，增强信息对象的表现效能和运用效能，以满足人们信息需求的过程。”

以上定义虽然表述不太一样，但实质上是一致的。在社会网络环境下，考虑到用户已经参与到信息组织中，我们认为，信息组织是以用户需求为导向，依据信息体自身的属性特征，

信息工作者或用户按照一定的原则、方法和技术，将杂乱无章的信息整理成为有序的信息集合的活动和过程。信息组织的结果是形成各种方便用户利用的有序化的信息检索系统，从而达到信息增值的目的。信息组织是信息管理活动的核心和基本环节。

社会信息作为一种公共交流的知识，是通过不同的形式进行传播的，包括口头方式、纸质方式、数字方式等。各种信息载体都是信息存在的形式，也是交流的主要形式。以记录形式进行信息交流能够克服面对面直接交流的时空局限，使得人类社会可以在继承已有经验的基础上向前发展，对人类社会的发展意义重大，是人类社会进步的重要条件。各种类型的信息资源既是产生公共知识的基础，又是社会发展的智力资源，是信息组织的主要处理对象。

需要说明的是，作为信息组织的处理对象，信息、信息资源或信息体在本书中与文献或资料的含义接近，通常是指一切以记录形式存在的信息载体，只是更侧重于对新型媒体形式的强调。信息范围广泛、种类多样，就信息的基本类型而言，可以是图书、期刊、报纸、会议文献、标准文献和档案等类型；就存在的形式而言，可以是各种传统的印刷型信息资源，也可以是缩微型、机读型资源、实体/非实体信息资源——如网络信息资源（或称数字信息资源）等。信息组织就是以各种信息媒体形式为对象进行的组织。就其处理的单元而言，信息组织存在着不同的层次：可以直接以信息资源的存在单元为处理对象。例如，可以以某图书、期刊、网站等为处理单元；也可以以期刊、报刊、网站中的个体，如论文、新闻、网站中的构成部分等为处理单元；还可以直接以其中的信息成为处理单元，进行组织和揭示。本书主要讲述记录型信息，包括网络信息组织（数字信息组织）。

需要强调的是，信息组织是信息检索与利用的基础。没有信息组织，就不会有信息检索系统，也不会有信息检索，从而也就谈不上对信息的有效利用。信息组织为人们获取信息提供了方法和工具。信息组织的目的是建立起信息资源收藏系统和检索工具，方便信息资源的开发和利用。信息组织是一种为了信息检索的需要对信息资源进行有序化组织的活动，是与信息检索活动密切联系的。

信息组织是信息管理活动的必然要求，起源于信息本身的自然无序状态。序是事物的一种结构形态，是指事物或系统的各结构要素之间的相互关系及这种关系在时间和空间中的表现，即事物发展中的时间序列及排列组合、聚类状态、结构层次等空间序列。当事物结构要素具有某种约束性且在时间序列和空间序列呈现某种规律性时，这一事物就处于有序状态，反之则处于无序状态。

信息的无序状态有两种类型：一是信息内容无序，即组成信息的各语言要素，如字、词、句、段落或章节等处于一种杂乱无章的状态，或是组成信息的各内容要素，如命题、观点、认识、推理等处于一种分散无序的状态，或是组成信息的各载体要素处于一种零散错位的状态，这种信息无序无法准确、科学地表述信息内涵、形成信息实体，也就无法被信息用户理解和应用；二是信息体及其相互之间的组织无序，即不同信息个体处于彼此毫无关联的自然状态，缺乏科学稳定的框架体系，没有信息的深层次加工，无法融合成一个可以科学地把握其信息内容、有效地查检其信息内容、完善地维护其信息形态、良好地排列其信息形态、充分地实现其信息价值的有序的信息集合体。

显然，本书中的信息组织要解决的主要第二种信息无序问题，第一种信息无序问题主要通过信息生产和信息搜集工作来解决。也就是说，信息组织是在信息搜集基础上进行的信息系统的信息整理和序化工作。一般，信息处于有序状态或无序状态是就一定参考系而言的。

信息组织是由来已久的一种人类社会实践。其在发展过程中，不断从相关学科的理论和方法中汲取营养，使自身逐渐得到完善。

### 1.1.2 信息组织的原理

信息组织全称“信息资源组织”（信息资源是可供人们直接或间接开发与利用的信息集合的总称），是使信息有序化和优质化的过程，即利用一定的科学规则和方法，通过对信息外在特征和内容特征进行揭示和描述，实现无序信息流向有序信息流转换，从而使信息集合达到科学组合，实现有效流动，促进用户对信息的有效获取和利用。本质上，信息组织是基于信息属性的一种整序方法。按照信息属性特征排列信息的工作称为信息整序。信息整序既是理论问题，又是方法问题。整序是信息组织工作的基本手段和直接目的。众多的、杂乱分布的信息只有经过整序，才能得以方便地、有效地利用，最大化地发挥信息效能。

信息组织的基本原理是：针对若干自然状态的无序信息资料（如若干图书或学位论文），将这些信息资料按照其某种属性特征（如著者、题名、类属、主题）排列成一个序列，并且需要使用信息资料的用户能将自己的信息需求转换成相应的信息资料属性特征，在排列后的信息资料序列中找到自己所需的信息资料。

信息整序通常是对信息记录的载体进行的，然而为了信息组织和检索的方便，信息组织的直接对象往往是信息载体的替代记录（亦称元数据）。替代记录反映了原信息载体的主要外部属性特征和内部属性特征，但在形式上比原始信息载体更简洁。如图 1-1 所示，《国家数字图书馆服务框架研究》一书的在版编目数据就是该书的替代记录和元数据；又如，图 1-2 是利用百度搜索该书的网页快照，也是该书的替代记录。

国家数字图书馆服务框架研究/张炜主编.——北京：国家图书馆出版社，2012.5  
ISBN 978-7-5013-4708

I. ①国… II. ①张… III. ①中国国家图书馆—数字图书馆—图书馆服务—研究 IV. ①G259.251

图 1-1 替代记录之一：图书在版编目数据



图 1-2 替代记录之二：百度搜索网页快照

替代记录和原始信息存在密切的关系。替代记录在信息组织和检索中的作用可以通过图 1-3 来说明。

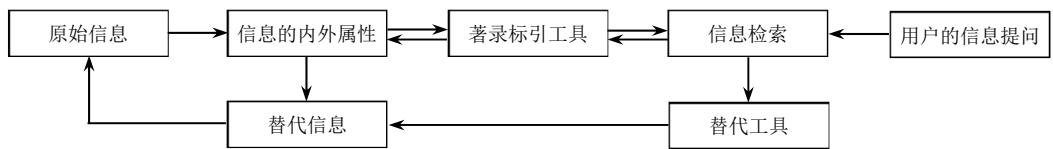


图 1-3 替代记录及其作用

信息组织与检索的实际过程是：信息工作者（含部分用户）在拿到原始信息后，根据信息组织工具对其进行加工，形成替代记录，并将替代记录存储在检索系统中。检索时，用户（或检索操作员）根据信息组织与检索工具对用户的信息提问进行加工，形成替代问题。检索的核心过程是通过替代问题和替代记录的匹配比较进行的。在找到与替代问题相关的替代记录后，再根据替代记录与原始信息的对应关系，查找到原始信息，以解决用户的原始问题。在网络信息组织中，这种替代机制同样存在。例如，可以通过搜索引擎概要结构了解搜索引擎内部的信息组织和替代机制，如图 1-4 所示。

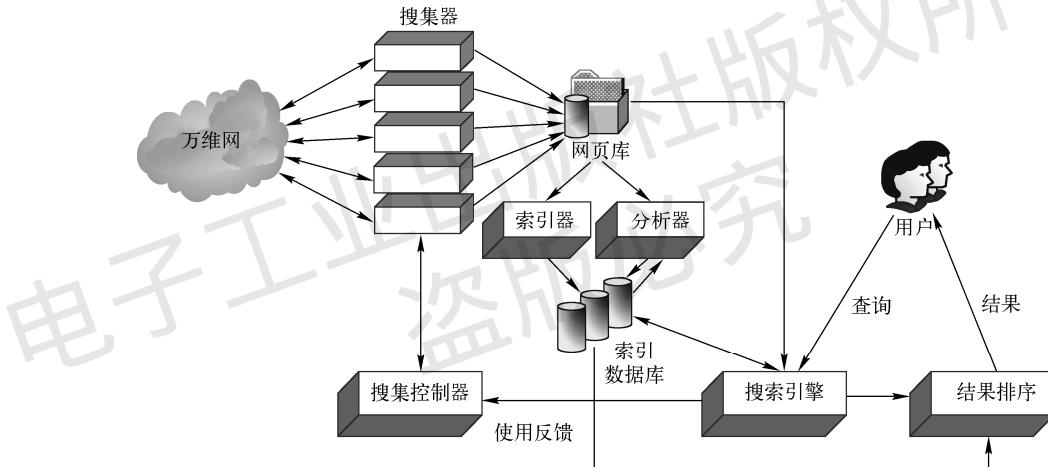


图 1-4 搜索引擎概要结构

### 1.1.3 信息组织的内容

信息组织一般包括信息筛选、信息分析、信息描述和信息揭示、信息序化和信息存储等内容。其中，信息筛选和信息分析是信息组织的基础准备工作，信息描述和信息揭示是信息组织的核心，信息序化和信息存储是前几个环节形成的有序信息集合的空间组织。信息组织通过对信息资源的序化和质化，最终实现向用户提供优质信息资源服务的目的，因此广义上，信息组织也可以归入信息服务的范畴。事实上，在社会网络环境下，信息组织与信息服务已经逐渐融为一体。

#### 1. 信息筛选

信息筛选是按照一定的判别标准从搜集到的、处于无序或部分无序状态的信息流中甄别出有用信息并剔除无用信息的信息活动，是整个信息组织过程的第一步。

## 2. 信息分析

信息分析是在信息筛选的基础上，按照特定的逻辑关系，从语义、语用和语法上对信息资源的内外部信息特征进行细化、挖掘、加工、整理并归类的信息活动。信息分析是信息描述和信息揭示的前提和基础，直接影响着信息组织的质量。

## 3. 信息描述和信息揭示

信息描述是根据一定的描述规则和技术标准，对信息资源的部分主题内容、形式特征及物质形态等描述维度进行分析、选择、记录的过程；信息揭示是对信息资源的内容特征进行深层揭示并转换成主题标识系统的过程。信息描述和信息揭示分为两种：一是著录，主要是对文献信息的外部属性特征（题名、著者、出处等）和部分内部属性特征（类属、主题、摘要等）进行描述的过程；二是标引，主要是对信息的内容特征给出信息内容标识（如分类号、主题词等）的揭示过程。信息描述和信息揭示是信息组织的重要内容，在信息组织中起着至关重要的作用。

## 4. 信息序化和信息存储

信息序化是在信息描述与揭示的基础上，按照一定的规则和方法，将信息资源描述和信息揭示记录（款目）进行有规律的组织排列，从而构成某种序列（亦称某种目录或索引，如著者目录或索引）的过程。信息存储是将经过加工整序后的信息资源按照一定的格式和顺序存储在特定的载体中的信息活动。信息存储是为了信息管理者和信息消费者快速、准确地识别、定位和检索有用信息。各种文献信息检索工具、光盘检索系统、网络信息检索工具等都是信息存储的重要方式。

# 1.2 信息组织的类型

信息源是人们生产生活、科学研究和其他一切活动中所产生的成果和各种原始记录，以及对这些成果和原始记录加工整理得到的成品。信息源是信息组织的对象，因此信息组织的类型基于信息或信息源进行分类。本文按照信息的表现形式、信息的加工程度、信息的传播载体、信息的认识层次和信息的存在环境对信息组织的类型进行划分。

## 1.2.1 按信息的表现形式划分

按信息的表现形式划分，信息组织可以分为文字信息组织、图像信息组织、音频信息组织、视频信息组织。

### 1. 文字信息组织

文字是人们为了记录事物、实现信息交流与通信联系所创造和约定的一种形象符号系统。广义的文字还包括各种编码，如 ASCII、国际电报与单元代码及计算机中的二进制数字编码等。这些文字、符号、代码均是信息的表述形式，其内容再现于它们的结构属性中。例如，

基本笔画的不同组合、字和字母的不同组合、二进制码“0”和“1”的不同排列等，分别代表不同的信息内容。这些文字信息是通过文字表达的信息内容的语义、语用和语法来组织的。

## 2. 图像信息组织

图像是客观对象的一种相似性的、生动性的描述或写真，是人类社会活动中最常用的信息载体，也是人们最主要的信息源。图像作为一种视觉信息，比文本信息更直观。图像本身有分辨率、形状、大小、颜色、深度、饱和度、亮度及存储格式的属性特征。在通常情况下，通过表现图像的各种属性，可以按照视觉的需要来组织图像信息，也可以按照对图像的存储格式的需要来组织。

例如，为了提高传输效率，Web 网页上的图像一般使用图形文件交换格式（Graphics Interchange Format, GIF）定义的 GIF 格式文件和联合图像专家组（Joint Photographic Experts Group, JPEG）定义的 JPG 格式文件，有时使用可移植的网络图像文件格式（Portable Network Graphics, PNG）定义的 PNG 格式文件。GIF 格式文件存储的图像是无损压缩的图像；PNG 格式文件是作为 GIF 的无专利替代品开发的，在 Web 上无损压缩和显示图像；JPG 格式的图像是有损压缩的图像，不过某些 Web 浏览器不支持 PNG 图像，而大多数 Web 浏览器不需要调用其他图像处理程序就能直接显示 GIF 和 JPG 图像。通常情况下，连续色调图像（如照片）应压缩为 JPG 格式，具有单调颜色或锐化边缘及清晰细节的插图（如文字）应压缩为 GIF 或 PNG 格式。

## 3. 音频信息组织

声音和语言是人类交流中最普遍使用的方式。声音的基本参数包括基音周期、共振峰、语音谱、声强等，这些参数由声源、声道和放射机构（嘴唇和鼻孔）导出。声音可以进行编码和压缩，可以通过 A/D 转换进行数字化并输入计算机处理，再通过 D/A 转换输出。我们一般按照听觉的需要或声音的记录方式来组织音频信息。

