

■ 李明鑫/著

基于知识链接的 数字资源整合研究

JIYU ZHISHI LIANJIE

DE SHUZI ZIYUAN ZHENGHE YANJIU



东北师范大学出版社

基于知识链接的 数字资源整合研究

JIYU ZHISHI LIANJIE

DE SHUZI ZIYUAN ZHENGHE YANJIU

李明鑫/著

东北师范大学出版社 长 春

图书在版编目 (CIP) 数据

基于知识链接的数字资源整合研究/李明鑫著—2版.
—长春: 东北师范大学出版社, 2015. 3

ISBN 978-7-5681-0376-3

I. ①基… II. ①李… III. ①数字技术—信息
管理—研究 IV. ①G203

中国版本图书馆 CIP 数据核字 (2015) 第 006888 号

☐ 策划编辑: 许革晨

☐ 责任编辑: 许革晨 ☐ 封面设计: 张 然

☐ 责任校对: 王 丹 ☐ 责任印制: 刘兆辉

东北师范大学出版社出版发行
长春净月经济开发区金宝街 118 号 (邮政编码: 130117)

网址: <http://www.nenup.com>

东北师范大学出版社激光照排中心制版
河北省廊坊市永清县晔盛亚胶印有限公司

河北省廊坊市永清县燃气工业园榕花路 3 号 (065600)

2015 年 3 月第 2 版 2015 年 3 月第 1 次印刷

幅面尺寸: 148 mm×210 mm 印张: 8.25 字数: 240 千

定价: 49.00 元

前 言

数字资源整合是一种对原生和转换的数字资源优化组合的存在状态。就是依据一定的需要，对分散无序、相对独立的数字对象进行类聚、融合和重组，使其重新组织为一个新的有机整体，形成一个效能更好、效率更高的新的数字资源体系。其目的是为了更好地揭示资源，以实现各类资源内容间的无缝关联和透明访问，同时将资源纳入到统一的文献信息服务体系中，做到资源与服务的无缝结合，从而为用户提供高效率、个性化的优质知识服务。

通过资源整合可以把不同类型、不同内容、不同组织方式以及不同获取手段的数字资源，以统一的形式提供给资源用户，缩短用户查找信息的时间，简化获取信息的途径。在对数字资源整合十几年的研究中，学者们已经注意到知识整合是数字资源整合发展的必然。资源整合体系包括数据整合、信息整合和知识整合三个层次。数据整合仅解决了异构数据库中信息实体的合并问题，无法反映信息实体间客观存在着的多种联系。如何在异构资源系统集成的基础上，揭示客观存在于不同资源系统中信息实体之间的关系，便成为信息整合要完成的任务。但信息整合的深度仅停留在信息实体层面，并没有揭示信息实体内部的概念和语义，因而从信息整合系统中所获得的还是信息的集合，而不是知识的集合。实现知识的有效获取，满足人们的知识需求，便成为知识整合要完成的任务。数字资源知识整合是整合的高级阶段，是以有效利用知识为目的，以知识管理为基础，知识组织方法为依托，在已有数据整合与信息整合的基础上，充分展示知识体系、知识内容、知识概念及概念关系

的一种整合方式。知识整合是一个动态的过程，最终目标是能够有效利用知识。

随着现代网络技术、信息技术和通信技术的发展，传统的信息服务模式受到了革命性的冲击，信息环境和信息服务发生了巨大变化，用户的信息需求开始从信息服务转向知识服务。相对于通过检索方式直接获取信息的信息服务方式而言，通过链接方式获取信息更符合人类通过“关联”获取信息的天性，更为直接和准确，更接近用户问题，因而逐步成为知识整合的一个主题。相应的，知识链接的需求被提出。知识链接是知识整合的基础，能够实现多个学科、多种类型、多种媒体、多数数据库存储的知识之间的无缝链接，满足知识的互联互通、准确定位和快速获取的功能。

