

语义网 与数字图书馆

The Semantic Web
and Digital Libraries

欧石燕 著

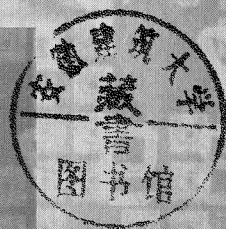


南京大学出版社

语义网 与数字图书馆

The Semantic Web
and Digital Libraries

欧石燕 著



本著作系国家社科基金重点研究项目“基于关联数据的学术文献内容语义发布及其应用研究(17ATQ001)”，国家社科基金一般项目“基于SOA架构的术语注册和服务系统构建与应用研究(11BTQ023)”和教育部人文社科基金一般项目“数据关联的语义数字图书馆研究(10YJA870014)”的研究成果之一。



南京大学出版社

图书在版编目(CIP)数据

语义网与数字图书馆 / 欧石燕著. —南京: 南京大学出版社, 2017.12

ISBN 978-7-305-19584-6

I. ①语… II. ①欧… III. ①语义网络—研究 ②数字图书馆—研究 IV. ①TP18②G250.76

中国版本图书馆 CIP 数据核字(2017)第 274024 号

出版发行 南京大学出版社
社 址 南京市汉口路 22 号 邮 编 210093
出 版 人 金鑫荣

书 名 语义网与数字图书馆
著 者 欧石燕
责任编辑 卢文婷

照 排 南京紫藤制版印务中心
印 刷 江苏凤凰通达印刷有限公司
开 本 787×1092 1/16 印张 22.25 字数 430 千
版 次 2017 年 12 月第 1 版 2017 年 12 月第 1 次印刷
ISBN 978-7-305-19584-6
定 价 66.00 元

网 址 <http://www.njupco.com>
官方微博 <http://weibo.com/njupco>
官方微信 njupress
销售咨询 (025)83594756

* 版权所有,侵权必究

* 凡购买南大版图书,如有印装质量问题,请与所购图书销售部门联系调换

前 言

自 20 世纪 90 年代以来,数字图书馆这一综合研究领域在世界各地蓬勃兴起并取得了巨大发展。随着数字图书馆基础设施建设和遗留资源数字化的基本告一段落,研究人员和从业者开始更多地关注如何保证在分布式异构的数字图书馆环境中,人们能够正确而全面地获得所需的信息与知识,这涉及当前数字图书馆关于资源组织和利用的几个难题:① 信息在局部范围得到组织但在整体上并不相互联系,形成了许多分散独立的信息孤岛;② 无法对不同的信息系统实现统一的访问;③ 无法通过机器对信息进行语义层面的操作。

进入 21 世纪以来,随着语义网的出现和发展,尤其是关联数据的兴起,图书馆和相关机构逐渐意识到语义网技术在解决数字图书馆资源发布、集成、语义互操作、智能检索等方面的巨大潜力,大力推广包括关联数据在内的语义网技术在数字图书馆中的应用。2010 年 5 月万维网联盟(W3C)成立了“图书馆关联数据孵化小组”,同年 8 月,国际图联(IFLA)也成立了“语义网兴趣小组”,分别从图书馆社区外部与内部共同推动语义网和关联数据在图书馆领域的应用,增强图书馆数字资源的全球互操作性。目前图书馆数据已经成为关联数据云的一个重要来源,大量书目数据(如 LIBRIS、WordCat.org)、词表数据(如 LCSH、AGROVOC)和学术论文数据(如 DBLP、CiteSeer)被发布为关联数据,约占整个关联开放数据(LOD)云的 13.3%。

我国图书情报界对语义网和关联数据的认识和研究则比较晚。自 2010 年起,国内才大量出现关于关联数据的论文,主要是对关联数据这一概念及其研究与应用现状进行介绍,或者对关联数据在图书馆

中的应用模式进行理论探讨,而实践研究(尤其是大规模研究)则比较少。在此背景下,本书作者欧石燕在2010年至2015年间主持并完成了两个研究项目:教育部人文社科基金项目“数据关联的语义数字图书馆研究”和国家社科基金项目“基于SOA架构的术语注册与服务系统构建与应用研究”,这两个项目属于国内图书情报领域较早关于关联数据的实践研究,本书主要是在这两个项目研究成果基础上编著而成,同时也包含了2017年国家社科基金重点研究项目“面向关联数据的学术文献内容语义发布及其应用研究(17ATQ001)”的部分最新研究成果。