## 4. 视频信息组织

一般意义上，视频信息可分解为一系列静态图像信息和声音信息，可以看成无数张图像帧按照一定的流向（顺序）排列并外加音频组配而成，对视频信息的组织实际上是对图像和音频信息的组织。视频信息由于具有时空冗余性，压缩效果受到较大限制，最常用的视频信息压缩技术为 DVD、MPEG-2（卫星直播电视用）和 MPEG-4（因特网传输用）。从单个视频整合的角度，视频信息组织主要通过视频编辑软件（如 Adobe Premiere、iMovie 和 Cyberlink 等）实现对视频信息非线性的分割、合并、转换、输入、输出，集合图片、音频、视频、文字等信息。一般意义上，对视频信息进行信息组织主要是从视频信息数据中抽取各种标识性辅助信息，如视频所有者、主题标签、时间长短、容量大小、画面清晰度等属性信息，达到特定目的的视频信息有序化。目前，基于内容的视频信息组织已被广泛应用。

### 1.2.2 按信息的加工程度划分

按信息的加工程度划分，信息组织可以分为一次信息组织、二次信息组织、三次信息组织。

## 1. 一次信息组织

一次信息是人们研究或创造性活动成果的直接记录，一般指公开出版的图书、论文、研究报告、会议文献、学位论文、专利文献和计算机网络上未经过再加工的数字化信息等。一次信息组织是采用一定的逻辑结构和语言规则，将大量零散的信息单元组织成能够反映信息所有者对事物和概念认识的一次信息（如图书、论文等）的信息组织方式。当前数字格式的一次信息大量存在，如计算机中的文本文件、图形文件、音频/视频文件、互联网的网页等。一次信息一般具有面广、量大、分散无序和良莠不齐等特点，有时很难获取，即使在网络上搜索，其查准率也非常低，逻辑相关性非常差。因此一次信息一般需要进行信息整序才能方便用户使用。

## 2. 二次信息组织

二次信息是对一次信息加工后产生的信息，如目录、题录、简介、文摘、索引和书目数据库等。二次信息组织是在一次信息的基础上加工整序的信息组织方法，主要采用选择、提炼和浓缩等方法，编撰和建立各种形式的目录、文摘、索引等检索工具与存取系统，其主要功能是规范和控制信息的流向，组织序化文本与数字化信息。二次信息组织的目的是建立各类信息检索系统，通过对一次信息的内容属性和外部特征进行有效揭示、描述、排序和存储来实现。信息的外部特征主要指信息记录的载体直接反映出来的特征，如题名、著者、信息类型、语言、时间和地点等；信息的内部属性主要指信息记录所载知识内容的学科类属性和主题属性。二次信息的形成是信息从分散、无序到集中、有序的控制过程。它的重要作用不仅在于报道，更在于为查找一次信息提供线索。

二次信息组织的产品很多，它们是人类分享信息的主要工具，可以以印刷型文献的形式出现，如传统的书目、索引、文摘等；也可以以数字化的方式出现，如各类书目型数据库（包括题录型和文摘型），互联网上的搜索目录和搜索引擎都是提供一次信息线索（包括网址和网页）的二次信息检索工具。

## 3. 三次信息组织

三次信息是根据特定的需求，对一次信息和二次信息进行加工、分析、改编、重组和综合概括生成的信息。根据其编制目的和方式，三次信息产品可以分为综合（或专题）研究型和参考工具型两大类，前者如专题述评、总结报告、动态综述、信息预测等，后者如手册、百科全书、年鉴、指南等。三次信息组织是在一次信息组织、二次信息组织的基础上，根据某学科、专业、主题或人物（类型）等需求，鉴别收集相关一次信息和二次信息，同时结合相关的数据统计与分析、数据挖掘方法和工具，以整序筛选、浓缩和综合概括，形成信息综述、信息汇编、信息述评或专题报告等的高层次信息组织方法。经过三次信息组织得到的信息产品在科技研发、企业竞争和政府决策等领域中发挥着重大作用。

### 1.2.3 按信息的传播载体划分

按信息的传播载体划分，信息组织可以分为文本信息组织、多媒体信息组织、富媒体信息组织。

## 1. 文本信息组织

一般，对文本的解释包括三方面：在语言文学领域，文本是一系列语句串联而成的连贯序列，通常是具有完整、系统含义的一个句子或多个句子的组合，一个文本可以是一个句子、一个段落或者一个篇章；印刷型文献也是文本信息，包括图书、期刊、会议资料、专门报告、专利资料、政府出版物、学位论文、产品样本、档案、标准和报纸等；在计算机领域，文本主要指文本文档，用于记载和存储文字信息，常见文本文档的扩展名有.txt、.doc、.docx 和.wps 等。显然，本书涉及的文本信息组织的对象是后两者。

印刷型文本信息组织以分类组织法和主题组织法为主要形式。分类组织法是语法信息组织和语义信息组织的综合，从学科角度集约信息，便于族性检索。主题组织法是建立在自然语言基础之上的语义信息与语法信息组织的综合，其词族索引和范畴索引展现了主题词之间的等级关系和学科关系，属于语义信息组织，而附表和英汉对照索引体现了语法信息的关系，主题组织法从事物角度集中信息，便于特性检索。计算机文本文档的信息组织则以计算机编码系统为准，主要通过对文本文档题名、作者、时间、格式、大小和存储地址等属性来进行标识。

## 2. 多媒体信息组织

多媒体信息组织是指利用多媒体工具、多媒体语言工具等手段以及各种信息组织方法，对文本、图形、图像、音频、视频等多媒体信息进行集中处理，以增强信息的表现力，提高信息的使用价值。多媒体信息组织广泛存在于电子出版物、多媒体教学软件等的设计和制作以及网络信息管理中。多媒体著作工具是能够集成处理和统一管理多媒体信息，使之能够根据用户的需要生成多媒体应用系统的工具软件，也被称为多媒体创作工具或多媒体写作工具等。多媒体著作工具提供了组织和编辑多媒体应用系统各媒体元素所需的框架，可以设计交互性用户界面，可以将各种多媒体信息组合成一个连贯的节目。

## 3. 富媒体信息组织

富媒体（Rich Media）是指具有动画、音频、视频和交互性的信息传播方法，包含下列常见的形式之一或者几种的组合：流媒体、音频、Flash 和 Java、JavaScript、DHTML 等程序设计语言。富媒体信息广泛应用于各种网络服务中，如网站设计、电子邮件、弹出式广告和插播式广告等。富媒体信息组织基于网络信息资源和网络信息技术，按照特定原则和方法收集、选择、描述、标引和存储富媒体信息资源，以更好地序化和优化富媒体信息资源。富媒体信息组织方法主要包括以下几种。

### (1) 超文本法

超文本是一种新型的信息组织方法，一大特征是信息的非线性排列，以节点为基本单位，节点间以链相连，将信息组织为某种网状结构。用户可以从任一节点开始，根据网络中信息间的链接，从不同角度浏览和查询信息。超文本法提供的非顺序性的浏览功能比传统的信息组织方式更加灵活、方便，且符合人们的联想思维方式。超文本的另一大特征是信息表达形式的多样性。超文本信息可以是文字、图形、图像、音频和动画等多种媒体形式，因此可称为“超媒体”。

### (2) 自由文本法

自由文本法主要用于全文数据库的组织，是对非结构化的文本信息进行组织和处理的一

种方式。自由文本法无须前控，不必用规范化语言对信息进行复杂的前处理，不是对富媒体信息特征的格式化描述，而是用自然语言深入揭示其知识单元，根据富媒体的自然状况直接设置检索点。自由文本法能够完整地反映出富媒体的全貌，通过计算机自动进行文献处理和组织。基于全文数据库的全文检索可以将任意字符作为检索标识，用户用自然语言即可直接检索未经标引的原始信息。

#### (3) 搜索引擎法

搜索引擎是网络上专门提供查询服务的一类工具，利用被称为 Robot、Spider 或 Worm 等名称的自动化处理软件，定期或不定期地在网络上爬行，通过访问网络中公开区域的每个站点，对富媒体信息资源进行收集，然后利用索引软件对信息进行自动标引，创建一个详尽的、可供用户按照关键词查询的 Web 页面索引数据库。这种数据库的内容一般有标题、摘要或简短描述、关键词和 URL、文件大小、语种及词出现的频率和位置等。搜索引擎法是目前网络上对富媒体信息进行组织的主要方式之一，著名搜索引擎有 Google、Baidu、AltaVista、Excite、WebCrawler 和 Lycos 等。搜索引擎法收集的信息虽然丰富广博，但良莠不齐，因而查准率低。

#### (4) 主题树法

主题树法是将富媒体信息按照某种事先确定的概念体系分门别类地逐层加以组织，先通过浏览的方式层层遍历，找到所需的信息线索，再通过信息线索连接到相应的信息资源。许多著名的网络检索工具如雅虎、搜狐等采用这种方式组织信息资源。主题树法提供了一个基于树浏览的简单、易用的网络信息检索与利用界面。信息检索由用户按照规定的范畴分类体系，逐级查看，按图索骥，目的性强，查准率高。采用树形目录结构组织信息资源具有严密的系统性和良好的可扩充性。当然，该方法也存在缺点，最突出的就是必须事先建立一套完整的范畴体系，由于富媒体信息资源的庞杂性及信息用户对体系了解的影响，无疑增加了用户的智力负担。另外，为了保证主题树的可用性和结构的清晰性，范畴体系的类目及每个类目下的信息索引条目都不宜过多，这大大限制了一个主题树体系能容纳的信息资源的数量。

#### (5) 学科信息门户组织法

学科信息门户（Subject Information Gateway, SIG）是针对特定学科或主题领域，按照一定的信息选择和评价标准及规范整合的资源描述和组织体系，对具有一定学术价值的网络信息资源进行搜集、描述和组织，并提供浏览、检索和导航等基本功能。学科门户组织法是由专业人员搜集富媒体信息，对信息源进行质量鉴别，使用受控语言和关键词进行必要的内容描述，从学科属性、研究方向和资源类型多重角度揭示。学科信息门户组织法融合了学科链接导航、搜索引擎法、主题树法、语义信息组织法和 Web 2.0 信息组织等技术和方法，是具有较高学术价值的综合性数据服务平台。

### 1.2.4 按信息的认识层次划分

从人类对信息的认识层次上，由于人类具有感受力，能够感知事物运动状态及其变化方式，由此获得的信息为语法信息；由于人类具有理解力，能够领悟事物运动状态及其变化方式的逻辑含义，由此获得的信息为语义信息；又由于人类具有明确的目的性，能够判断事物运动状态及其变化方式的效用，由此获得的信息为语用信息。对这三种信息进行组织就是信息组织的基本方法。

## 1. 语法信息组织

语言学中的语法学是研究语言符号之间的结构规则的，主要包括词典构成和变化、词组和句子的组织，研究的语言内容属于形式范畴。信息组织借用了“语法”二字，是指按照形式特征组织信息。最常见的语法信息组织有如下4种。

### (1) 字顺法

字顺法是历史最悠久、使用最广泛的一种信息组织方法。它的意义在于从字、词角度集中排列有关信息，可满足用户检索的一般要求。字顺法的具体操作有音序法、形序法和两法并用三种形式。从字顺法的发展史来看，目前音序法逐渐占据了主导地位。

### (2) 代码法

信息用代码集中排列，人们易于接受又便于管理，所以随着信息量的激增和信息多样化，代码法也从无到有，日益显示出其重要性。代码一般使用拉丁字母和阿拉伯数字，如专利代码和商品代码。

### (3) 地序法

获取有关地域方面的信息是用户需求的一个重要方面，因为地域本身就代表了某方面的特殊信息。同时，跨地域的比较研究日益增多也促使地域信息需求的增多。地序法一般有文字法和图文法两种形式。

### (4) 时序法

获得历史信息和从历史角度获取信息历来是大众的普遍需求。时序法有广泛的应用范围，如编纂工具书、著书立说，甚至写一篇文章等。

## 2. 语义信息组织

语义学（Semantics）中，语义具有研究语言符号与它代表的对象之间的结构关系之义。语义信息组织是专门研究信息的内容特性的信息组织方法，常见的语义信息组织有如下2种。

### (1) 分类法

人们的分类对象可概括为三种：一是实物，如商品；二是概念，如知识；三是概念与实物的结合体，如文献。于是有三种分类方法，在信息组织的实践中，可以结合使用。知识分类法是人类认识客观世界的科学方法，对其他两种分类法有着指导意义。文献分类法以知识分类为基础，结合了文献实体属性和信息利用的实际需求。实物分类法带有更多的专业（或行业）特性和效用原则。

### (2) 主题法

主题法是字顺法在语义信息中的特殊应用，既采纳了字顺法直截了当、便于检索的优点，又兼顾了相同内容聚集的特点，是人们从内容角度更直接获取信息的有效方法。最常见的主题法有标题法、单元词法、叙词法和关键词法等。

## 3. 语用信息组织

语用信息是借助语用学的特有含义来研究随环境与使用者的不同而不断变化的信息群。常见的语用信息组织法有如下2种。

### (1) 权重值信息组织法

权重值信息组织法是按照信息的重要性来组织信息。例如，报纸在版面安排上，总是把

最重要的信息放在头版头条的位置；电视节目的安排，总是把重要节目放在黄金时间播出。城市规划、行业决策和质量评估等其他方面也常用到这种方法。

## (2) 概率组织法

概率组织法是在未全知信息的情况下对信息进行组织的方法，根据事件发生的概率大小序化信息，可以在一定程度上规避风险，如预测文体活动胜负、期货交易等。

### 1.2.5 按信息的存在环境划分

在我国信息化建设过程中，建立和逐步完善一个能够使信息资源得以充分开发和有效利用的信息环境具有重要意义。信息环境是指与人类信息获得有关的一切自然、社会和心理因素的总和。信息环境主要包括与人类信息活动有关的人的要素、信息的要素、技术的要素和社会诸要素（社会、政治、经济及政策法律等）。可见，信息存在环境既包括技术环境，也包括社会环境。如从是否使用网络技术的角度，信息可以分为网络信息和非网络信息。从信息存在的社会环境的角度，信息组织类型可以有以下4种。