知识链接能够有效发现知识之间的各种关联，能够将具有同一关系、隶属关系和相关关系的知识单元按照特定的需求有序地联系起来。知识链接是基于知识关联的信息组织，能显示知识元或知识信息群之间网络、结构、互动、交叉、演化或衍生等诸多复杂的关系。通过语义关联把知识体系连接为清晰的知识网络结构，为知识检索和知识共享提供知识资源，可以方便用户即时获取有效的知识。

20 世纪 90 年代末期开始，期刊价格上涨导致信息机构期刊采购量下降，新型的科研成果出版模式如电子印刷（e-print）、机构知识库等大量出现，与知识链接直接相关的新的技术标准如 DOI/CrossRef、OpenURL/SFX 的大量出现、用户日益增长的对知识全面链接的期待变得迫切等等，意味着依靠互联网有效的知识链接技术以及相关的标准和规范来解决信息的传播和服务所面临的问题，是无法回避的挑战和发展机遇，抓住了它可能前程无量，错过了它则可能后患无穷。知识链接服务是一种面向知识创新的服务体系，将集合多种知识链接方式来实现全球范围内不同类型知识资源的无缝、开放的链接，成为下一步知识整合的关键所在。

随着互联网的不断普及，数字资源的不断涌现，关联数据的逐步发展，以及信息技术和网络技术的日新月异，用户的知识信息需求和知识学习环境都发生了巨大变化，基于知识链接的数字资源整合作为知识整

合的一种重要整合方式，必将成为组织信息资源以及用户获取知识信息的重要手段。在这样的环境下，本书提出基于知识链接的数字资源整合思想，不仅有助于促使数字资源整合从局部走向整体，更会促进知识的利用，为知识创新的实现提供重要的促进作用。

本书在写作中吸收和借鉴了相关学术领域研究的最新成果，在此向这些研究成果的作者致以谢忱。书中所述为近年本人在教学与研究中的初步成果和一些思考，错误之处在所难免，恳望学界同仁和读者批评指正。

目 录

1	绪论	1
1.1	研究背景与意义	1
1.1.1	研究背景	1
1.1.2	研究意义	3
1.2	国内外研究现状	5
1.2.1	数字资源整合的研究现状	5
1.2.2	知识链接的研究现状	11
1.3	研究目标与结构	17
2	数字资源整合概述	18
2.1	数字资源整合的必要性	19
2.2	数字资源整合的关键技术	21
2.2.1	整合技术的基础——Metadata	21
2.2.2	异构数据库资源整合技术	32
2.2.3	链接访问技术	37
2.2.4	元数据整合技术	38
2.2.5	基于 Web 的资源整合技术	39
2.3	数字资源整合方式	40
2.3.1	基于 OPAC 的整合方式	40
2.3.2	基于数字资源导航的整合方式	41