教育部人文社科基金项目“数据关联的语义数字图书馆研究”的研究目的是将语义网的最新进展——关联数据技术——应用到数字图书馆中,构建一个语义数字图书馆原型,实现数字图书馆中以文献资源为核心的各类资源的语义关联和整合,并支持具有语义功能的浏览与检索。其学术价值在于:探索语义网和关联数据技术在数字图书馆中的应用模式,提出了一个基于本体与关联数据的语义数字图书馆资源描述与组织框架,这一理论模型对关联数据在我国数字图书馆中的应用具有重要指导意义。其应用价值在于:提供了采用语义网和关联数据技术构建语义数字图书馆的总体解决方案,能够用来解决数字图书馆中资源描述、发布、集成、语义互操作、语义检索等各类实际问题,可在各类数字图书馆中推广应用。

国家社科基金项目“基于SOA架构的术语注册与服务系统构建与应用研究”的研究目的是采用语义网技术和Web服务架构,设计和构建面向中文词表的术语注册与术语服务系统,并探索术语服务的应用机制。其学术价值在于:制定了术语注册与术语服务的相关标准与规范,并基于面向服务架构(Service Oriented Architecture,简称SOA)原则设计了术语注册与术语服务的体系架构,为我国面向中文词表的注册与服务系统的开发提供了理论指导。其应用价值在于:

① 对术语注册与术语服务系统中关键组件的实现技术进行了探索,

为相关系统的开发提供了实践经验和技术解决方案;②对术语服务在信息检索中的应用模式进行了探索,对相关应用具有借鉴和指导意义。

国家社科基金重点项目“面向关联数据的学术文献内容语义发布及其应用研究”的目的是对语义出版模式下学术文献内容的关联数据构建和应用进行研究,旨在通过对学术文献的结构和内容进行知识建模,对文献中的语义元素进行抽取和语义标注,实现文献知识内容的模块化和结构化发布,从而促进知识理解和交流、提高科研效率,解决大数据环境下信息过载的难题,并有助于信息机构实现从信息服务到知识服务的提升。

除项目内容外,本书还对语义网和关联数据技术在国内图书馆领域的应用进行了总结和梳理。当前,图书馆界采用语义网和关联数据技术主要是对文献资源的外部特征(主要是书目特征)进行语义描述、语义关联和语义发布,从而在文献资源的整合和检索方面有了很大提升。一方面,能够实现书目数据之间的语义整合与互操作,以及文献资源与其他相关资源之间的无缝语义链接;另一方面,提供具有语义功能的检索,能够基于概念对图书馆各类资源进行多角度的复杂查询。但是,这仍旧只是停留在元数据层面对文献资源进行描述、组织与检索,向用户提供的还只是信息服务,无法上升到文献的内容与知识层面。在未来工作中,如何深入文献资源内容层面对其中的知识内容进行表示、发布、关联、整合和推理是值得考虑和研究的问题。一方面采用自然语言处理和文本挖掘技术,对文献资源的结构和内容进行解析与处理,从中提取出有用的知识内容;另一方面采用语义网和关联数据技术对知识内容进行表示和发布,并对相关知识进行关联与整合,从而进行知识发现,向用户提供高层级的知识服务。

随着 e-Science 的兴起,数字图书馆不仅要作为信息库而存在,更重要的是要成为人类知识的巨大宝库和人类信息交互与共享的平台,能够为用户提供决策支持、专家咨询、智能信息检索、知识管理、信息

推荐等多种功能,语义网技术的出现和发展使建立更加智能的语义数字图书馆系统成为可能,在数字图书馆领域具有非常广阔的应用空间。

本书的全部内容由作者执笔完成,博士研究生周宇收集了相关文献资料,硕士研究生左丹进行了排版和校对工作,在此一并对两位同学致谢!

