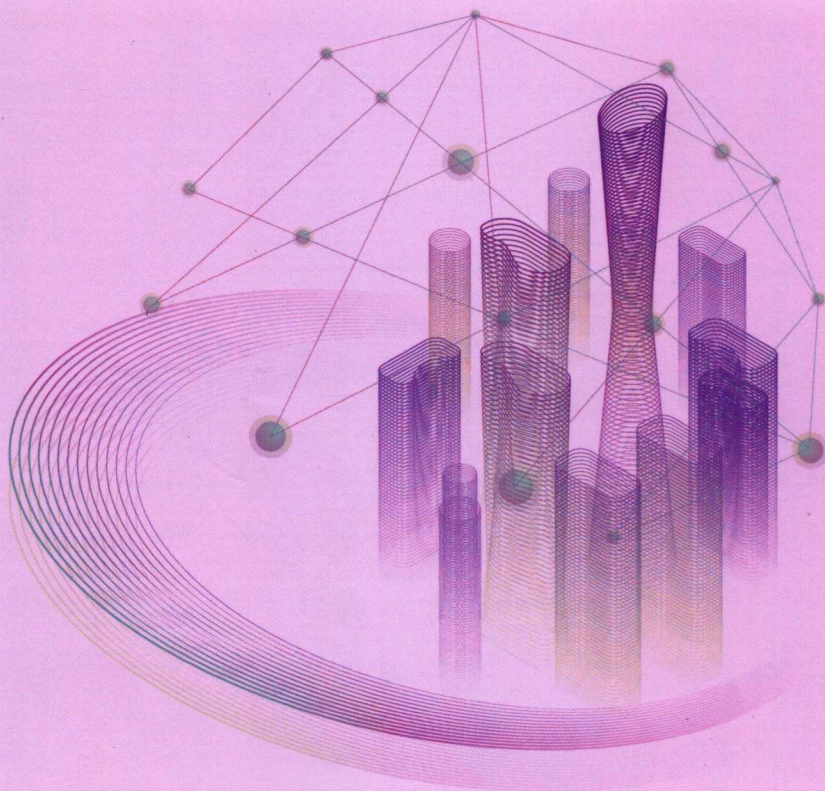


面向网络社区的 领域知识聚合研究

陈 果 ◎ 著



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

非外借

江苏省社会公共安全科技协同创新中心项目资助

面向网络社区的领域知识聚合研究

陈 果 著



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

· 北 京 ·

图书在版编目 (CIP) 数据

面向网络社区的领域知识聚合研究 / 陈果著. —北京: 科学技术文献出版社, 2019. 6

ISBN 978-7-5189-5224-3

I. ①面… II. ①陈… III. ①数字图书馆—图书馆服务—研究 IV. ①G250.76

中国版本图书馆 CIP 数据核字 (2019) 第 026080 号

面向网络社区的领域知识聚合研究

策划编辑: 孙江莉

责任编辑: 李 鑫

责任校对: 张叫噪

责任出版: 张志平

出 版 者 科学技术文献出版社

地 址 北京市复兴路15号 邮编 100038

编 务 部 (010) 58882938, 58882087 (传真)

发 行 部 (010) 58882868, 58882870 (传真)

邮 购 部 (010) 58882873

官方网 址 www.stdp.com.cn

发 行 者 科学技术文献出版社发行 全国各地新华书店经销

印 刷 者 北京虎彩文化传播有限公司

版 次 2019年6月第1版 2019年6月第1次印刷

开 本 710×1000 1/16

字 数 269千

印 张 17

书 号 ISBN 978-7-5189-5224-3

定 价 78.00元



版权所有 违法必究

购买本社图书, 凡字迹不清、缺页、倒页、脱页者, 本社发行部负责调换

前 言

本书从网络社区用户知识交流特征出发,提出一种用户需求的知识聚合方案。网络社区知识组织的主要依据是用户交流内容中的主题及其对应的领域概念关联,因此网络社区知识聚合的核心是构建面向特定领域的概念关联体系。传统知识组织体系不能很好地满足网络社区中的知识聚合需求,因此本书研究的重点为通过融合已有的领域背景知识体系和基于UGC内容挖掘生成的知识关联体系,构建领域多元概念关联体系,以此为支撑实现面向网络社区的多维知识聚合。本书主要从以下几个方面对网络社区的领域知识聚合进行了研究。