#### 1. 政务信息组织

政务信息，也称为政府信息，是涉及国家政府各级部门和单位的一切制度及事务范畴的信息总和。政务信息为政府部门决策提供依据，在信息化和网络化高度发达的当今社会，政务信息组织对政府部门具有举足轻重的意义。政务信息组织应以有利于政务信息的“公开、公正、及时、透明”原则为准，做好政务信息的选择、分析、描述、揭示和存储等工作。目前，电子政务是政务信息组织的主要方式，是政府机构应用现代信息技术，将政府管理和服务通过信息技术进行集成，在互联网上实现政府组织结构和工作流程的优化重组，超越时间、空间与部门分隔的限制，全方位地提供优质、规范、透明和符合国际水准的管理和服务。

#### 2. 经济信息组织

经济信息是反映经济活动实况和特征的各种消息、情报、资料和指令等的统称。在宏观经济、中观经济和微观经济活动中都存在大量经济信息，反映和沟通各方面经济情况的变化，从而帮助人们调控和管理生产，实现管理环节间的联系。经济信息又分为计划信息、控制信息、生产与经营信息、统计信息等。通常，经济信息组织形式多是从市场调研与竞争情报等活动中开始的，也有部分表现为经济类文献数据库（如国研网等）。经济信息组织以具体的经济利益为目标，对相应经济信息进行组织与处理，形成高经济价值的信息资源。

#### 3. 科技信息组织

科技信息是记录科学技术发展的信息集合，集中反映各时代、各领域、各学科和各行业科学技术发展水平与方向。科技信息往往具有知识性、创新性、敏感性和机密性等特点。科技信息组织是以相应领域的科技信息数据库为主，包括中外文文献数据库，涉及电子图书、期刊、学位论文、会议文献、标准文献、专利文献、研究报告和技术说明书等形式，科技信息组织产品是目前信息组织最重要的实施对象。

#### 4. 文化信息组织

随着我国对公共文化服务的重视及其政策的逐步出台，文化信息的组织与管理日益凸显其重要性。文化信息积淀在人类文化发展整个历程中，记录一个国家或民族的历史、地理、风土人情、传统习俗、生活方式、文学艺术、行为规范、思维方式和价值观念等内容。目前，系统性的文化信息组织工作在我国仍处于初级阶段，但随着国家、组织或个人对文化信息组织的日益重视和研究探索，文化信息组织在未来将为国家文明建设、公共文化服务提升带来巨大能量。

当然，以上 4 种信息组织方式依据信息所存在的社会环境来区分，仅具有相对意义。其他存在环境下的信息及其组织方式也是大量存在。需要说明的是，非网络环境和网络环境的信息及其组织方式在前面已有介绍，在此不再赘述。以上各种类型的信息组织往往是相互联系、相互交叉、相互渗透的，在一定条件下可以相互转化。

### 1.3 信息组织的作用

由于现代信息技术和互联网络的迅速发展，信息生产、再造、存储和传播空前便捷，导致信息量急剧增加，随之而来的是信息质量参差不齐、信息污染日益严重、信息垄断与不公，“信息爆炸”和“知识爆炸”成为现代社会的重要特征。这些现象造成了人类与信息之间的两种矛盾：一种是知识和信息的海量性、无限性与人的精力、时间的有限性的矛盾，另一种是知识和信息的无序性、污染性与人类使用的选择性的矛盾。

美国未来学家奈斯比特在《大趋势》一书中曾说过，“我们淹没在信息中，却渴求知识。”为什么当各类信息像洪水一样向我们涌来时，我们仍然缺乏所需的信息呢？这是因为在知识经济和信息化社会中，“失去控制和无组织的信息不再是一种资源。”

具体说来，信息组织对于信息交流与共享、重组与增值有如下作用。

#### 1. 控制整序作用

随着人类文明的发展和人类对世界的探索，信息资源数量不断增加。尤其是 20 世纪 90 年代以来，在现代信息技术发展和互联网兴起的带动下，信息规模呈指数级增长，信息的混乱程度日益加剧，导致用户获取信息所需的成本越来越大。由于缺乏控制，信息集合不可能为用户提供高效快速的检索，信息数量越大，混乱程度越高，当这种混乱达到一定程度后，信息检索几乎无法进行。因此，手工检索和计算机检索都离不开信息组织对信息的控制和整序，网络环境下的信息组织的作用更为突出。当前，人们开发研制的各种网络信息组织和检索工具，对于网络信息的快速有效查询具有重要意义。

#### 2. 提升品质作用

随着信息数量的急剧膨胀，信息质量日益恶化，大量的无用信息、虚假信息和垃圾信息充斥着各类信息交流渠道，极大地妨碍了人们对有用信息、正确信息的吸收和利用。事实上，没有组织或不加控制的信息不仅不能创造财富，还会对人们利用信息构成严重的危害。信息组织的作用之一就是通过对海量信息的选择、评价、整理、集成和重组等一系列工作，提升

信息品质，把良莠不齐的信息加工成为精良有序和高度集成的信息检索系统，为信息增值（如信息的分析研究）打下基础。

### 3. 传播利用作用

信息组织克服了混乱的信息流带来的信息查询和利用上的困难，把分散的、无序的信息加工为系统的、有序的信息流集合，使其成为真正意义上的信息资源，一方面便于信息机构通过各种方式向用户提供信息服务，另一方面便于用户自主地、随时随地查找和利用信息，从而充分发挥信息的效用。信息的交流实质上是信息传播的一种方式，或者说是信息传播的结果。序化的信息有利于信息的传播，研究表明，按一定规律排列的信息更易于人们理解和记忆固化，因而传播效能会更大。例如，报纸上的广告经过分类处理后，由于查找方便，读者接触到它的可能性明显增加。

### 4. 节约成本作用

当前，信息获取与筛选的成本日益增大，用户的时间成本将大于信息管理和信息服务的价值。如果任其发展，将大大增加用户社会各项活动的总成本。信息组织将多余的和虚假的信息过滤掉，将有用的信息精选出来并加以序化，使得信息交流和存取等环节的效率不断提高，成本不断下降，从而节约了社会各项活动的总成本，使社会在同样的时间里能够创造出更多的财富。

## 1.4 信息组织的发展

作为信息活动的重要组成部分，信息组织是与人类社会同时产生并同步发展的。信息组织活动具有悠久的历史，在当今网络信息时代对信息组织需求的有力推动下，其理论研究方兴未艾，其实践工作不断丰富。

信息组织的发展过程是信息管理整体发展过程的组成部分。在发展阶段上，信息组织的发展与信息管理发展基本保持一致。由于在原始信息管理时期几乎谈不上任何形式的信息组织活动，信息的传播通常是不正规、自发性的。因此，在此仅把信息组织的发展区分为三个阶段，即古代信息组织时期、近代信息组织时期和现代信息组织时期。需要指出的是，信息组织发展阶段的演化并不是一个阶段对另一个阶段的全面否定，而是一种扬弃、一种理论与方法的完善和扩展。即使是最早的清册职能也并未在现代信息组织时期被彻底废止，而是被融进了一种更为综合的组织活动之中。

人类的信息组织活动源远流长。了解信息组织的历史渊源和发展沿革，可以使我们从更深远的角度认识今天的信息组织活动，集成和汲取先人在长期信息组织实践中积累的有益经验。同时可以使我们高瞻远瞩、明确方向，使信息组织的理论和实践在科学地解决现实问题的基础上具有前瞻性，适应不断发展的社会需要。

### 1.4.1 古代的信息组织

古代信息组织时期是指近代工业革命之前的漫长的历史时期。信息组织最早是在人类思

维认识领域中进行的，从广泛的意义理解，可以说，自从有了人类思维活动，就有了最原始的信息组织。随着人类社会活动的日益复杂和多样化，人类大脑的思维、分析、推理、归纳、类比和聚类等功能得到不断加强，人们不但能够在纷纭的信息中选择，将同类的信息加以集中，不同的信息加以区分，而且能够由此及彼，发现事物间的内在联系。人类大脑在对信息的分析、综合、比较、抽象、概括、类比、判断和推理等活动中形成了最原始的信息组织方法。正如美国学者沃尔曼（R.S. Wurman）所比喻的：“人完完全全是由存储器组成的，这种存储器需要具有一个组织框架，该组织框架来源于个人的洞察力——具有精密过滤镜的眼睛、拥有独特频率响应能力的耳朵，以及通过思考找到个人兴趣的内容间相互联系的大脑。”尽管后来许多信息组织的活动和研究远远超越了人的认识领域，而延伸至社会信息的生产、加工处理、存储、传播和检索的广阔空间，但一切信息组织方法都离不开人类思维活动这个基础，都是在此基础上发展起来的。

信息符号作为表达概念的标识、信息组织的工具，它的产生和广泛运用对信息组织的发展起到了关键作用。人们对信息的内容和形式的组织是通过对信息符号的组织完成的。人类最早用手势、体姿符号来表达和传递信息，后来出现了有声语言符号，它比体姿符号能表达更多的意义，不同的语言表达不同的意义，其中就包含了信息组织的过程。

为了克服体姿语言和有声语言不便保留、稍纵即逝的局限，我们的祖先很早就在寻求信息记录的形式，例如，曾采用“结绳”方法来记录事物的数量和特点：“事大，大结其绳；事小，小结其绳。”这是人类以实物作为符号来记录、描述事物的早期应用。后来，人们又开始用图画来表达思想、记录知识，并在此基础上产生了最初的文字。文字的产生和运用使信息可以脱离人脑而被记录、存储于外界物质载体上，使信息变得容易识别、存储、组织、传播和交流，这是人类认识和处理信息能力的一大飞跃。

承载信息符号的各种载体是信息组织活动的物质基础，随着社会生产力的发展，信息记录的载体材料和信息记录方式不断改进。人们将代表某种意义的符号刻记在龟甲、金石上，书写在绵帛、竹木上，或者记录于泥版、羊皮上，进而书写在纸张上，刻录在光盘、磁带上，为信息组织的社会化创造了条件。

## 1. 中国古代信息组织的萌芽与发展

原始的信息组织是自发的、不规范的信息组织。当人类的信息记录积累到一定数量，并被有意识地汇集、排序时，就开始出现社会信息组织活动的萌芽。在我国的殷商时代，有大量记录事物的甲骨文文献。据殷商考古发掘发现，窖穴中的甲骨文排列有一定顺序，甲骨上除了有卜辞之类的记录，还刻有占卜者或经管卜官邸名字。这说明，殷代文献的管理就已经包含着目录工作的雏形，可以看作我国古代目录工作的起源。

我国古代正式的信息组织活动主要体现于对文献的整理，通过“部次甲乙”来满足人们即类求书、因书究学的需要。

最早的文献整理记录可以追溯到春秋时期孔子整理“六经”及以后各种图书目录的编制。孔子做学问时有感于大量文献的零散无序、不便使用，于是对鲁、周、宋等国的文献进行了收集、选择、校订和分类编排，将虽已成书但分散混乱的文献重新加以组织，使之成为结构合理的有序整体。如孔子对《诗经》305篇的组织整理、对《尚书》按体裁的分类排序，开创了我国私人整理文献的先河。

汉代刘向、刘歆父子编制了我国第一部大型分类目录《七略》，解释了当时国家收藏的自先秦以来所有的图书，这是我国历史上第一次系统的、大规模的文献信息组织活动。刘氏父子将当时收集的所有藏书分门别类著录，按学科内容分为六艺略、诸子略、诗赋略、兵书略、数术略和方技略六大类，加上“辑略”部分共七类，形成了我国第一部综合性的国家书目。此后，各代又有多种有影响的目录出现，如魏晋南北朝、唐宋元明所编纂的官修目录、史志目录、佛经目录等。南朝王俭的《七志》、梁朝阮孝绪的《七录》采用了七分法，郑樵在《通志·艺文略》中创造了独具一格的十二大类三级类目的分类体系。在我国使用最久、影响最大、占主导地位的是四分法。西晋时，国家书目《中经新薄》最早提出四部分类方式，后逐步改易，至唐魏征所编的《隋书·经籍志》确立了经、史、子、集四部分类系统。自唐以来，各种官私书目大多采用了四部分类。清代《四库全书总目》收录了先秦至清初重要书籍 10254 种，著录书名、卷数、著者书籍来源并有内容提要，组织成经、史、子、集四部 44 类，代表了我国古代目录的最高成就。这些目录系统揭示了我国历史文化典籍及其所记录的知识，为存储、查找信息提供了方便。这时的信息组织活动已经从单纯的个人行为发展到既有个人又有国家组织的社会活动，其范围和影响力不断扩大。

在主要应用分类组织法的同时，字顺主题法和索引法也初露端倪。《四库全书总目提要》中将颜真卿看作采用声韵字顺的方法来排列事物主题的鼻祖，称“昔颜真卿编《韵海镜源》为以韵隶事之祖。”这说明，我国在唐代已出现了以事物或对象名称作标题的信息组织方法。

类书被视为我国主题法的滥觞。目录学家姚名达认为，类书可以视为主题目录的扩大，如果删其繁文，仅存书目，也就成为现代最先进的主题目录了。我国最大的类书《永乐大典》就是一部带有主题法性质的检索工具。全书共 22877 卷，采用了“用韵以统字，用字以统事”的组织编排方式，与主题法的字顺系统原则完全吻合。

中国古代索引是在字书、韵书和类书的基础上发展起来的，曾被称为“通检”“备检”或“串珠”等，形象地说明了它的作用和特征。魏代建安年间类书《皇览》就具索引功能，被认为是中国古代索引的起源之一。我国唐代林保的《元和姓纂》、宋代黄邦先的《群史姓纂韵谱》、陈思的《小字录》等即为早期的姓名索引，宋、明、清等时期还编纂了一些专书索引。特别是清代史学家、目录学家章学诚在其《校讎通义》等书中明确提出了一系列较为重要的索引理论和索引方法。