2.3.3 基于跨库检索的整合方式 42

2.3.4 基于超级链接的信息资源的整合方式 43

2.3.5 基于知识本体的数字资源整合 44

2.4 数字资源整合的发展趋势 45

2.4.1 数字资源整合范围与规模不断扩大 45

2.4.2 知识整合是数字资源整合的主流趋势 46

3 知识关联及其分析方法 48

3.1 知识关联的内涵 48

3.1.1 知识的界定 48

3.1.2 知识单位 50

3.1.3 知识关联的内涵 56

3.2 知识关联的类别与特征 59

3.2.1 知识关联的类别 59

3.2.2 知识关联的特征 64

3.3 知识关联的理论基础 66

3.3.1 唯物辩证理论 66

3.3.2 系统论 68

3.3.3 小世界现象 70

3.3.4 科学统一性原则 72

3.3.5 科学继承性原则 74

3.4 知识关联的方式 75

3.4.1 链表结构 75

3.4.2 树结构 79

3.4.3 图结构 81

3.5 知识关联的分析方法 88

3.5.1 引文分析法 88

3.5.2 决策树分析法 90

3.5.3 语义网络知识推理 93

4 知识链接——资源整合的目标	95
4.1 知识链接概述	95
4.1.1 知识链接的概念	95
4.1.2 知识链接的产生背景	98
4.1.3 知识链接的作用与功能	103
4.2 知识链接的类别与特征分析	106
4.2.1 依据链接范围进行划分	107
4.2.2 依据链接机制进行划分	108
4.2.3 依据链接对象进行划分	111
4.3 知识链接的发展进程与趋势	118
4.3.1 知识链接发展历史	118
4.3.2 知识链接研究进展	123
4.3.3 发展趋势	128
5 知识链接的构建分析	133
5.1 基于文献单元知识链接的构建	133
5.1.1 引文的概念	134
5.1.2 引文的关联类型	135
5.1.3 引文链接	136
5.1.4 引文链接的构建模型	139
5.1.5 引文链接构建中存在的问题分析	144
5.2 基于信息单元知识链接的构建	147
5.2.1 特征信息链接	148
5.2.2 特征信息链接构建模型	149
5.2.3 特征信息链接构建中存在的问题	152
5.3 基于知识单元知识链接的构建	158
5.3.1 知识单元概述	158
5.3.2 知识单元的表达	164
5.3.3 知识元链接理论	168

5.3.4 知识元链接的构建模型	173
5.4 知识链接的构建策略	178
5.4.1 多层次规划知识链接的目标	178
5.4.2 跨系统组织知识链接资源	179
5.4.3 分步骤建立知识链接机制	179
6 基于知识链接的数字资源整合模式	181
6.1 起源于出版商的 DOI+CrossRef	181
6.1.1 DOI 概述	181
6.1.2 DOI 系统组成	184
6.1.3 DOI 系统的应用	189
6.1.4 DOI 系统在图书馆中的应用	194
6.1.5 CrossRef 概述	199
6.1.6 CrossRef 的功能与影响	206
6.2 起源于数据库商 CNKI+知网节	212
6.2.1 CNKI 及知网节技术	212
6.2.2 知网节知识链接	214
6.2.3 知网节的功能	218
6.2.4 存在的问题	220
6.3 起源于图书馆的 OpenURL+SFX	222
6.3.1 OpenURL 框架概述	223
6.3.2 OpenURL 应用现状	231
6.3.3 OpenURL 应用中的问题及相关进展	236
6.3.4 OpenURL 框架的代表性技术 SFX	239
结语	248
参考文献	250

1 绪 论

1.1 研究背景与意义

1.1.1 研究背景

随着信息技术的快速发展，我们现在已经处于具体且形象的知识世界当中。彼得·德鲁克指出：“在新的经济体系内，知识并不是和人才、资本、土地相并列的社会生产资源之一，而是唯一有意义的资源，其独到之处，正在于知识是资源的本身，而非仅是资源的一种。”在知识经济环境下，知识已成为机构的重要资源，知识管理也成为现代管理的核心问题。知识管理作为管理学界的一个热点研究方向，已经扩大到多个学科共同研究，在图书馆学、情报学和计算机科学等领域尤为显著。互联网技术的发展让人们对知识资源的相互关联和无缝链接有了一定的要求，这些资源除了包括同质资源外，还包括异质的、异构的以及异地的知识资源。人们希望能够在不同的知识平台上获取广泛链接的学术资源，并且要求将不同平台内的相关学术资源进行集成整合，实现资源检索的统一性和知识服务的协调性。这就需要对相关的学术资源进行一定的关联性分析，通过 OpenURL、网页、引文等方式实现多元、多角度、多途径的链接，实现不受出版商、图书馆等机构以及数据库收录范围限制的具备开放性和动态性的各种知识资源的有效链接、数据的跨库整合和数据集成。在这种背景下，知识链接在数字资源整合中的价值被

突显出来，越来越多的学者开始展开对知识链接的研究。知识链接是知识服务的基础，能够实现多个学科、多种类型、多种媒体、多数据库存储的知识之间的无缝链接，满足知识的互联互通、准确定位和快速获取的功能。