第一,研究网络社区知识聚合的基础问题。首先,在分析网络社区用户交流特征的基础上,讨论了其知识聚合需求;其次,在回顾网络社区知识聚合发展路径的基础上,重点探讨了概念知识关联的特征与形式,并以此为基础提出了网络社区知识聚合的模式。

第二,以领域知识分析为视角,讨论基于领域概念关联的网络社区知识聚合模型构建问题。在引入领域知识分析视角的基础上,进一步讨论了以领域知识为背景的领域概念关联的特征,以及其在知识组织中的作用机制。在此基础上,提出了包括领域多元概念关联体系构建流程和基于概念关联的网络社区知识聚合实现方案在内的总体模型。

第三,研究领域结构化概念关联体系的构建。一方面,通用



性和专业性的结构化知识组织体系（分类法、主题词表、本体、主题图等）为构建面向特定领域的结构化概念关联体系提供了基础，因此从已有的结构化概念关联体系中提取领域概念单元、合并概念关联是最直接的方案；另一方面，大多数领域都缺乏可利用的有针对性的结构化知识组织体系，因此从其他资源途径获取领域概念单元和概念关联有其必要性。本书讨论了基于用户用词领域概念单元抽取、基于资源内容挖掘的领域概念单元抽取、基于资源挖掘的领域概念关联识别等问题；并进一步阐述了如何将这些从其他资源途径获取的领域概念关联体系基本元素与已有的概念关联体系进行合并，包括概念映射和标准化、概念关联合并等问题。领域结构化概念关联体系的形式化是后续知识利用的基础，通过对比分析 SKOS 和 OWL 2 两种语言对概念关联体系的表述细节，本书选定基于 OWL 2 语言的领域结构化概念关联体系形式化方案。为了检验上述思考和方案的有效性，本书开展了基于主题词表和百科内容挖掘的心血管领域结构化概念关联体系构建实验。

第四，研究领域共现型概念关联体系的构建。其思路是从网络社区用户交流内容中挖掘概念关联。文本图模型为概念单元在用户交流内容中的共现关系表示提供了理论基础，通过探讨基于文本图模型的领域概念共现网络构建流程，本书将领域共现型概念关联体系划分为两大阶段。一阶段是由用户交流文本集向概念共现关系矩阵转化，核心步骤包括文本集构建、中文切分词与同义词合并、领域术语识别、概念共现关系提取与共现矩阵构建；另一阶段是概念共现关系矩阵向概念共现网络的转化，核心步骤包括概念共现关系强度的标准化、概念相似度计算、基于相似度



的概念关联边强度映射和概念共现网络生成。根据上述思路和方案,本书以著名的医学社区丁香园为例,选定其中影响力较大的心血管论坛作为对象,构建了基于用户交流内容挖掘的心血管领域共现型概念关联体系。

第五,研究领域多元概念关联体系的融合。本部分研究开展的前提是探讨结构化概念关联体系和共现型概念关联体系融合的机制,在对两者优劣势分析的基础上提出基于优势互补的融合原则,并制定将结构化概念关联体系中的概念内涵、概念和关联类型、层次组织结构赋予共现型概念关联体系的融合方案。融合方案的实现主要体现为体系间元素映射,包括概念术语映射和语义关系映射,在语义映射部分重点探讨了语义相似度融合方案和细粒度关系发现两个问题。此后,通过对多元概念关联体系融合后的数据关系进行梳理,确立了基于关系数据模型的多元概念关联体系形式化方案。在实验部分,对已构建的心血管领域结构化概念关联体系和共现型概念关联体系进行了融合。