在实践的基础上，我国古代一些学者对信息组织的原理和方法进行了理论探讨和阐述。春秋末期，一些大思想家已经注意到事物分类的普遍性，从“类”的角度去认识事物和进行推理。如墨子不仅将“类”作为分析事物的依据，还把“类”与“故”（理由）联系起来，作为立辞和辩说的基本原则。宋代目录学家郑樵从理论上研究了文献的分类问题，提出了一系列类分图书的原则和方法。他在《通志·校讎略》中指出：“类书，犹持军也。若有条理，所多而治；若无条理，虽寡而纷”，“类例不明，图书失纪”，“类例既分，学术自明”，生动阐明了文献分类的意义。清代章学诚则进一步主张文献分类应随文献增长的变化和学术演变而改变，主张编制目录不仅要对文献分门别类地加以编排，还要“辨章学术，考镜源流”。他大力提倡采用“互著”“别裁”（相当于现在的参照法和分析著录法）的方法对文献进行著录，并阐述了索引的方法和功用，主张将各种文献中的人名、地名、官阶、书名等按韵编排，详著出处，以便分类时按韵查找。阐述这些思想的专著《校讎通义》成为我国古典目录学理论的集大成之作。

## 2. 国外古代信息组织的萌芽与发展

在国外，最原始的“目录”是以泥版的形式出现的。约公元前3000年，在两河流域巴比伦王朝的一座寺庙的附近，有大批泥版文献被集中在一起，并按主题排序。这说明，当时对这些记录有信息的甲骨和泥版的收集和储藏是有意识的、经过整理的，已有了简单的标识著录法和排列法。

在古希腊，一些哲学家很早就提出了知识分类思想，并构筑了自己的知识分类体系。例如，亚里士多德将知识区分为理论知识（逻辑学、物理学、数学、形而上学）、实践哲学（伦理学、经济学、政治学）、创造哲学（史学、修辞学、艺术）三大门类，对后世分类法中学科的划分产生了深远影响。中世纪的欧洲修道院图书馆和大学图书馆分别建立了以宗教观点和学科设置为基础的分类体系。印度的一些古老经文如《吠陀经》将人类全部知识分为四大门类，也包括许多朴素的分类思想。

古代国外的正式分类活动主要可以追溯到公元前6世纪。当时亚述巴尼拔皇宫收集了大量的泥版文献，上刻主题标记，并分为人间、天上两大类，其下再分子类。公元前250年左右，古希腊学者卡利马科斯（Callimachus）为当时规模最大的亚历山大图书馆编制了长达120卷的藏书目录《皮纳克斯》（Pinakes，又名《各科著名学者及其著作目录》），将藏书分为戏剧、诗歌、法律、哲学、历史、修辞学、医学和杂著等大类，其下再按字母或年代顺序排列，并附对每部著作的评价，成为古代最早的目录之一。

被称为“目录学之父”的德籍瑞士学者格斯纳（C. Gesner）将当时能搜集到的全部拉丁语、希腊语和希伯来语的各种著作1.5万种编织成大型书目——《世界书目》，全书四卷，包括著者字顺目录、分类目录和主题字顺索引，其中第二卷将知识分为21个大类、250个细目，较全面地反映了当时的科学发展水平，成为西方第一部检索系统较为完备、详尽著录的综合性大型书目。13~14世纪的《圣经》语词索引直接以自然语言中的字、词作为标目，按字顺序查检，基本具备了主题法的要素。虽然当时主题和索引方法应用很少，但它成为近现代主题法和索引发展的渊源。

公元16世纪前后，随着罗马帝国的衰亡，欧洲社会逐步进入宗教统治时期，宗教文献及其记录与检索宗教文献的工具——目录、索引、文摘在西方国家有了迅速的发展。17世纪末，西方国家不仅出现了综合性与专题性的信息目录，英国和德国还诞生了具有现代意义的国家书目。1605年问世的《英国牛津大学图书馆目录》首次提出了基本款目的概念，为信息组织的理论探讨迈出了第一步。古代信息组织的主要活动是书目工作，这是一种对现有著述或文献特征进行识别、选择、著录与排序的信息整序工作。书目工作主要研究“已知”文献，对原始信息记录具有较强的依赖性。

总之，随着古代文明的发展，人类的文献生产、制作、积累与利用逐步形成规模，面对数量日益增长的文献，如果不加以组织与管理，人类将无法有效利用。基于此，早期的信息组织为适应人们管理与利用文献的需要而诞生。可以说，一定数量文献的生产与积累是信息组织产生的基础，而大量文献的产生与积累是信息组织发展的首要条件。没有一定数量文献的产生及文献资源的积累，就失去了信息组织产生、发展的物质基础。

## 3. 古代信息组织的特点

总体上，由于古代社会生产力低下，信息的数量和对信息的需求都非常有限，在此基础

上进行的信息组织活动基本处于初始阶段，呈现出以下特征：

- ① 以手工编撰书目的信息组织形式为主流，包括一书目录和群书目录、国家书目等，同时出现了以数序、音韵、类名和语词为排列方式的类书及从文献正文中摘录片段的文摘方法。
- ② 信息组织活动以个体劳动为主，虽然偶尔存在着简单的协作关系，但并未形成一种长期固定的专业机构，如我国古代《七略》和《七志》的编撰。
- ③ 信息组织着眼于文献的收藏管理，主要是为了管理一定范围内的文献，并非着眼于文献检索和利用；其信息组织的主要方式限于记录与登载文献的基本特征，或者按照学科或事物范畴排序；其成果形式类似文献清册，即文献书目。
- ④ 信息组织的主要对象是图书和档案文献。
- ⑤ 信息排序多采用分门别类的方式，国外偶尔采用主题（catchword）方式揭示和排序等。总之，不管古代的信息组织其形式与成果如何简单，但其“辨章学术，考镜源流”的功能和分类揭示与排序的手段对后来的信息组织具有深刻的影响。

### 1.4.2 近代的信息组织

近代信息组织时期是指从近代工业革命和印刷工业机械化到第二次世界大战（20世纪40年代中期）的一段时间，其特点是文献生产、积累与利用形成了较大的规模，信息组织的职能工作的主体由个体逐步趋向社会专门机构（如图书馆等）。

#### 1. 国外近代信息组织的发展（17世纪起）

近代工业革命极大地促进了社会信息活动的发展。以蒸汽机为核心的动力技术与活字印刷术的结合，进一步提高了文献生产的效率；交通工具的进步既密切了世界各国的联系，又为文献资料的传播提供了便利条件；以电力技术为基础的电信技术则为人类信息交流创造了新的手段。与此同时，近代科学的发展也为近代信息管理和信息组织活动开辟了广阔的舞台，提供了丰富的内容。科学研究活动从科学家个人的自发研究成长为有组织的社会事业，科学交流从自发组织的科学团体发展成为正规的学术管理机构，使科学劳动的成果成倍增加，文献信息的数量和需求也急剧增长。因此，图书馆作为社会上最早出现的信息管理和组织机构在这一时期获得了快速发展，文献组织的基本方法（如编目法、分类法和文摘索引法等）成为近代信息组织的主要方法。

近代以来，西方学术逐步进入发展时期，按研究对象划分的一门门自然科学开始从包罗万象的哲学中分离，人们的知识结构和思维结构发生了很大变化。新科学的产生和技术的发展在图书分类中得到了反映，出现了表达力学、古生物学、细菌学、人类学、胚胎学、生态学和社会学等学科新概念的大类、小类和细目，出现了以学术发展为基础的知识分类体系。近代西方一些哲学家，如培根、康德、黑格尔、霍布斯和孔德等构造的不同知识分类体系对信息分类的思想产生了深远影响。其中较具代表性的是培根的分类思想和知识分类体系。他从心理特征出发，将知识分为三类：历史（记忆知识）、诗歌（想象知识）和哲学（理性知识），并在其下分出细类。这一分类体系后来被用于18世纪狄德罗编制的法国百科全书的组织，不少图书馆和藏书家也根据它来编制图书分类目录。恩格斯早在1858年就注意到了科学分类问题，他认为：“每门科学都是分析某一个别的运动形式或者一系列互相关联和互相转化的运

动形式的，因此科学分类就是这些运动形式本身依据其内部所固有的次序的分类和排列。”

随着学科的不断变化、知识记录的不断增多、大型图书馆藏书量的剧增，专业图书馆也得到了发展。在这种情况下，原有的分类法已不适用。在旧分类法中没有包含的、由于科学和技术的发展而出现的新的知识门类的大量涌现，不仅需要确定新知识及新学科在图书分类表中的位置，还需要深入研究，使类表扩展到图书分类结构，以便对几十万、几百万册图书和藏书目录进行组织。

1876 年，美国图书馆学家、教育家杜威编制了《杜威十进分类法》(Dewey Decimal Classification, DDC)，建立了结构完备、等级分明的分类体系和主题索引，体现了当时信息组织的最高水平。这一时期的分类法还对各种分类技术（如标记符号、类目索引）进行了探索，努力使类目体系具有“伸缩性”和“适应性”，类目符号具有简易性和助记性，出现了一批有影响力的综合性分类法，如克特的《展开式分类法》、奥特勒和拉封丹的《国际十进分类法》(Universal Decimal Classification, UDC)、《美国国会图书馆分类法》(Library of Congress Classification Schedules, LCCS) 等。在杜威之后，编制统一的国际分类法的思想已经成了各国图书馆员关注的问题。

在对大量文献检索的过程中，人们愈来愈感到传统的分类法已不能完全满足需要。除了对分类的族性检索要求，读者还开始关注对具体主题的特性检索。于是一些专家对以满足特性检索为目的的主题法进行了探讨。18 世纪后期出现了主词索引和字顺分类目录。主词索引直接在书名中抽取自然语言形式的关键词作为文献主题标目，成为后来标题目录的先导。字顺分类目录将字顺方法与分类体系相结合，在每级类目下以字顺方式排列同位类，并逐步演变为将类目直接按主题字顺方式编排，也为主题目录的发展创造了条件。

主题法的真正形成和发展是在 19 世纪后半期，最早期的类型是传统的标题法。1856 年，英国克里斯塔多罗的《图书馆编目技术》提出应用文献题名中的“词”作为字顺标题系统中表达内容主题的标题，成为标题法的先导。在这部书中，他提出了“主词”（即关键词）这一概念，并介绍了用轮排方法编制的曼彻斯特公共图书馆目录书名语词索引的步骤。1876 年，美国图书馆学家克特发表了《字典式目录条例》，在传统主词款目和字顺分类目录的基础上，明确规定了标题的意义和处理方式，制定了标题选择和使用的一系列原则和方法，从而完成了字顺分类法向字顺主题法的转变，标志着现代主题法原则的确立。1895 年，美国出版了第一部标题表——《美国图书馆协会标题表》，并在其后出版《美国国会图书馆标题表》(Library of Congress Subject Heading, LCSH)。主题目录在美国公共图书馆获得了广泛应用，引起了许多欧洲分类理论专家的注意。

同时，更具信息报道和指引意义的文摘和索引在此时期发展起来。1830 年，世界上第一部科技文摘杂志《药学总览》在德国问世，并附有索引。各国相继出版了一系列文摘杂志和著者、题名、语词、分类索引，如 1848 年美国普尔的《普尔期刊文摘索引》、1851 年美国的《纽约时报索引》。为开展有关研究和交流，英国于 1856 年成立了世界上第一个索引学会。

## 2. 中国近代信息组织的发展（鸦片战争后）

在漫长的封建社会发展过程中，我国的信息组织活动一直以藏书楼的建设为主体，与其相适应，四部分类法具有稳固的主导地位。鸦片战争以后，由于西学输入、社会变革和教育的倡导，反映新学科、新技术、新思想、新文化的信息内容日益增多，我国旧有的四部分类

法已不能适应。一些学者开始以新的思想和方法编制目录，打破了四部分类法一统天下的局面。一批反映西学、新学的书目问世。例如，梁启超的《西学书目表》摒弃了传统的四部分类法，而采用了自编的新分类法体系，所收入的西学翻译书籍分为学、政、杂三大类 23 小类，初步具备了现代分类法将知识划分为自然科学、社会科学、综合性图书三大部类的雏形。1904 年刊印的《古越藏书楼书目》完全打破了四部分类法的传统，将浙江绍兴古越藏书楼的所有藏书分为学、政两大部，容纳了反映新学科的书籍。

19 世纪末 20 世纪初，一些学者开始翻译、引进国外分类法。1909 年，孙毓修在《教育杂志》上连载《图书馆》，介绍《杜威十进分类法》，引起了广泛关注。一批留学国外学习图书馆学的学者归国后，不但积极从国外引进新的分类法及其理论，而且着手对我国旧有分类法进行改造，编制了一系列中西合璧的新型分类法，以寻求解决我国文献分类问题。至 20 世纪上半叶，我国引进的国外分类法有 23 种，包括著名的《杜威十进分类法》(DDC)、《美国国会图书馆分类法》(LCCS)、《国际十进分类法》(UDC)、《克特展开制分类法》(EC) 等。这对我国的信息组织体系结构和方法有相当大的影响。在学习研究国外分类法的基础上，我国相继出现了仿杜、补杜及自编的各种分类法，达三四十种。比较有影响的如 1917 年沈祖荣、胡庆生编制的我国第一部以标记符号代表类目的《仿杜威书目十类法》、1929 年刘国钧自编的《中国图书馆分类法》、皮高品的以杜威法为基础改编的《中国十进分类法及索引》等。其间，虽然反映旧学的四部分类体系仍在继续起着作用，但总体上，我国文献分类法的发展过程是一个引进与继承、融合与创新的过程。

在信息分类的实践活动不断发展的同时，我国图书馆学家在理论上对分类标准、分类原则、分类方法等也进行了深入探讨。例如，1920 年之嘉在“科学分类之历史”中论述了科学分类史及一系列国外学者的分类思想和科学分类体系，杜定友、查修等系统研究了我国历代书目及图书分类体系，戴志骞率先提出了图书分类法的编制原则。1929 年 1 月，中华图书馆协会在第一次年会上通过了编制标准分类法的四项原则：“中西分类一致；以创造为原则；分类标记须易写易记，易识易明；须合中国图书情形。”这反映了当时的基本分类思想。