在发达国家，经济发展更多依赖的是知识和技术的生产、扩散、应用及创新，而我国目前正在努力从“制造大国”向“制造强国”和“设计大国”方向发展，也更加需要充分利用已有的知识资源。但是，一方面日益庞大的信息资源容量使得文献、专利等显性知识范畴内的知识文献在浩瀚的知识海洋中逐步隐性化，人们在获取知识的同时，不得不动阅读大量的无用知识、垃圾知识，很多精力消耗在知识的查找、定位以及提取上，不能一步到位，迅速找到自己需要的知识进行消化吸收并服务于设计、开发等知识增值环节；另一方面，网络信息资源的积累却在相当长的时间内处于“只填充，不维护”的恶性环境中，大量的无用、过时、重复的知识资源充斥其中，有价值知识资源的查找、定位以及提取变得很困难，严重降低了知识资源的利用效率，而且这种现象存在越来越严重的趋势。

早在 1955 年，美国的尤金·加菲尔德博士首次提出了“引文索引”的概念，即将一篇文献作为检索词，通过收录其所引用的参考文献和跟踪其发表后被引用的情况来掌握该研究课题的来龙去脉，从而迅速发现与其相关的研究文献。这可以看作通过知识链接进行文献整合的最早形式。同期的知识组织方法，如体系分类法、标题词法、单元词法、关键词法以及叙词法等对知识管理起到了很大的作用。但是，面对剧烈增长的文献数量，此类以文献单元（文献、标准、专利等）为管理单位进行知识组织研究的方法无法有效地归结文献内容，这就使每个人都去重新学习自己感兴趣的每篇文献，从中提取出对自己有用的东西，甚至有时发现不了对自己有用的内容，从而造成脑力资源的严重浪费，也降低了知识资源的利用效率。因此，单纯以文献单元作为知识管理的基础是远远不够的。20 世纪 70 年代后期，美国专家指出：文献数量膨胀之后，知识的控制单位将从文献深化到文献中的数据、公式、事实、结论等最小的独立的“知识元”，知识元可以被称为文献管理的最小单位。知识

元不仅可以用于情报管理中的文献处理,而且可以表示其他种类知识载体,如专利等,将其中所涉及的概念、论据、论证以及创新点等知识核心以知识元的方式呈现,以此作为知识管理、知识评价及知识发现的最小单元。按照这样的理论,如果将文献管理的单位缩小至知识元级别,人们在获取知识的时候就可以剔除尽可能多的垃圾信息,使信息的纯度更高,知识资源的利用效率也就更明显地提升了。但目前对于知识元的理解存在差别,而且建立在知识元基础上的知识链接也并未得到足够的重视。

数字图书馆是一个数字化系统,它将分散于不同位置、不同载体的信息资源以数字化形式储存,以网络化形式连接,及时提供文化、信息的传递与服务,实现网络用户的资源共享,其关键是资源的数字化和传递的网络化,其实质则是形成有序的信息空间,提供可靠的信息保障。但是客观上,大量数字资源来自不同的检索平台、不同的数据库类型、不同的检索语言、不同的语种、不同的服务形式,并且各数据库的内容也存在着一定的重复和交叉,因此用户在使用时,不但要掌握多种系统的使用方法,而且要反复登录不同数据库,重复输入检索策略,既浪费了大量的时间和精力,检索效率也比较低。

要解决上述问题,就必须进行数字信息资源整合。而将知识链接与资源整合集成,建立基于知识链接的数字资源整合属于资源整合的高级阶段,也是今后的发展趋势。通过知识关联关系分析,借助链接的手段直接将相关资源提供给用户,提高整合的效果,促进用户的资源利用,通过对异构的数字资源进行整合,对比较有效的整合方式进行研究,最终制订出一套的适用的规范标准,这些工作在资源整合的研究方面的意义是重大的。

1.1.2 研究意义

1. 理论研究意义

情报学的理论、方法和技术是进行数字资源建设的重要支柱。信息与情报是情报学的核心问题,由事实、数据、信息、知识、情报五个链环构成。此链环中,“信息”的下游是面向物理属性的,上游是面向认

识属性的。作为中心环节的“信息”既有物理属性也有认知属性，因此成为这一链环的核心。这条链环是从信息的形态出发，它们之间的关系是并列、转化或层次关系。链环所要表达的思想是怎样将信息转化为情报，而本研究谈及的知识链接是知识的分析和深度加工的过程。其要素有三个：链接源、链接目标和知识关联。知识链接技术是数字资源信息处理技术的重要技术，能够通过思维抽象和思想升华使知识增值。