第六,研究基于多元概念关联体系的网络社区知识聚合实现问题。在探讨了多元概念关联体系对网络社区知识聚合作用机制的基础上,提出了以知识单元粒度和知识聚合形式为维度的多维知识聚合模式。此后,分别对不同知识单元粒度的知识类聚和知识共聚方案与实现方法进行了探讨。知识共聚方案是本书有别于传统知识聚合形式的创新点,以概念单元的多维度和细粒度关联为支撑,可实现网络社区分面检索与导航、主题多维推荐、知识元链接体系构建和资源关联发现等功能。另外,专门针对丁香园社区心血管论坛的知识聚合方案实现进行了实验。

目 录

第一章 绪 论	1
1.1 研究背景	1
1.2 研究意义	2
1.2.1 理论意义	2
1.2.2 现实意义	3
1.3 国内外研究现状分析	4
1.3.1 网络社区中用户交流特征与内容挖掘研究	5
1.3.2 领域概念关联体系构建	8
1.3.3 基于领域概念关联的知识组织研究	11
1.3.4 知识聚合相关研究	14
1.3.5 研究述评	18
1.4 研究内容、方法和创新点	19
1.4.1 研究内容	20
1.4.2 研究方法	23
1.4.3 研究创新点	24
第二章 面向网络社区的知识聚合模式	25
2.1 网络社区中的信息交流特征与知识组织需求	25
2.1.1 网络社区用户交流的内容特征	25
2.1.2 网络社区中的知识组织需求	27
2.2 面向网络社区的知识聚合发展	30
2.2.1 从信息聚合到知识聚合	31
2.2.2 从馆藏资源知识聚合到网络社区知识聚合	32
2.3 知识关联与概念知识关联	34
2.3.1 知识关联的内涵	35



2.3.2 知识关联的类型	36
2.3.3 概念知识关联及其类型	38
2.4 网络社区的知识聚合模式	41
2.4.1 网络社区知识聚合的知识单元粒度	41
2.4.2 网络社区知识聚合的形式	47
第三章 领域分析视角下的网络社区知识聚合模型	51
3.1 知识组织中的领域分析视角	51
3.1.1 领域知识的内涵	51
3.1.2 领域知识分析范式及其对知识组织的影响	53
3.2 领域知识分析视角下的概念关联	56
3.2.1 领域知识背景对概念关联的作用	57
3.2.2 领域概念关联的特征	59
3.3 领域概念关联对网络知识组织的作用机制	60
3.3.1 领域概念关联在知识类聚中的作用机制	61
3.3.2 领域概念关联在知识链接体系中的作用机制	64
3.4 基于领域概念关联的网络社区知识聚合模型构建	68
第四章 领域结构化概念关联体系构建	71
4.1 基于概念关联的结构化知识组织体系	71
4.1.1 结构化知识组织体系及其类型	71
4.1.2 概念类聚体系	72
4.1.3 概念关联体系	76
4.2 领域结构化概念关联体系的构建流程与关键问题	80
4.2.1 领域结构化概念关联体系的构建流程	80
4.2.2 领域概念单元的获取	82
4.2.3 领域概念关联的抽取	87
4.3 结构化概念关联体系间的语义映射	91
4.3.1 结构化概念体系的概念映射与标准化	91
4.3.2 结构化概念体系间的概念关联合并	92
4.4 领域结构化概念关联体系的语义形式化	94
4.4.1 SKOS 语言体系	94