在主题法方面，虽然 20 世纪 30~40 年代我国已有少数标题表问世，如多源 1939 年的《标题表》、吕绍虞 1935 年的《中文标题总录初稿》，但在当时没有得到实际使用。

20 世纪初，西方现代索引技术传入我国，30~40 年代中国兴起了一个编纂索引的高潮，如哈佛—燕京学社引得编纂处洪业等主持编制了一系列索引，叶圣陶的《十三经索引》、王重民等的《清代文集篇目索引》和开明书店编纂出版的《二十五史人名索引》等，并有钱亚新的《索引与索引法》和洪业的《引得说》等索引研究著作出版。

### 3. 近代信息组织的特点

① 信息产品进一步丰富，信息组织与管理作为一种相对独立的职业得以发展。这一时期，图书馆、档案馆和情报部门等文献机构普遍开展了文献信息组织活动，并开始面向社会公众提供信息服务。国内外出现了一批致力于研究信息组织的专家。

② 用户需求成为信息组织关注的问题，“用户中心论”在信息描述、揭示与检索点形成与排序过程中得到认同与贯彻。除了围绕着文献的保存开展，信息组织的活动开始注重根据社会需要提供信息服务，开始重视对信息外在特征和内容属性的全面描述、揭示，并为用户提供更多的信息检索点。

③ 文献分类法的建立和完善。1876 年，美国图书馆学家、教育家杜威编制的《杜威十进分类法》(DDC) 建立了结构完备、等级分明的分类体系和主题索引，体现了当时信息组织的最高水平。1933 年，印度图书馆学家阮冈纳赞编制了世界上第一部分面组配式分类法——《冒号分类法》，并系统提出了分面分类理论。对现代分类法和主题法的发展产生了巨大影响。在此过程中，我国的分类法在引进国外分类法的基础上推陈出新，实现了从四部分类法向以科学分类为基础的体系分类法的转变，并结合我国文献的特点，力求解决中外文献、古今文献的统一分类问题。

④ 信息组织技术与方法的逐步完善、信息组织的对象从载体单元逐渐转向知识单元，索引和文摘组织法的面世打破了书目形式长期一统天下的局面，对以满足特性检索为目的的主题法进行了进一步探讨。1830 年，世界上第一部附有索引的文摘杂志《药学总览》在德国问世。此后，各国相继出版了一系列文摘杂志和著者、题名、语词、分类索引，如 1848 年美国普尔的《普尔期刊文摘索引》、1851 年美国的《纽约时报索引》。

⑤ 主题组织法和机械化信息组织手段的发展。1895 年出版了第一部标题表——《美国国会图书馆协会标题表》(LCSH)，20 世纪 50 年代美国的陶伯 (M. Taube) 以字面上不能再分的词汇单元——元词作为标识，以字面组配表达文献主题，并结合比孔卡等设备的使用，开创了在检索阶段匹配检索的后组式检索方式。这标志着继标题法后，一种新的主题法——单元词法问世。这种方法的基本原理在美国海军兵器中心、美国专利局、杜邦公司等得到应用，为后来发展的计算机自动检索开辟了道路。

⑥ 信息描述与揭示的内容日臻完善，信息存取系统的检索途径增多。如美国《化学文摘》(CA)，除了有著者索引和主题索引，还增加了环系索引、分子式索引和专利号索引等。

近代信息组织的发展与变化为现代信息组织的发展与完善奠定了基础。

### 1.4.3 现代的信息组织

现代信息组织时期是指以电子计算机技术、通信技术和网络技术等一系列现代信息技术的诞生、发展及在信息组织中广泛应用为主要特征的时期，即从 20 世纪 50 年代至今。20 世纪中期以来，随着科学技术的飞速发展，人类生产信息、积累和利用信息的活动形成了空前的规模，出现了“知识爆炸”、新技术层出不穷的现象。以电子计算机、网络技术和多媒体技术为主的现代信息技术的发展不仅把信息组织推上了一个新的发展水平，也将信息组织的技术手段带入到一个新境界，信息组织不但成为现代科学技术与国际经济发展的重要组成部分，而且已成为人类社会文明程度的重要标志之一。

大体上以 20 世纪 90 年代初期互联网的商业化为分界线，现代信息组织可以分为基于计算机应用的信息组织和基于网络的数字信息组织两个阶段。

#### 1. 基于计算机应用的信息组织阶段

这一阶段信息组织发展的特点从宏观和微观两方面得以体现。

1) 宏观上的主要特点是信息组织的技术化、社会化、产业化和标准化

(1) 信息组织技术化

自 1946 年世界上第一台电子计算机在美国诞生后，计算机技术便被应用于信息组织工作。如 1954 年，美国海军兵器试验站首次建立了世界上第一个计算机书目存取系统。卢恩

(Luhn) 首创了题内关键词索引，开创了计算机编制索引之先河。20世纪60年代以后，书目工作自动化有了进一步发展，首先是机读目录 (Machine Readable Catalogue, MARC) 研制成功。1966年，美国国会图书馆成功地开发了机读目录 MARC I，1968年推出了 MARC II，并正式向国内外发行。机读目录的研制成功极大地推进了信息组织现代化的进程。20世纪70年代以来，信息组织的现代化领域不断扩展，从文献编目自动化走向文摘自动化、检索网络化、翻译自动化、索引自动化和自动抽词标引，其实现方案也层出不穷。例如，英国于1974年推出新型的索引系统——保持上下文索引系统，在世界上产生了很大反响。我国上海交通大学的王永成教授等人对自动文摘的研究已取得令人鼓舞的成绩。

另一方面，自20世纪80年代以来，除电子计算机以外的其他信息组织技术也发展很快。例如，1978年第一个光盘问世，1979年第一个实用光缆通信系统在美国芝加哥投入使用；20世纪70年代末80年代初，可视数据存取系统在英、法等国家实际使用。所有这一切极大地促进了信息组织手段与技术的现代化。

#### (2) 信息组织社会化

信息组织社会化起源于图书馆联机编目工作。早期的文献编目采用手抄、雕刻等手工方式，随后转换为打印模式。20世纪50~60年代，计算机逐步应用于编目领域，美国国会图书馆先后研制成功 MARC I 和 MARC II 后，采用电子及光电技术进行编目成为60年代后期崭新的计算机编目形式。MARC 的优点是：可促进文献编目标准化；加速编目流程，加速书目信息的存储与传递速度；可一次输入，多次输出；可为编目社会化和网络化提供条件。60年代中期，MARC II 试制成功，标志着联机编目社会化和网络化的开始。美国是世界上实现联机编目最早的国家，其中最著名、影响最广的是联机计算机图书馆中心 (Online Computer Library Center, OCLC)。到80年代中期，全美已有5000多个图书馆，约94%的资料利用这个联机系统进行编目，各图书馆同时向该系统输入各自的原始目录记录，供其他单位合作使用。美国是社会化联机编目的先驱，为世界各国编目工作网络化提供了宝贵经验。

#### (3) 信息组织产业化

信息组织活动产业化孕育于20世纪50年代，至70年代中期进入稳步发展时期。由于信息组织产品可以转化为商品，大量文摘索引服务社和书目数据公司的产生与发展，使具有特殊社会经济价值的信息组织产品的生产和流通出现了产业化和商业化趋势，主要表现为两方面。一方面，信息组织的数据库产品逐步应用于各领域，包括企业、政府机构、金融行业、科研单位、商业领域等，这就为其产业化提供了市场。数据库研制与服务由早期分散的、小规模形式逐步发展成为集中的、多数据库的大规模服务，并在全世界形成了几十个大型的跨国数据库服务中心，建成了国际性的数据库销售体系。另一方面，商用信息存取系统自70年代起先后投入营运，如国际联机存取系统 DIALOG、STN 和 OCLC 等。与此同时，传统的印刷型文摘书目等检索刊物纷纷转为磁带型、网络型书目数据库，如《化学文摘》(CA)、《生物学文摘》(BA) 和《科学文摘》(SA) 等均同时出版印刷型、缩微胶片型、磁带型和网络版。书目数据库信息组织产品开发的产值逐年增加。

#### (4) 信息组织标准化

标准化是此阶段的重要特点之一。信息描述与著录经过100多年的实践，从随意编目到一国和跨国统一编目，直到20世纪才迈开国际标准化的步伐。其中，文献著录标准 ISBD 计划自1984年基本完成后，很快成为世界各国编制文献著录标准的主要蓝本。此外，LCC 和

DDC 这两个分类法由西方国家如英国、美国、加拿大等英语使用国的文献分类组织统一使用，经过数十年的多次修订，已成为国际公认的较为成熟、质量较高、被世界各国广泛采用的文献分类体系。我国目前统一使用的《中国图书馆分类法》经过不断修订，其科学性和逻辑性也不断加强。不仅如此，文献主题标引工作标准化于 20 世纪 80 年代在西方国家普遍推广，除了 LCSH 被广泛采用，分类主题一体化的模式也备受信息组织理论研究与实践领域的关注。我国的一体化分类主题词表《中国分类主题词表》于 1994 年正式出版发行，加快了我国文献标引工作的标准化进程。在版编目（Cataloguing In Publication, CIP）则是我国于 20 世纪 90 年代启动的一项信息组织标准化的重要措施。

2) 微观上的主要特点是对分类法的改造、主题法的发展、分类主题一体化、自然语言检索系统的兴起及对自动标引和分类的探讨

### (1) 分类法的改造

早期的分类法如杜威十进分类法、克特展开式分类法均属于列举式分类法。这种分类法的树型结构虽有结构明晰的特点，但不便揭示复杂、专指、细小的主题，也不便容纳新的学科主题。为此，分类学家们进行了多方面的改进。1933 年，印度的阮冈纳赞编制出世界上第一部分面组配式分类法——冒号分类法后，其中体现的分析—综合的分类理论引发了分类领域分面研究的热潮。在其影响下，一系列专业分面类表问世。例如，英国分类法研究小组在深入探讨分面分类理论的基础上，编制了十多种专业的分面分类法，该小组成员奥斯汀据此编制了保留上下文索引系统（PRE-Context Index System, PRE-CIS）。1955 年，该分类法研究小组向英国图协和联合国教科文组织提出“需要以分面分类作为一切情报检索方法的基础”的著名备忘录，在 1957 年的国际分类研究研讨会上受到一致肯定，从而奠定了分面组配理论的基础地位。

同时，传统分类法被不断增加分面组配成分，朝分面组配方向改造。如 DDC 的通用复分类表从无到有，从 1 个增加到 7 个，专类复分和仿分也在增加，第 20 版则用分面分类的方法对音乐类进行了全面的改造。1976 年，英国分类法研究小组的成员米尔斯（L.J. Mills）对美国图书馆学家布利斯 20 世纪 40 年代初编的《书目分类法》（Bibliography Classification, 即 BC1）进行了全面的分面改造，使其由原来的等级列举式分类法发展成一部大型的分面组配式分类法——《布利斯书目分类法》（Bliss Bibliography Classification, 即 BC2），成为列举式分类法彻底分面改造的典范。

我国现代分类法的发展经历了从 20 世纪 20~30 年代模仿杜威法、50~60 年代模仿苏联分类法，到全面探索我国适用分类法的理论和方法的过程，初步形成了具有中国特色的分类体系和分类理论。新中国成立以来，先后编制了多部大型综合性分类法和数十种专业分类表，如《中国人民大学图书馆图书分类法》《中国科学院图书馆图书分类法》（简称《科图法》）《中国图书馆分类法》（简称《中图法》）《中国档案分类法》《中国标准文献分类法》等。尤其是大型综合性分类表《中图法》广泛吸取了国内外各种分类法的优点，以科学分类为基础，同时在四次修订过程中不断扩大分面组配技术的使用范围，逐渐增加复分、仿分方法，引入并扩大冒号组配方法的应用，成为目前我国文献信息组织使用最广泛的标准化分类体系。

### (2) 主题法的发展

陶伯（M. Taube）开创的单元词法问世后，这种方法的基本原理在计算机检索系统中得到应用。几乎与单元词法同时，1947 年至 1950 年，美国的穆尔斯（C.N. Mooers）在研究组

配分类法的基础上，提出了一种新型主题法——叙词法，并创造了“叙词”“叙词法”“情报检索”“情报检索系统”等术语。他认为，叙词是表达概念的一种标识符号，是十分自由和独立的观念成分，应能以任何组合或次序来规定一次检索。20世纪60年代，为适应计算机在图书馆及情报工作中的应用，叙词语言吸收了标题法、单元词法、关键词法和分类法等检索语言之长，逐步取代了元词法成为现代情报检索语言的主流。叙词以概念组配取代字面组配，并广泛揭示概念间关系，使文献信息的揭示更加准确。1959年，美国杜邦公司编制了第一部叙词表。

在我国，1950年程长源的《中文图书标题法》出版，填补了新中国中文标题语言的空白。20世纪70年代，我国开始大规模编制和使用主题法。1971年，航空部情报所编制的《航空科技资料主题表》第2版问世，成为我国的第一部叙词表。此后，《电子技术汉语主题词表》《常规武器专业主题词表》《国防科学技术主题词典》《原子能科技资料主题词典》《机械工程主题词表》《汉语主题词表》陆续出版。尤其是1979年出版的《汉语主题词表》，作为“748工程”（汉字信息处理系统工程）配套项目，历时5年编成，不仅成为世界上最大规模的叙词表，还为后来我国叙词语言的大发展做了人才和理论准备。