本研究以知识关联为中心，围绕知识关联方式，探讨从物理属性到认知属性的揭示方法。整个过程包括出版商、数据库商、图书馆内部以及之间的知识链接模式。通过对比其他学者提出的数字资源整合方案，提出本文基于知识链接的数字资源整合内容。继而，从理论上构建数字资源整合模式。

2. 现实意义

(1) 加强不同类型知识之间的联系

研究知识关联的类型，通过一定的链接方式将不同类型的知识联系起来，可以揭示知识之间的内在关系，最大限度地保持知识体系的完整性，使之成为一个有机的整体。它打破了不同类型知识之间存储与检索方式不同的壁垒，可以全方位地为用户提供知识资源保障。

(2) 提高知识资源的利用率

通过具有链接功能的数字资源整合系统，为用户节省了大量的检索时间，使用户可以致力于对知识资源本身的利用。同时，由于各类知识资源之间交叉链接的实现，用户可以得到许多以往可能忽略掉的资源，这样无论从效率方面还是效果方面都大大地提高了知识资源的利用率。

(3) 实现真正意义上的数字资源整合

数字资源整合就是信息资源优化组合，根据用户的需要，把相互独立的信息资源之间的关系进行融合和重组，从而形成有机的整体。在知识经济时代，基于知识链接的数字资源整合研究是以数字图书馆为中介，以知识为中心，围绕知识关联，在不同资源之间提供便捷访问通道，推进知识组织，使信息资源能够被高效吸收和利用，在知识型和创新型社会中将发挥越来越大的影响和作用。主要表现在：一方面，剔除重复信息，提高知识含量。避免重复建设和投入，删除垃圾信息和没有

价值的信息,构建完整的知识体系。另一方面,实现效益最大化,满足用户需求。基于知识链接的数字资源整合是信息资源效益最大化的保障,用户可以根据需要在一个界面以统一的标准检索到相关的知识,达到用户的最佳满意度。

1.2 国内外研究现状

本研究将知识链接与数字资源整合进行集成,此提法在现有研究中还不能直接找到相关文献,因此下面分别从数字资源整合的研究现状以及知识链接的研究现状进行介绍。

1.2.1 数字资源整合的研究现状

数字资源整合近年来作为一种研究方法,在理念和情报学等的研究领域进行了大量的论述。从简单的了解、逐渐的认识一直到实际的应用整个过程,数字资源整合的思想精髓和轮廓在我们的眼前逐渐地展现。近年来,人们对于数字资源整合的实际研究和理论理解达到了前所未有的热度。与此同时,数字资源的整合以及对信息检索结果的可视化、元数据、网格等也成为情报学领域当今的前沿课题。

1.2.1.1 国内现状

在国内的图书情报界,以“整合”为研究内容开始于2000年。台湾的陈昭珍教授所著《电子图书馆整合检索之理论与实践》一书分析了数字资源整合检索的需要与模式,介绍了Z39.50协议在图书馆数字资源整合中的应用,并对今后整合检索发展趋势作了一些宏观上的展望。在该研究内容发展初期,在图书情报类杂志上刊登的相关文献数量较少,如:2002年文献大约10余篇,2003年文献大约30篇。随后几年,在图书情报类杂志上刊登的相关文献数量逐年增多,每年的发文量能达到上百篇。可见,有关数字资源整合的研究越来越受到图书情报界,特别是国内学者的重视。从研究内容看,文献的研究内容也越来越深入,从理论研究到模型构建,再到系统、技术实现的实践等。具体研

究主题涉及以下几个方面的内容:

1. 整合理论方面

整合理论的研究集中于初期, 主要围绕数字资源整合的概念提出数字资源整合标准、整合原则和整合的必要性。在马文峰发表的《数字资源整合研究》^①中, 谈到了资源整合的含义、理论基础、原则、主要特征、动因、缘起、立足点和整合方式。该文对数字资源整合的基础理论进行了较为全面的阐述, 指出数字资源整合是数字图书馆建设的关键, 也对以后其他学者的研究起到了很大的引导和借鉴作用, 数字资源整合的研究也逐步走向成熟, 并开始向其他领域渗透。王善平^②从理论上探讨了数字资源整合中标准化的重要性, 认为实现数字信息资源整合的关键在于标准化, 所以应围绕标准化来推进信息资源的整合。

2. 整合技术方面

数字资源整合技术并非只涉及许多计算机界研究课题, 也涉及图书情报界的研究课题。信息资源整合技术在研究初期, 多是和网络技术有关, 随后开始在整合系统的标准和特点、自动标识及整合界面等进行了研究。金更达^③探讨了整合检索界面的设计, 并提出了智能型 Agent 界面的设计思想和实现方法。李勇文、张晓林^④分析了跨网关整合检索和整合浏览的机制, 并介绍了涉及的计算机技术。

3. 整合方式方面

许萍华^⑤总结国内数字资源整合方式主要有四种: 基于 OPAC 的整合, 基于资源导航的整合, 基于系统的整合和基于服务的整合。赵建

① 马文峰. 数字资源整合研究 [J]. 中国图书馆学报, 2002 (4): 64—67.

② 王善平. 论数字信息资源的整合与标准化 [J]. 情报资料工作, 2002 (6): 19—21.

③ 金更达. 网络资源界面整合和 Agent 界面实现探讨 [J]. 大学图书馆学报, 2002 (1): 30—34.

④ 李勇文, 张晓林. 跨网关整合检索和整合浏览机制 [J]. 图书情报工作, 2002 (9): 74—78.

⑤ 许萍华. 国内数字资源整合理论研究和开展现状分析 [J]. 图书情报工作, 2005, 1 (5).

平^①认为数字资源整合的方式可分为宏观和微观两个方面：宏观上分为实体法和虚体法；微观上分为物理整合和逻辑整合；基于数字图书馆的整合方式；面向内容的整合方法；面向对象的整合方法；面向任务的整合方法。其中，基于数字图书馆的整合方式为四个层次：汇合整合方式；组合整合方式；重组整合方式；一体化综合整合方式。

4. 整合模式方面

章成志、苏新宁^②在《信息资源整合的建模与实现方法研究》中给出信息资源整合的基本框架，信息资源整合的三维模型及各个维度的含义和功能，给出信息资源整合实现的四层体系结构，分别说明各个层次的整合方法，并提出信息资源整合的原则和方法。王爱丽^③将数字资源整合模式分为四种：基于 OPAC 系统的资源整合；基于数据源的整合；基于知识整理的数字资源整合；OAI 标准整合模式。

5. 知识整合方面

当知识逐渐成为数字图书馆发展的主题，将知识相关理论和方法应用到数字资源整合中是一种必然趋势。围绕基于知识的整合研究内容分为：基于知识组织的整合研究、基于知识管理的整合研究、基于知识本体的整合研究、基于知识服务的整合研究和知识需求的整合研究、基于知识集成的整合研究等。马文峰在《基于知识组织理论之上的数字资源整合》一文中，对比信息组织与知识组织差异性，数字资源整合是数字图书馆建设的关键，应将知识组织理论与方法运用于数字资源整合实践中。邓三鸿等人^④结合知识集成的内容、特点、要求，提出了知识管理理念下的数字资源整合策略：基于主题的资源整合，面向问题的资源整合和基于知识地图的资源整合。

① 赵建平. 数字信息资源整合方式比较研究 [J]. 情报科学, 2008 (12): 1787-1791.

② 章成志, 苏新宁. 信息资源整合的建模与实现方法研究 [J]. 现代图书情报技术, 2005 (10): 60-63.

③ 王爱丽. 数字资源整合技术与模式研究 [J]. 图书馆学研究, 2006 (3): 60-62.

④ 邓三鸿, 金莹, 秦嘉杭. 基于知识管理的数字资源整合策略 [J]. 情报科学, 2006, 24 (10): 1489-1493.