4.4.2	OWL 2 语言体系	98
4.4.3	结构化概念体系关系数据模型的语义形式化	103
4.5	实验：基于主题词表和百科内容挖掘的心血管领域 结构化概念关联体系构建	106
4.5.1	基于医学主题词表的心血管基本概念关联体系构建	107
4.5.2	基于百科内容抽取的心血管概念关联体系构建	112
4.5.3	结构化概念关联体系间语义内容映射	115
4.5.4	心血管概念关联体系的 OWL 2 形式化表示	121
第五章	领域共现型概念关联体系构建	125
5.1	基于文本图模型的概念共现关系表示	125
5.1.1	文本中的概念共现关系	125
5.1.2	文本图模型及其形式化表达	126
5.1.3	文本中概念共现关系的边权重计算	127
5.1.4	基于文本图模型的领域概念共现网络构建流程	128
5.2	领域概念共现矩阵的生成	129
5.2.1	网络社区文本集的构建	129
5.2.2	中文切分词与同义词合并	129
5.2.3	领域术语识别	131
5.2.4	概念共现关系提取与共现矩阵构建	133
5.3	领域概念关联网络构建	134
5.3.1	概念共现关系强度的标准化处理	134
5.3.2	基于共现关系的概念相似度计算	135
5.3.3	基于相似度转化的概念共现网络生成	136
5.4	实验：心血管领域共现型概念关联体系构建	137
5.4.1	用户交流内容采集和预处理	137
5.4.2	基于 ICTCLAS 的用户交流内容切分词处理	139
5.4.3	同义术语映射与信息合并	143
5.4.4	融合 TF-IDF 和词长加权的领域术语提取	145
5.4.5	领域术语的共现关系提取与矩阵生成	148
5.4.6	概念相似度计算与概念共现网络构建	149



第六章 领域多元概念关联体系融合	153
6.1 多元概念关联体系的融合机制	153
6.1.1 基于优势互补的多元概念关联体系融合原则	153
6.1.2 多元概念关联体系融合的主体问题	157
6.1.3 多元概念关联体系融合的形式问题	160
6.2 多元概念关联体系融合中的概念术语映射	161
6.2.1 同义术语的映射	162
6.2.2 同形异义术语定位	163
6.2.3 概念相关信息的合并	165
6.3 多元概念体系融合中的语义关系映射	165
6.3.1 语义关系映射规则	166
6.3.2 等级关系映射与关系强度计算	167
6.3.3 细粒度相关关系的映射与发现	169
6.4 多元概念关联体系融合的形式化表达	171
6.4.1 多元概念关联体系融合后的数据关系	171
6.4.2 基于关系数据模型的多元概念关联体系融合的形式化	172
6.5 实验：心血管领域结构化概念体系与共现型概念 体系的融合实现	175
6.5.1 概念映射与信息合并	175
6.5.2 术语等级关系映射与相似度融合计算	177
6.5.3 术语细粒度相关关系映射与相似度融合计算	178
6.5.4 术语细粒度相关关系发现	180
第七章 基于领域概念关联的网络社区知识聚合实现	184
7.1 基于领域概念关联的网络社区知识聚合模式	184
7.1.1 领域概念关联体系在网络社区知识类聚中的作用	184
7.1.2 领域概念关联体系在网络社区知识共聚中的作用	185
7.1.3 网络社区多维知识聚合模式	187
7.2 基于领域概念关联体系的网络社区知识类聚实现	188
7.2.1 聚类方法	188
7.2.2 网络子群划分方法	192



7.2.3 基于领域概念关联体系的网络社区主题类聚实现	196
7.2.4 基于领域概念关联体系的网络社区文本类聚实现	198
7.3 基于领域概念关联体系的网络社区知识共聚实现	199
7.3.1 基于领域概念关联体系的网络社区分面检索与 导航体系实现	200
7.3.2 基于领域概念关联体系的网络社区主题多维关联 推荐体系实现	201
7.3.3 基于领域概念关联体系的网络社区知识元链接体系实现 ...	202
7.3.4 基于领域概念关联体系的网络社区资源关联发现	203
7.4 实验：丁香园社区心血管论坛的知识聚合实现	206
7.4.1 心血管论坛中的主题分面导航与检索体系实现	206
7.4.2 心血管论坛中的主题多维推荐体系实现	209
7.4.3 心血管论坛中的知识元链接体系实现	211
7.4.4 心血管论坛中的非相关文献知识发现思路验证	213
第八章 总结与展望	215
8.1 研究总结	215
8.2 研究不足	217
8.3 研究展望	218
附录 实验核心代码	220
附录1 用 JAVA 调用 ICTCLAS 分词模块	220
附录2 名词提取及处理（计算 TF 值、IDF 值，生成名词共现 矩阵）	222
附录3 领域术语的共现关系提取	228
附录4 领域术语的共现关系统计	232
参考文献	236