### （3）分类主题一体化

以往分类法、主题法作为信息组织的两种独立的方法各成体系，在使用过程中，标引和检索需要分别进行。计算机在信息组织中的应用进一步揭示了分类系统与主题系统的联系，促成了分类主题一体化的发展。1969年，英国学者艾奇逊（J. Aitchison）研制成功世界上第一部分类主题一体化的《分面叙词表》，将一部分面分类表与一部字顺叙词表结合，通过严格规范，使每个词汇同时出现在分类表与叙词表中，实现了两种检索语言的兼容。在其影响下，英美等国陆续出版了一批分类主题一体化词表，如《伦敦教育分类法》（第2版）《建筑工业叙词表》《基础叙词表》等。此外，印度、日本、苏联、德国等许多国家（和地区）开展了分类主题一体化的研究和实践活动。

20世纪80年代，我国先后用手工和计算机编成《常规武器分面叙词表》和《教育分面叙词表》。这是我国图书情报界编制一体化词表的最早尝试。此后，我国陆续编制、出版了十余部一体化词表，包括三部大型词表——《中国分类主题词表》《农业科学叙词表》《社会科学发展叙词表》和七部中型词表，使一体化语言理论与应用研究进一步深入。其中，《中国分类主题词表》作为国家哲学社会科学“七五”规划重点项目，对推进汉语分类主题一体化词表的发展起到了重大推动作用。

### （4）自然语言检索系统的兴起

分类法和主题法采用规范化的人工语言系统控制文本信息，虽然规范严谨，但难以掌握，词表的稳定性强，无法适应新学科、新事物的变化。因此，它们的使用长期被局限于科研和专业领域，很难成为大众信息检索工具。在新形势下，人们希望有更方便、更具亲和力和大众性的语言系统，以满足日益丰富的、动态的信息检索需求。

20世纪50年代，卢恩在前人探索的基础上，将计算机用于关键词索引的编制。其后，各种直接以自然语言为标识的检索系统相继出现。20世纪60年代初，美国匹兹堡大学健康法律中心率先建立起第一个全文检索系统，以电子文本为处理对象，通过计算机以自然语言的语词或语词的组配进行检索，广泛探索了自由文本的检索方法，后来，图书馆界的书目检索系统吸纳了这一方法，将文本检索用于标题和文摘检索。

20世纪70年代，一批采用句法—语义分析技术的自然语言理解系统脱颖而出，在语言分析的深度和难度方面都比早期的系统有了长足的进步。进入20世纪80年代后，为了适应知识管理的需要，智能信息检索系统开始应用于自然语言的处理。自然语言容易被普通检索者接受，但存在着表达概念过分自由、语义无关联、词汇无控制等问题，影响了检索效率。于是，一些专家提出了编制后控词表等方案以改进检索效果。

#### (5) 对自动标引和分类的探讨

自动标引始于20世纪50年代后期。1957年，卢恩在对自动标引和自动编写文摘研究的基础上，发表了关于文献自动标引的论文，提出了基于词频统计的抽词标引法，率先进行了自动标引的探索。从60年代后期到70年代末，自动标引研究取得了很大进展，提出了概率统计标引法、句法分析标引法及各种加权模型等，建立了一批应用与实验系统。1962年，博科(H.Borko)等人利用因子分析法进行了文献的自动分类。此后，各国学者进行了大量有关文本词频统计分析、句法分析、语义分析等实验研究，并在主要的国际联机检索系统中实现了全关键词自动标引和检索。主关键词标引也建立了许多实用系统。20世纪70年代美国国防部文献保障中心(Defense Documentation Center)采用的机助标引系统、90年代美国NASA宇航信息中心使用的机助赋词标引系统等，都是结合自动标引研究成功建立的人机结合的实用系统。20世纪80年代末，日本庆应义塾大学文学系的图书情报专业和日本IBM东京基础研究所合作开发了一个自动分类专家系统，采用国际十进分类法(UDC)实现了图书资料的自动分类。1987年，《杜威十进分类法》(DDC)的联机编辑系统建成，并完成了对DDC第20版的自动编辑。至今，自动标引形成了抽词标引和赋词标引两大主要类型。几乎所有大型分类法都实现了计算机辅助编辑修订和管理。

我国自20世纪80年起开始从事汉语自动标引和分词实验研究，并逐步达到了科技文献自动分词的实用水平。20世纪80年代中期起，我国在大量有关自动分类理论研究的基础上，相继开发出一批机辅分类标引系统、自动分类标引系统，如“计算机辅助分类主题标引系统”“文书类档案的计算机分类标引算法”“中文文本自动分类系统”“中文技术文档的自动分类系统”“图书分类专家系统”和“中文地质资料分类专家系统”等。

## 2. 基于网络的数字信息组织阶段

这一阶段，信息组织的主要特点是：在以计算机、现代通信技术为代表的信息技术的推动下，传统的信息组织方法不断创新和完善，向自动化方向发展；同时，在网络环境下，信息组织面对多元化的信息载体和多元化的信息需求，又需要不断拓展新的应用领域，探讨新的数字信息组织方法。这两者之间存在着密切的联系。

### 1) 网络环境下传统信息组织方法的延伸和发展

#### (1) 利用元数据对数字信息体进行描述

元数据(Metadata)是关于数据的结构化的数据，专门用来描述数据的特征和属性。一个元数据款目构成一个信息资源的基本数据，是检索系统的基本构成单元，可以代表信息资源用来组织目录、索引、数据库和搜索引擎等检索系统。信息描述的目的就是以元数据为中介，对信息资源进行各种操作。元数据的基本类型有：以详细记录为目的的元数据——机读目录(MARC)，以发现为目的的元数据——都柏林核心(DC)，以网络查询为目的的元数据——元搜索引擎。

MARC 即机读目录，是以代码形式和特定结构记录在计算机存储介质（磁带、磁盘、光盘）上的，用计算机识别和阅读的目录。MARC 作为一种以详细记录为目的的元数据，是电子化管理、发布类表、词表、人名、机构名等规范文档的标准方式。目前，美国国会标题词表（LCSH）、国会图书馆分类法（LCC）都提供 MARC 版本并可以在网上查询。用 MARC 格式表示的知识组织系统可以植入联机公共目录查询系统（OPAC），与书目数据统一管理。

都柏林核心（DC）元数据，是为了描述数字信息资源、支持数字信息资源检索而建立的元数据模式，是一种以资源发现为目的的元数据，旨在为数字信息的著录提供一种更简单、更有效的方式。DC 元数据通过电子资源提供者对数字资源属性信息的描述，粗略地对资源内容进行编目，来帮助人们尽快在网上发现所需的资源。DC 元数据避免了搜索引擎结构过于简单而 MARC 格式又过于复杂等问题，不需要进行专业化训练就能对数字信息进行恰当的著录，降低了编目成本，能迅速适应数字信息资源的巨量增长，在数字信息组织中得到了广泛的应用。

元搜索引擎是一种以网络查询为目的的元数据，使用自动索引软件来发现、收集并标引网页，建立数据库；以 Web 形式提供给用户统一检索界面，供用户输入关键词、词组或短语等检索项；代替用户在数据库中查找出与提问匹配的记录并返回结果且按相关度排序输出。其特点是：由自动索引软件生成数据库，收录信息广、加工速度快，能及时向用户提供新增信息。元搜索引擎由检索请求提交机制、检索接口代理机制和检索结果显示机制三部分组成。检索时，直接输入关键词或词组、短语，无须判断类目归属，使用比较方便。常用的搜索引擎有谷歌、百度等。但是，元搜索引擎依赖于数据库选择技术、文本选择技术、查询分派技术和结果总和技术，用户界面的改进、调用策略的完善、返回信息的整合和最终检索结果的排序是未来元搜索引擎研究的重点。

### （2）虚拟图书馆（专题指引库）

虚拟图书馆就是根据特定的目标，选定信息资源的学科领域，对有关网站网页进行搜索和收集，加以鉴定核实，并对核实后的网址进行合理组织，使之能够提供检索、浏览和链接的信息集合。与搜索引擎的主要区别在于，虚拟图书馆属于专题性和学科专业性的，其系统性和易用性强。虚拟图书馆对网络资源的组织是优于搜索引擎的关键环节。该环节由专业图书馆员把关，在自动系统的协助下，利用某种分类法和主题词表，对收集来的原始资源进行描述和组织，改善了搜索引擎采用自然语言标引的根本缺点。这方面的实践和研究集中在重点学科导航库建设、专业网络资源导航库建设和热门站点链接或相关站点推荐方面。重点学科导航库以学科为单元，对互联网的相关学术资源进行搜集、评价、分类和组织的序化整理，并对其进行简要的内容揭示，建立分类目录式资源组织体系、动态链接学科资源数据库的检索平台并发布于网上，为用户提供网络学科信息检索线索的导航系统。例如，我国“高等教育文献保障系统”（CALIS）提出了构建重点学科导航库系统，建议其内容可分为 7 项，并规定各子项目必须有分类浏览功能，以主题树浏览方式组织信息。

### （3）分类法在数字信息组织中的应用

国内外学者就图书馆分类法特性、具体分类法与搜索引擎分类体系比较（类目涵盖范围、揭示深度、类表结构和功能）进行研究，认为传统分类法的知识系统性、标识语言的通用性、族性检索能力和扩检、缩检功能是其他情报检索语言不具备的。它在网络中的应用主要如下。

其一，用于数字信息资源组织。以文献分类法为工具的网络资源检索服务系统，从学科

的角度揭示网络信息，成为组织网上学术性知识内容的主要应用模式。例如，欧洲科研与教育信息服务发展计划于 1997 年 8 月在“互联网资源描述与发现”专题报告中介绍了这方面的应用——17 个网上服务系统使用 DDC 组织资源、5 个使用 UC、5 个使用 LCC。国外在这方面的实验还有：使用 DDC 的万维网杜威分类法 (CYBERDEWEY)、使用国会图书馆 (LC) 的万维网虚拟图书馆 (WWW.Virtual Library)。研究表明，分类法能够提供网络信息资源框架，并作为网络环境下的浏览工具。

其二，建立网络信息资源分类法。现有众多的中文搜索引擎均采用各自分类体系组织网络资源，用户必须熟悉不同的分类体系才能较快地检索信息，这给用户带来了极大不便。专家们提出，要建立网络信息分类法，提供网络资源的统一分类体系。有的专家提出，建立网上信息的知识分类系统的基本结构与编制方法；有的专家提出，以《中图法》为蓝本，综合国内外优秀分类法和搜索引擎的优点，编制我国的数字化分类法；有的专家提出，从最终用户检索需求出发改造《中国分类主题词表》。

其三，人工神经网络 (Artificial Neural Network, ANN)。人工神经网络是根据人类的生物神经系统结构设计的计算机系统，在信息组织中可以用于自动分类。国外有些信息检索系统已经采用了使用人工神经网络的自动分类系统。目前，分类上应用最广泛的人工神经网络模型称为自组织映射 (Self-Organizing Map, SOM)，可以实现 Web 文档的自动聚类，在此基础上进一步，即利用 SOM 网络实现索引词聚类，就可以实现超文本链接的自动生成。

#### (4) 主题法在数字信息组织中的应用

近年来，国内外学者及研究机构已认识到主题法在数字信息组织中的重要作用，主要集中在以下几方面。

其一，关键词法的应用。关键词法具备以下优点：在标引时不必查表，选词、标引速度快，成本低；不依赖专职标引人员，可由作者或机器自动标引；不存在人为性或滞后性，能及时应用最新的提法及最新的词汇等。因此，目前由搜索引擎软件自动建立的网络信息资源索引数据库支持的就是关键词检索。但是关键词检索的查准率低，所以人们提出使用后控词表（标引不控制+检索控制）的模式改进关键词法的检索性能。

其二，主题词表的应用。少数搜索引擎提供主题词检索方式，在用户界面上可直接浏览主题词表，从中选取主题词，作为搜索引擎的检索提问。用户可以在检索界面中修改检索提问，也可返回到主题词表界面重新选择主题词。其共同特征是词表内具有超文本导航。

其三，标题词表的应用。标题词表在网络信息组织中的应用可以分为两种情况：一是检索前使用，即通过标题词表规范用户的检索表达式，用户可以首先在网络信息组织工具提供的词表中检索到标准的标题词及相关联的词汇，以该词作为检索词，单击表中超链接即可得到检索结果；二是检索后使用，即在给出用户所用检索表达式，得出检索结果的同时，提供相关词作为用户进一步检索的线索，用户可自由进行扩检和缩检，从而提高检索效率。

#### (5) 主题图的应用

主题图是一种新型的数字信息组织方法，可以提供最佳的信息资源导航。主题图利用了主题索引的概念及网站的特点，将主题、联系和范围三者紧密结合起来，用格式表单来控制信息的获取和浏览，并详细描述各种浏览层次，实现对复杂知识管理关系的模拟，以便帮助用户更有效地浏览数字信息资源。由于主题图吸收了各种知识组织方法的长处并采纳了本体论和语义网的部分思想，因此能对数字环境下的信息资源进行有效的组织和管理。主题图具

有良好的信息检索功能，表现为：可支持现有的搜索引擎在资源层面实现检索；主题图概念可看成一个图或树，支持可视化人机交互式检索；主题图可看成本体，提供一定程度的概念之间的关系描述，利用概念间的关系，提供一定程度的智能化检索。

#### (6) 本体的应用

本体（Ontology）的概念源于哲学，即对世界上客观存在物的系统描述，一般称为本体论。许多研究人员从不同的问题和研究角度出发，对于本体给出了不同的定义。Studer 等在对本体进行深入研究后，给出了一个被广为接受的定义，提出“本体是共享概念模型的明确的形式化规范说明”。这个定义有 4 层含义：“概念模型”是指识别反映某些现象的相关概念的抽象模型；“明确的”是指所使用的概念及它们之间的联系都被明确定义；“形式化”是指本体是计算机可读的；“共享”是指本体中反映的知识是其使用者共同认同的。

某种意义上，本体同叙词表一样是一种控制词表，是一种知识组织工具。虽然本体和叙词表同是知识组织工具，但在形式化水平、概念抽象和语义关系表达等方面存在着明显差异。本体的结构特性使其在理论上具有超越叙词表性能的可能性，并为它在一个更广阔的范围内获得有效应用奠定了基础。事实上，本体的应用范围比叙词表更广泛，信息组织与检索不过是它的一个适宜应用的领域而已。从实现虚拟组织信息共享出发，有学者提出了一种基于本体的两阶段视图映射关系构建方法，以使得所构建的视图映射关系既能保证较高的信息查询效率，又能保证具有良好的可扩展性。目前，学者们正在积极研究如何基于叙词表来构建本体的问题。