图表目录

图 1-1	基于概念关联的网络社区知识聚合相关研究体系	19
图 1-2	本书研究的总体框架	21
图 2-1	“手机”检索结果引导中的知识类聚和知识共聚方式差异	49
图 2-2	基于句子相似度的知识类聚和基于句子逻辑关系的 知识共聚对比	50
图 3-1	概念“铁”在引入化学领域背景知识后的内涵	58
图 3-2	维基百科（中文）中条目“时间”的内容组织	65
图 3-3	维基百科（中文）中对与“时间”相关概念的组织	66
图 3-4	Google 中检索“Wuhan University”展现的知识图谱内容	67
图 3-5	基于领域概念关联的网络社区的知识聚合模型	69
图 4-1	《中国分类主题词表》网络版界面	77
图 4-2	基于概念语义关联的结构化知识组织体系构建流程	81
图 4-3	领域概念单元获取时需要兼顾的术语类型	82
图 4-4	基于非结构化资源的中文领域概念术语抽取流程	87
图 4-5	基于统计方法的概念关系提取流程	89
图 4-6	基于概念映射的等同关系合并	92
图 4-7	基于等级定位的等同关系发现	93
图 4-8	结构化概念关联体系中的层级关系横向合并、纵向合并	93
图 4-9	层级关系冲突时的“从细不从宽”策略	94
图 4-10	OWL 2 的主要构成模块及模块间关系	98
图 4-11	OWL 2 中可用的 XML Schema 数据类型	100
图 4-12	结构化概念关联体系语义化的关键步骤	104
图 4-13	概念关联体系关系数据模型与 OWL 2 数据模型的元素 对应	104
图 4-14	CMesh 中的主题导航结构	108
图 4-15	心血管领域对应的主题词表内容（部分）	108



图 4-16	心血管相关器官组织的层级关系	109
图 4-17	主题词内容页面中的款目词与参照词提取	110
图 4-18	“高血压”疾病相关概念术语与关联采集	114
图 4-19	两个结构化概念体系中“冠心病”概念的等同关系合并	118
图 4-20	两个结构化概念体系中“冠心病”概念的层级关系合并	119
图 4-21	两个结构化概念体系中“冠心病”概念的相关关系合并	120
图 5-1	基于文本图模型的领域概念共现网络构建流程	128
图 5-2	文本中共现关系的邻接矩阵生成方式	133
图 5-3	内容采集代码结构和最终入库结果	138
图 5-4	切分词后的名词术语频次分布	143
图 5-5	丁香园社区心血管论坛中用户交流内容的主题共现关系 强度 (2 以上) 分布	148
图 5-6	心血管领域概念关系共现频次矩阵 (部分)	149
图 5-7	基于等价系数转化的心血管领域概念关系共现 矩阵 (部分)	149
图 5-8	等价系数矩阵转化为 Pearson 相似矩阵的结果 (部分)	150
图 5-9	Pearson 相似矩阵转化为网络边强度的结果 (部分)	151
图 5-10	部分概念的关联网络可视化 (含弱相关关系)	151
图 5-11	部分概念的关联网络可视化 (不含弱相关关系)	152
图 6-1	基于优势互补的网络社区概念关联体系融合方式	161
图 6-2	同义词匹配中的缩写语与全称匹配流程	163
图 6-3	基于语义上下文的同形异义术语定位示例	164
图 6-4	基于语义上下文的同形异义术语定位流程	164
图 6-5	直接等级关系映射和间接等级关系映射	167
图 6-6	等级关系映射与关系强度融合示例	168
图 6-7	等级关系映射与关系强度融合流程	169
图 6-8	细粒度相关关系的直接映射方式示例	170
图 6-9	基于约束规则和强共现的细粒度相关关系映射方式示例	170
图 6-10	领域知识涉及的实体和实体关系	172
图 6-11	领域多元概念关联体系的关系模型	173
图 6-12	同义词映射后多元概念体系间概念的信息合并	176
图 6-13	“冠心病”在结构化概念关联体系中的等级关系	177



图 7-1	基于领域概念关联体系的网络社区多维知识聚合模式	187
图 7-2	K-means 聚类流程示例	189
图 7-3	典型的层次聚类结果可视化示例	191
图 7-4	GN 算法的一个示例	194
图 7-5	Fast unfolding 算法的示例	196
图 7-6	基于领域概念关联体系的网络社区主题类聚流程	197
图 7-7	基于领域概念关联体系的网络社区文本聚类流程	198
图 7-8	基于领域概念关联体系的网络社区知识共聚方式	200
图 7-9	基于概念类型的检索、导航系统分面式改造	201
图 7-10	基于细粒度概念关联的多维主题关联	202
图 7-11	网络文本中的知识元链接示例	203
图 7-12	网络社区中基于领域概念关联体系的知识元链接体系构建 ..	204
图 7-13	网络社区中基于领域概念关联体系的资源关联发现	205
图 7-14	丁香园社区中主题导航方式	207
图 7-15	丁香园社区中按疾病推荐的结果	209
图 7-16	心血管论坛中用户交流内容示例	211
图 7-17	领域概念关联体系指导下的用户内容切分词和概念识别	212
图 7-18	典型的知识元内容页示例 (冠心病)	212
图 7-19	用户交流内容中的知识元链接标引	213
表 3-1	领域知识分析的 10 种典型方法	55
表 3-2	概念“水”在不同专业领域中所侧重的本质特征差异	57
表 4-1	结构化知识组织体系的类型划分	73
表 4-2	基于词形规则模板的层级关系抽取	89
表 4-3	心血管领域主题词对应的款目词和参照词 (部分)	110
表 4-4	基于语法规则模板和词性规则模板的概念关联提取	113
表 4-5	融合百科结构化信息与内容模板的概念关联提取 结果 (部分)	114
表 4-6	标准化后的同义术语概念映射 (部分)	117
表 5-1	Pearson 相关系数值、相关性强度和概念节点边强度对应	136
表 5-2	不同内容长度下无意义帖子的数量抽样统计	138
表 5-3	从结构化概念关联体系中补充的分词词典 (部分)	140



表 5-4	心血管领域用户交流内容中 ICTCLAS 识别的前 100 个新词 ...	144
表 5-5	基于字面匹配的同义术语映射结果 (以“冠心病”为例)	144
表 5-6	TF-IDF 值排名前 50 位的领域术语	146
表 5-7	术语长度加权后的 TOP 50 高领域权值词	147
表 5-8	最终选定领域术语的频次分布	148
表 5-9	以 Pearson 相关系数为依据筛选的“冠心病”相关概念	150
表 6-1	共现型概念关联体系中概念术语的类型统计	176
表 6-2	“冠心病”与相关概念在结构化概念关联体系中的 语义关系强度	178
表 6-3	“冠心病”与相关概念在共现型概念关联体系中的 Pearson 相似系数	178
表 6-4	多元概念关联体系中“冠心病”等级关系的相似度融合	178
表 6-5	“冠心病”的细粒度相关关系	179
表 6-6	多元概念关联体系中“冠心病”细粒度相关关系的 相似度融合	179
表 6-7	心血管领域概念关联约束条件	180
表 6-8	与“冠心病” Pearson 相似度大于 0.20 且类型明确的概念	181
表 6-9	多元概念体系融合中的细粒度概念关联发现	183
表 7-1	心血管社区顶层分面导航结果	208
表 7-2	点击“冠心病”后所得的动态分面导航结果	208
表 7-3	基于领域概念关联体系的“冠心病”主题多维推荐结果 (共现关系相似度)	210
表 7-4	基于领域概念关联体系的“冠心病”主题多维推荐结果 (融合相似度)	210
表 7-5	“冠心病”与“冠脉造影”在用户交流文档中出现情况	213