#### 2) 网络环境下新的数字信息组织方法与手段不断涌现

网络环境下新的数字信息组织方法与手段层出不穷，下面仅从三方面进行介绍。

##### (1) 从总体规划层面，信息组织分为信息构建、信息网格、概念地图等

信息构建（Information Architecture）源于从建筑学视角来解决信息组织和利用的问题，通过合理地组织、标识信息并构建信息环境，以改善信息浏览和信息检索的过程与效果。信息构建的核心理念是关注用户、以人为本。信息构建是信息用户、信息内容与信息组织三者的交集。在对信息组织、导航、标识和检索环节的设计实施过程中，应当时刻关注用户需求、用户习惯和用户利益；可视性和可描述性也是信息构建关注的焦点，要求深刻理解信息环境，明确信息在哪里、这些信息属于谁、如何利用这些信息及组织信息的目标是什么等问题。信息构建的主要活动是组织信息内容、生成信息结构和设计信息界面，其直接目标是建立一个清晰的、易于理解的信息结构，最终目标是提供给信息用户一个良好的信息空间环境。

网格技术是近年来国际上兴起的一种信息技术，是互联网信息技术发展的新趋势，包括计算网格、知识网格和信息网格等。将网络引入信息管理领域，便产生了所谓的信息网格。信息网格就是利用现有的网络基础设施、协议规范、Web 和数据库技术，为用户提供一体化的智能信息平台，其目标是创建一种架构在操作系统和 Web 之上的基于网络的新一代信息平台和软件基础设施。在这个平台上，信息的处理是分布式、协作和智能化的，用户可以通过单一入口访问所有信息。信息网格追求的最终目标是能够做到服务点播（Service On Demand）和一步到位的服务（One Click Is Enough）。目前，对信息网格的研究主要集中在体系结构、信息表示和元数据、信息连通的一致性和安全性等方面。

概念地图是针对特定主题的个人结构化知识的一种图示方法，也是语义网的可视化表示方法。概念地图用节点表示概念，用连接线和连接词表示概念之间的关系。概念地图具有三

个特征：一是命题，由两个以上的概念及其关系构成表达意义的陈述；二是等级结构，按照宽泛概念在上、具体概念在下的顺序排列形成等级结构；三是交叉关系，不同分支中的概念之间形成的连接关系。概念地图的构建包括四个步骤：第一，概念选取，即列出关于某主题的所有重要概念；第二，概念分类；第三，定位中心概念；第四，连接交叉概念。概念地图的构建过程即是知识创新的过程，利用概念地图可以沉淀隐性知识。概念地图的构建过程也是学习的过程，可以将学习中涉及的资源链接到概念地图，实现知识结构与相关资源的整合；同时，概念地图表示的知识结构遵循人类的认知和学习过程，因此可以用于用户检索所需的显性知识。

### (2) 按信息服务方式，信息组织分为智能检索、个性化服务、信息可视化等

智能检索以文献和检索词的相关度为基础，综合考查文献的重要性等指标，对检索结果进行排序，以提供更高的检索效率。智能检索的结果排序同时考虑相关性和重要性。相关性采用各字段加权混合索引，分析更准确。重要性是指通过对文献来源的权威性分析和引用关系分析等实现对文献质量的评价，这样的结果排序更加准确，更能将与用户愿望最相关的文献排到最前面，提高检索效率。

个性化服务是指为了方便用户利用各种数字化资源，充分利用各种智能化技术对不同类型、不同特点的数字化资源进行整合，实现信息资源、信息技术、信息内容的集成，使目前信息资源组织系统的公共用户界面变得简单、友好，并且使用户利用同一检索表达式或检索词对各种数字化资源进行同步检索，实现同一主题信息资源的一步到位的检索与查询的信息组织方法。同时，个性化服务可以根据某些用户特定的信息需求定制具有个性化特点的用户界面，来提供符合其特定需求的具有个性化特点的信息和信息服务。个性化的信息服务是以信息资源整合和信息服务集成系统的建立为基础的，如果没有完备的资源整合体系作为后盾，无缝的、贴切的、高效的、主动的和一站式的信息服务模式就是无法实现的。

信息可视化利用计算机支撑的、交互的对抽象数据的可视表示来增强人们对这些抽象信息的认知。信息可视化的过程就是从信息维映射到可视维的过程。通用信息的可视化一般可分四个步骤：抽取、转换、映射（定义）、隐喻。当今信息时代，人们常用“信息爆炸”来形容信息量猛增的特征与趋势。要处理和应用这浩如烟海的信息，需要用先进的处理方法和有效的工具。信息可视化就是将信息转换成二维和三维图形、图像、动画形式的技术方法和有效工具。用户通过这些可视形式进行观察、交互。图 1-5 为《中图法》22 个基本大类展开的可视化检索界面。

### (3) 按语义组织方式，信息组织可分为语义网技术、网络知识组织系统等

语义网并非另一个独立的 Web，而是 Web 的延伸。其中，所有信息都有定义完好的含义，更利于人与机器之间的合作。语义网便于解决网络信息爆炸性增长所带来的巨大的处理和搜索问题。语义网的核心问题是如何有效地表达语义信息，能够被计算机有效处理。传统的数字信息组织以数据层面的组织和信息层面的组织为主。在语义网环境下，数字信息的组织以知识组织为主。语义网技术和本体技术的结合使信息组织进入更深的层次。语义网将改变传统互联网只是实现计算机硬件和网页的连接，而数据和信息资源分散在各网站，分别对信息进行解释、交换和处理的问题，更关注网络结构和语言的设计，使分布于全球的孤立的数据库融合，最终使用户独立运用网络上庞大的信息资源。本体使知识组织体系从传统的树型结构向网状结构进化，为各类不同知识体系的结构和结合方式提供理论基础，很可能极大

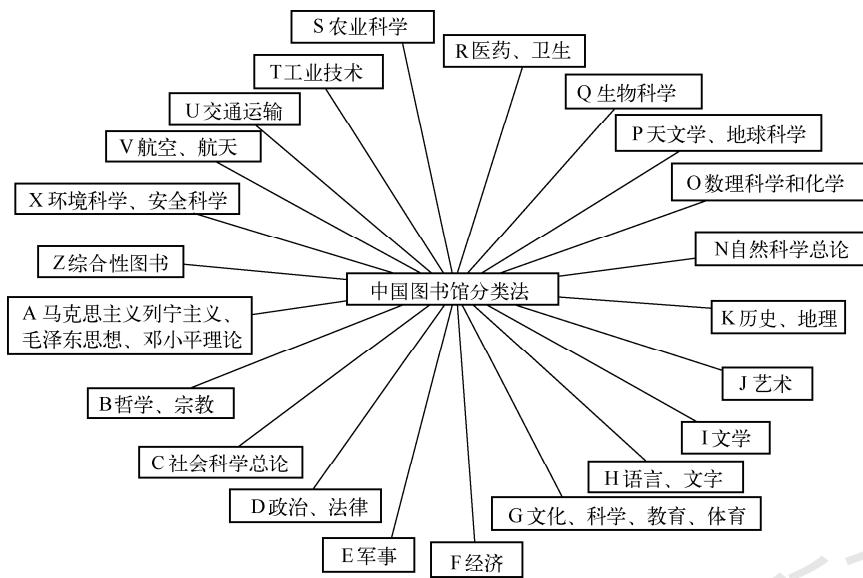


图 1-5 《中图法》可视化检索界面

促进信息/知识检索和导航功能的更新换代。

知识组织系统是各种对人类知识结构进行表达和有组织阐述的语义工具的统称，包括传统图书馆建立在文献单元基础上的分类法、标题表、叙词表及更泛指的情报检索语言、标引语言，也包括网络时代建立在概念单元或知识单元基础上的知识地图、语义网、概念本体等。从以文献单元、数据单元为基础的知识组织系统发展到以本体为基础的语义网络知识组织系统，反映了知识组织系统研究的方法和技术不断提高、应用范围不断扩展。网络知识组织系统的出现由两方面的共同作用所致：一是随着人类信息活动由纸质环境向数字环境迁移，传统知识组织系统的数字化、网络化势在必行；二是网络环境下信息量急剧增加，需要相应的语义工具实施对网络信息资源的组织，改进检索性能，实现对网络资源的深度挖掘和智能利用。网络知识组织系统代表了知识组织系统的发展趋势：数字化、网络化、语义化、协议化和自动化。网络知识组织系统的发展已经超越了“知识组织系统在网络环境中的应用”这一初级阶段，成为网络环境下新生的“语义工具”。网络知识组织系统将为图书馆知识组织带来生机。从关键词查找发展到内容检索，其作用十分关键。

#### 1.4.4 信息组织的未来发展趋势

互联网及信息技术的发展使信息传播方式发生了巨大的变化，对信息组织理论和技术则是一个重大挑战。面对庞杂无序多样化的数字信息，未来的信息组织应该以用户为中心，遵循实用性和易用性的原则，在继承优秀传统的同时，充分利用新的理论、方法、手段和技术，朝着更符合用户需要的方向发展，具体方向如下。

##### 1. 以需求为导向的信息组织方法和技术不断加强

###### 1) 基于个性化服务的信息组织

随着信息技术的发展和用户需求的变化，个性化信息服务的趋势愈来愈强劲。这是目前网络信息组织努力发展的重点之一。信息组织个性化努力可以包括多个层面的努力，如专

门搜索引擎的编制、综合性检索工具中特殊资源服务功能的建立、各种检索选择的提供（如大型搜索引擎中提供的对域名、语言、资源类型、检索网站的选择）、检索偏好设置的提供、检索优化的采用等，这都属于改进检索个性化努力的内容。但是，更深层次上的检索个性化则要利用网络搜索引擎的能力，发展检索个性化技术，通过对用户检索习惯和相关特点的分析，有针对性地提供相应的服务。

目前，基于网络的个性化信息组织与服务的思想在国外网站设计或信息服务系统中已经有一定的发展。许多门户网站和信息提供者推出了个性化定制服务系统，如 MyYahoo!、MyLibrary! 等。又如，Google 等使用的结合检索地点的资源提供技术等在一定程度上反映了检索个性化的一种努力。国内许多网站为了满足国内不同地区中文用户和英文用户的不同需要，将其主页以多种版本（简体中文、繁体中文、英文）组织，体现了以用户为中心的思想。面向不同对象的信息组织也受到了图书馆界的重视，开发出了 MyGateway、MyLibrary 等个性化定制服务系统，如美国北卡罗来纳州立大学图书馆的 MyLibrary、麻省理工学院和中国科学院都建立了基于用户的数字图书馆。国内部分高校图书馆也引进或开发了个性化定制服务系统。

在科学界，对信息组织的要求则更为专深、复杂，随着信息管理相关技术在各学科的逐步应用，新知识发现可以通过信息管理相关技术得以实现，使得科学家们对信息组织有了浓厚的兴趣和更高的期待。以行业领域数据挖掘为主的大数据和云计算成为新时期信息组织的重要研究领域，它们以处理海量多类型数据、高密度经济和研究价值等优势在学术研究、电子商务、数字图书馆、天文科学、大气科学、生命科学、农业科学、医疗卫生科学等多重领域广泛应用。

随着 Web 数据库技术、信息推送技术和智能代理技术等个性化信息服务所需的支撑技术的逐步成熟，网络信息组织将发展成一种信息代理服务，即根据网络信息的属性和用户需求，对网络信息进行加工、整理、排列和组合，使之有序化，以满足用户对网络信息的需求。

## 2) 信息检索工具的易用化

用户对各类信息、知识有日益增多的需求，但目前复杂的检索语言、词表结构和计算机技术却令许多人望而却步。以往的信息检索系统多是面向专业人员的，而普通用户要适应计算机系统的键盘和编码方式，还需要经过一个再学习的过程。尤其是对于中国用户来说，以西方使用方式和英语为基础设计的计算机给他们的使用增加了困难。为使更多的人能享受到信息化带来的好处，使更多的普通人也能不费力地从检索系统中查到所需信息，需要提供更简便易行的信息组织和检索工具。目前，已提出的解决信息组织工具易用化的方案如下。

① 自然语言的使用。用户不必事先掌握复杂的词表结构和检索规则，用自然语言标引检索，做到“所想即所得”，开发出便于中国人识别和检索的中文信息组织和检索系统。

② 用户界面可视化。通过用户易于理解的点线图、直方图、拼图、网状图及动画、三维技术和虚拟现实技术等表现复杂数据及其相互关系。例如，利用具有导航功能的信息地图，将复杂、抽象的语义以直观的方式呈现给用户，与搜索引擎相比，这样不仅可以更清楚地揭示信息方位，还可以体现信息间的联系，帮助用户从全局了解信息的分布情况。

③ 用文字识别、声音识别、触屏输入等更简单的人机界面代替目前的键盘输入等操作，清除用户学习的障碍。由于用户通过终端的屏幕获得网络信息资源，在进行信息组织时，还要考虑到对显示器屏幕的有效利用，使用户更方便地利用网络信息。随着计算机智能化和超

文本检索技术的完善，用于信息组织的智能软件会有很大发展。将来用户从信息组织到信息检索的过程有望像使用电视机、电话那样方便。

### 3) 信息内容揭示的深入化

数字时代用户对信息检索的需求不再仅仅满足题名、作者、主题词等传统条件下有限的检索点，而更注重实际内容（如目次、提要、文摘、全文、知识点等）的检索。这就对信息组织提出了更深入的要求。

① 组织的对象深入化。迅速普及的网络技术和数字技术使任意层次的任意信息元素、信息单元和信息集合体系正在逐步以计算机可识别和可理解的方式被定义、描述、指向、链接、传递和动态组织。网络信息组织的对象不仅停留在对信息特征的描述上，也深入到知识单元，扩大标引广度，增加数据库的标引深度，通过多层次、多方位的描述和分析来揭示与组织网络信息资源，以促进网络信息资源的合理利用。