第一章 绪 论

1.1 研究背景

社会网络环境下,用户的知识获取、交流和利用越来越依托于一定的网络社区。基于用户兴趣或特定目标而形成的具有领域专业性的网络社区发展十分迅速,目前已涌现出一大批具有代表性的网络社区。例如,在用户电影、书籍、音乐等兴趣交流基础上形成的豆瓣网,在用户对电子消费产品交流基础上形成的中关村在线数码论坛,以经济管理类专业人员为基础形成的人大经济论坛,以专业医生业务交流为基础形成的丁香园医学社区等。

一方面,网络社区具有开放性和目标性,社区用户之间通过知识交流获得一定的网络社会资本,进而对其知识贡献和知识交流行为起到了促进作用,网络社区知识增长十分迅速,知识累积量相当可观;另一方面,网络社区中的用户交流内容(UGC)通常依托于现实问题的解决,包含丰富的、兼具时效性和专业性的第一手经验知识,因而尽管在表达形式、严谨性等方面存在一定欠缺,但仍然是当前各领域宝贵的“知识库”,网络社区中资源已成为用户解决现实问题时使用最频繁的网络资源类型。对网络社区中的用户交流内容经过一定的整理、挖掘和组织后,可实现更深层的知识服务。例如,百度知道平台开展的问答配对服务;豆瓣社区在用户电影交流基础上整理挖掘的电影库和进一步开展的推荐服务等。

在当前大数据背景下,为促进数字化信息资源的有效利用,包括计算机学科、图书情报学科在内的诸多学科学者都致力于开展知识关联挖掘研究,因而涌现出一大批有针对性的理论与方法,如本体、潜在语义关联分析、复杂网络等,将传统的数字资源组织和服务推向更深层次的基于语义关联的知识组织与服务。以图情学科为例,近几年涌现出一大批以“聚合”为题的各种国家项目,推进了相关理论方法的深入和突破,并取得了丰富的成果。然而,图情学科的相关研究都是围绕数字学术资源展开,以提供学科知识服



务为主。网络社区中用户产生的内容与学术资源特征迥异，特别是其内容表达存在随意性、缺乏有效的内容标注和外部属性等。显然，以学术资源为对象的知识关联挖掘、知识聚合理论和方法，在一定程度上并不适用于网络社区。由此提出了本书所关注的几个重要问题：面向网络社区的知识关联挖掘和知识聚合，与面向学术资源的有何不同？当前针对学术资源挖掘和聚合的理论方法，有哪些可以借鉴，哪些需要改进？如何引入并改进其他领域的理论方法，以将其拓展到网络社区用户内容组织上？

另外，在数字资源的深层聚合中，知识关联的挖掘和利用是关键，而知识关联又以知识概念间的关联最为基础。在揭示数字资源内容主题、计算资源相似度、实现资源语义聚类、推荐相关资源、构建资源导航体系中，领域概念不仅是最基本的知识单元，还是表征其他知识单元（文档、用户、群体、类目）内容的关键特征项。因此，无论是针对何种类型的数字资源（学术文献、用户交流内容），要实现深层知识聚合，必须深入到语义关联层面，而概念间的语义关联挖掘则是各种数字资源共有的。在此背景下，需要以网络社区内容为对象，从领域概念关联的层面出发开展知识聚合研究。

综上所述，“面向网络社区的领域知识聚合研究”主要集中于解决面向网络社区的领域概念知识体系构建和基于领域概念关联的知识聚合实现等问题。

1.2 研究意义

从领域概念关联的层面出发进行知识聚合研究，其意义不仅在于对已有知识聚合理论和方法的拓展与深入，更重要的是具有现实应用价值。

1.2.1 理论意义

数字资源的知识关联挖掘和知识组织一直是当前多个学科的研究重点。然而，目前对于网络社区中用户交流内容的资源挖掘和组织研究较少，探讨面向网络社区用户内容的深度挖掘和内容聚合，是网络社区发展的必然要求，也是对当前相关研究的拓展和深入，因而具有一定理论研究意义。

（1）将知识聚合理论与方法拓展到网络社区知识组织

一方面，当前知识聚合理论方法多以数字学术资源为对象，相应的资源和服务对象在一定程度上具有特殊性，由此产生的理论方法在普适性上有待