② 网络信息挖掘。目前，网络上有 90% 的可用信息是非结构化信息，且数据往往是经常变动的和不规则的。网络信息挖掘在已知数据样本的基础上，通过归纳学习、机器学习、统计分析等方法得到数据对象间的内在特性，采用信息过滤技术，在网络中提取用户感兴趣的信息或更高层次的知识和规律。除了处理传统数据库中的数值型的结构化数据，网络检索处理更多的是文本、图形、图像、WWW 信息资源等半结构化、非结构化的数据。人工智能技术和信息推送技术促进了对网络信息的挖掘与深层次揭示，以更好满足不同用户的各种需求。

③ 向知识组织发展。未来的网络信息组织要更严格地控制信息的质量，对网络信息进行有效评价和筛选，为用户提供有价值的信息，而不是大量的无用信息，目的是向人们提供便于利用的、可以帮助解决问题的、序化的知识，实现从信息层次到知识层次的根本转变，组织的知识包括显性知识与隐性知识。目前的知识组织主要以文献单元和数据（各种事实、概念、数值的总和）单元为基础，但都是静态的、列举式的。未来的知识组织将以专家系统为基础，具有动态联系、判断、分析、比较、推理等新型知识处理与组织功能。

## 2. 信息组织工作的标准化和合作化

### 1) 信息组织标准化

在信息组织数字化和网络化过程中，必然遇到数据库、联机系统、检索系统和检索语言的兼容化和标准化的问题。例如，许多数据库的类型不同所引起的主题内容、学科领域、数据来源、文档格式、文档记录方式、标引规则及检索语言、输出格式等的不同给用户的使用带来了众多麻烦，影响了信息资源的共享。网络信息组织中的文件、搜索引擎、编目、学科信息门户等都涉及标准化问题，因此，解决信息组织和检索系统的兼容和标准化问题成为信息组织发展的主要趋势。

#### (1) 图书情报机构的参与

国际图联制定了《IFLA 书目记录的功能要求》。OCLC 一直注重研究、宣传与推行书目活动标准。国际知识组织协会 (ISKO) 制定了知识与信息组织的相关标准，并在其主页发布。美国国会图书馆于 2002 年 5 月提出了“元数据输入与传输标准”，并设立了 Z39.50 维护机构和 MARC 办公室。美国图书馆协会的兴趣与活动之一是标准化与指导，并有专门网页提供标准化信息。我国在文化和旅游部召集下，本着统一的规划、统一的技术标准及统一的运行规则等原则，组建了“中国数字图书馆工程建设联席会议”来协调工程的资源建设和标准规范。

## (2) 国际和各国家标准机构的推动

SC9 是国际标准化组织 (ISO) 的 ISOTC46 的分委员会，负责发展和维护关于文献展示、识别与描述的国际标准，并经常在其网站公布关于电子文献（包括网络信息资源）书目控制等方面的 ISO 国际标准草案。ISO 还成立了元数据工作组，负责元数据的标准与规范工作。美国全国标准化协会下设全国信息技术标准委员会，从事有关元数据的命名、标识、定义、分类和注册等工作，还成立了信息基础设施标准座谈小组。欧盟和英国的相关机构有信息社会标准化系统 (Information Society Standardization System, ISSS) 和英国标准协会的向用户传递信息解决方法部。另外，互联网管理机构、不同学科领域的学（协）会、公司、民间自发组织甚至个人，都在为网络信息组织及其相关的信息交换、信息检索、通信协议等方面标准的制定与推行做出积极努力。万维网联盟 (W3C) 是万维网上最有影响的互联网标准的认定机构，在网络信息组织领域，该机构认可的网络资源描述语言为 XML，资源描述框架为 RDF，元数据标准为都柏林核心元数据 (DC)，日期和时间格式为 W3CDTF。

国内外在网络信息组织标准化方面已取得一些进展，但仍有必要形成一系列标准与准则，使网络信息组织活动有规则可依。

### 2) 信息组织合作化

互联网是一个无主管的分散型互连结构，网络信息的通畅流动与有效利用要求各方面的整体配合。网络信息的组织是一项涉及面很广的持久性工作，需要全球性合作，以保证准确、及时地报道网上信息及其变动情况，提高信息的质量并实现规模效益。

图书情报界已经开展了网络信息组织合作化的有关活动。国际图联发起了“全球书目控制和国际机读目录核心活动”，并编辑出版了《国际编目与书目控制》杂志。2002 年 7 月，在西班牙召开的第 7 届国际知识组织大会的主题是“21 世纪知识表示与组织的挑战：跨越边界的知识一体化”。OCLC 一直在书目控制与资源共享的协作中扮演着国际中心的角色，国际图书馆协作体联盟则是将各种协作体组织起来的组织机构。例如，Renardus 是欧洲范围内开展的合作项目，由欧洲用户友好信息社会项目发展而来，得到了欧洲各国家图书馆、大学研究和技术中心、全欧洲主题信息网关的支持和合作。自 2000 年 1 月以来，这些合作者们正努力构建一个单一的基于 Web 的“中介服务”，目标是为遍布欧洲的分布式网络科学和文化信息资源提供跨库检索和跨库浏览。

网络信息组织的合作还将超出图书情报界，拓展到整个信息生产链的其他所有参与者，包括网站内容创作者、出版商和信息资源系统的合作和集成者等。美国的网络图书馆是世界上最大的全文电子图书的收藏者与服务提供者，也是出版商、发行商、图书馆与读者有效结合的典范。《出版机构与国家书目服务连接》是由英国、荷兰、挪威、法国、西班牙 5 个国家图书馆等 9 个机构合作的项目，解决了出版机构书目数据与 MARC 数据的双向转换，有利于实现将出版机构提供的数据作为电子出版物网络元数据的构想。

## 3. 信息系统的互操作和信息组织大众化

### 1) 信息检索系统的互操作

在分布、异构、变化的网络环境下，数字信息资源数量巨大且高度分散，而信息源、数据格式、用户需求高度异质，对如此庞杂的信息进行格式化和结构化很困难，用户要找到真正符合需要的信息也很困难。如何将广泛分布的、自治的、异构的信息资源和信息检索系统

联合起来，向用户提供统一、透明的服务并实现信息系统的互操作，成为当前和未来信息组织研究实践的热点，引起普遍关注。

系统互操作是指分布信息检索系统之间能无缝地交换、共享信息资源和信息服务，并能在不损害各分布系统自主性的同时构成一个虚拟的集成系统。在信息资源组织与服务的集成中，信息系统互操作的目标是向用户屏蔽分布的、异构的各信息系统间的差别，实现用户对多个信息系统的交叉浏览和交叉检索，提供统一入口的多个信息系统的检索和浏览服务，实现信息共享。

信息资源组织与集成服务要解决互操作的以下主要问题：屏蔽分布的各信息系统之间的差别，为用户提供一个一致的服务，在统一界面上进行的跨仓储的服务对于用户来说是透明的；为信息资源和信息系统提供一种灵活的集成机制，这种集成机制必须允许各相对独立的信息系统自由增加新的服务，或对以前的服务进行修改；信息资源整合和集成服务协议的制定，包括元数据协议、数字对象存储协议、信息搜索协议、付费协议、信息资源服务的运行管理协议等；开发信息资源整合与集成服务系统高层协议中间件，实现分布子系统间各项服务的互操作。

## 2) 信息组织大众化

Web 2.0 及相关概念的引入改变了现有万维网信息提供的模式，使信息组织不再局限于专业人员的范畴，而是走向大众化。

目前，网络上的信息检索和传输系统大都采用 C/S 模式，服务器总是被动地等待客户机的信息需求做出相应的响应。在 Web 2.0 中，用户已经不再简单地只是信息的获取者，也是信息的提供者。Web 2.0 发展了一系列新的形式，包括博客、互动标签（Tag）、社会网络交往系统（Social Networking Services, SNS）和资源订阅系统（Resource Subscription System, RSS）等。除了使信息大量增加，这类新形式更加重视用户作用的发挥，极大地促进了信息的传输和交流，而且在信息资源表述和关系揭示上具有重要作用，表现在以下三方面。

其一，通过用户的参与，改进资源揭示。最典型的是标签中用户个性化标注的使用，使用户的检索表述得以反映。

其二，利用用户交流和使用揭示信息资源之间的联系。SNS、RSS 都具有这一作用。这些联系是用户根据需要自动建立的，融入了用户的智力判断。

其三，通过交互方式的应用实现整合，例如，标签中某特定用户个性化标注的使用有可能是不适用的，但通过在使用过程中用户的调整和整合，适用的表述会得到保留；RSS、SNS 中联系的建立等具有同样的情况。尽管个体的处理会出现错误，但作为整体、作为统计学意义上的多数的判断，必然有其合理性。

Web 2.0 为资源处理引入了新的因素，改变了资源的数据特点，实质是用户的参与和用户智力判断的加大，而这一变化是结合使用需要进行的，是以最为自然的、人为性最少的方式实现的。Web 2.0 的出现将对网络资源的组织方法、处理技术等产生深刻的影响，包括资源搜索软件中采集策略的设置、检索提供中排序因素的确定、相关性揭示中相关性权值的设置，并有可能衍生出新的检索系统，如基于互动标签的浏览系统、基于 RSS 或 SNS 的相关揭示系统等。因此，Web 2.0 的发展可能给上述各方面的发展带来不同程度的影响和促进。

Web 2.0 提供了信息自组织功能，其兴起和应用创造了一个全新的信息空间。Web 2.0 将复杂的技术移至后台，在简单规则的约束下，用户广泛参与。知识信息的生产、传播和利用

在多元化、多样化、个性化和去中心化模式下实现了自组织和有序化。从博客信息交流社区的形成到维基百科的协同组织编辑及社会化书签产生的分众分类等，无不体现了 Web 2.0 的信息自组织功能和序化特征。

#### 4. 信息组织的智能化和语义网格化

##### 1) 信息组织智能化

网络环境下，先进的信息技术将部分地代替人脑进行信息组织中的信息识别、信息分析综合和信息重组，进而实现智能程度更高的“知识组织”，即用高度模拟人脑思维机制与习惯的方法来组织知识。

人工智能技术将促进网络信息的深层次挖掘和揭示，更好地满足用户的不同需求，如系统自动运行、不断更新用户资料库、提供个性化的主动服务等。计算机将能理解人们用自然语言输入的信息并正确回答用户提出的有关问题，还能自动对输入的信息做摘要，能用不同词语复述输入的内容，还能把用某一种自然语言表示的信息自动翻译为另一种自然语言。其中，使用最广泛的智能化技术是超文本、专家系统、数据仓库、知识挖掘等。

目前，有关信息组织智能化的一些关键技术的研究如自然语言理解、知识的表示、知识的获取等已取得一定成果，出现了自动抽取、收集和产生元数据的软件。例如，利用智能化的书签软件可以从 URL 中解析出元数据，并自动进行分类标引，还可通过监测用户的信息搜索与浏览过程，自动获得用户的需求与兴趣信息，把合适的信息提交给用户，并允许用户订购感兴趣的新资源或更新已有的信息资源。随着网上自动分类、自动标引、自动编制分类表与词表、自动漫游技术、信息类别的自动判别技术和信息推拉技术的逐步发展完善，会有越来越多的网络信息资源被自动地追加、组织到相应的位置，方便用户及时准确地检索。同时，信息组织技术的智能化将使信息组织工作变得十分简便，它不再只是专家和技术人员的事情，更多的用户将成为信息组织活动的直接参与者。

##### 2) 信息组织语义网格化

近年来，网络结构管理中的一个新进展是语义网格（Semantic Grid）技术的发展。网格计算（Grid Computation）及其相关的语义网格技术是着重从计算机网络结构的角度来研究和开发计算资源及进行语义处理的。网格计算最初是针对复杂的科学计算伴随着互联网的发展提出来的，被看成一种新型计算模式。这种计算模式利用计算机网络，把分散在不同地理位置的计算机组织成一个“虚拟的超级计算机”，每台参与计算的计算机被看成一个“节点”。这样，整个计算就是由成千上万“节点”组成的“一张网格”，所以被称为“网格计算”。网格计算在较长的一段时间里都是独立于语义网而发展的。现在语义网格技术也采用了语义网的许多技术标准，如 RDF/PDFS 语言标准用于对应的网络资源描述。语义网格技术正在呈现出与语义网技术相结合的趋势，为网络资源的自动计算及其组织体系提供了一个值得关注的方向。

随着语义网格概念的提出，国内外一些学者试图将这种新技术应用于数字信息的组织研究中。英国 Ali Shiri 在 e-Science 项目报告中提出，利用本体，为数字资源构建基于网格的语义框架。斯洛维尼亚 Liubljana 大学的 Ziga Turk 指出，语义网格技术对于识别和标注概念和术语、本体描述、体系结构构建及用户需求分析起着至关重要的作用，并指出语义网格技术在互操作、数字资源和虚拟组织等方面扮演着重要的角色。

总之，未来的信息组织发展的趋势将会呈现跨学科、跨国界、跨地域和不同文明之间高度融合，技术应用高度智能化，在不同文化背景中深度兼容化，走向标准化，理论研究高度深入化，信息服务个性化和泛客户化，信息资源和背景高度复杂化，所涉及内容向纵深化方向发展。

## 本章小结

信息是社会发展的重要前提和主要资源，是人类的宝贵财富。信息资源是现代社会最重要的战略资源之一，它的开发利用水平是国家综合国力的体现。信息组织就是根据人类社会发展的需求，以各类信息源为对象，通过对其内容特征等的分析、选择、标引、处理，使其成为有序化集合的信息增值活动，是信息资源开发和利用的重要环节。

## 习题 1

1. 信息组织的含义和特点是什么？
2. 信息组织包括哪些内容？
3. 按信息的运动状态和方式，信息组织可以分为哪几个层次？
4. 中外信息组织发展历史有哪些异同？
5. 信息组织研究前沿领域有哪些？