欧石燕

于南京大学仙林校区

2017年1月

目 录

第 1 章 绪论	1
1.1 语义网的诞生	1
1.2 语义网发展历程	2
1.3 语义网研究状况分析	4
1.3.1 国际研究状况	4
1.3.2 国内研究状况	17
1.4 本书组织结构	20
参考文献	21
第 2 章 语义网相关标准与技术	22
2.1 引言	22
2.2 标准化语义网技术	25
2.2.1 资源描述框架 RDF	25
2.2.2 资源描述框架模式 RDFS	29
2.2.3 本体描述语言 OWL	31
2.2.4 RDF 查询语言 SPARQL	35
2.2.5 规则交换格式 RIF	39
2.3 其他语义网技术	42
2.3.1 简单知识组织系统 SKOS	42
2.3.2 丰富网页摘要标记格式	46
2.3.3 GRDDL	52
2.4 关联数据	55
2.4.1 关联数据概述	55
2.4.2 关联数据中 Web 资源的访问	57

2.4.3 关联数据的发布	60
2.5 语义网的主要功能	62
2.6 语义网应用分析	66
2.6.1 语义网应用定义	66
2.6.2 语义网应用调研与分析	67
2.7 总结与结论	72
参考文献	73
第3章 语义网在数字图书馆中的应用	80
3.1 引言	80
3.2 书目语义模型与本体	81
3.2.1 来自图书馆界的模型与本体	81
3.2.2 来自工业界的模型与本体	95
3.2.3 语义出版与参考本体	100
3.3 图书馆关联数据的发布	111
3.4 基于语义网技术的图书馆信息资源整合	112
3.4.1 基于本体的信息资源整合	113
3.4.2 基于关联数据的信息资源整合	114
3.5 基于语义网技术的数字图书馆	116
3.5.1 语义数字图书馆项目	116
3.5.2 语义数字图书馆实例	117
3.6 总结与结论	126
参考文献	127
第4章 数据关联的语义数字图书馆构建	133
4.1 引言	133
4.2 语义数字图书馆资源描述与组织框架	134
4.3 文献资源的语义化描述	137
4.3.1 核心元数据本体的设计	138
4.3.2 核心元数据本体的扩展	144
4.3.3 语义书目元数据的应用	149
4.4 受控词表的语义化描述	150

4.4.1	相关研究	150
4.4.2	基于 SKOS 的词表语义化表示方案	152
4.4.3	词表语义化转换的实施	156
4.5	其他相关资源的语义化描述	163
4.5.1	个人和组织机构	163
4.5.2	地点	166
4.5.3	事件	168
4.5.4	时间	170
4.6	图书馆关联数据的构建与发布	172
4.6.1	图书馆关联数据的构建	172
4.6.2	图书馆关联数据的发布	174
4.6.3	图书馆关联数据的访问	179
4.7	实验测评	184
4.7.1	查询特定作者的文献	184
4.7.2	查询特定主题的文献	186
4.8	总结与结论	188
	参考文献	189
 第 5 章 面向图书馆关联数据的自动问答		193
5.1	引言	193
5.2	相关研究	194
5.3	系统概览	198
5.4	命名实体抽取与索引构建	199
5.4.1	命名实体抽取	200
5.4.2	索引构建	202
5.5	本体词汇抽取与索引构建	203
5.5.1	本体词汇抽取	205
5.5.2	同义词扩展	206
5.5.3	本体词汇索引构建	207
5.6	问句预处理	210
5.6.1	汉语分词与命名实体识别	210
5.6.2	时间日期标注	212

5.6.3	命名实体标注	213
5.6.4	本体词汇标注	214
5.7	问句分类	217
5.8	简单句的问答过程	221
5.8.1	本体兼容三元组的生成	221
5.8.2	简单句 SPARQL 查询的构建	222
5.9	复杂句的问答过程	227
5.9.1	复杂句的分解	228
5.9.2	复杂句 SPARQL 查询的构建	235
5.10	实验测评	238
5.11	总结与结论	241
	参考文献	242
第 6 章	术语注册与术语服务	245
6.1	引言	245
6.2	国外术语注册与术语服务	249
6.2.1	FAO 注册系统与术语服务	249
6.2.2	OCLC 术语服务	255
6.2.3	开放元数据注册	260
6.2.4	NKOS 注册	265
6.2.5	HILT 项目	267
6.2.6	STAR 项目	270
6.2.7	ADL 地名表与叙词表协议	271
6.2.8	PoolParty	273
6.2.9	分析比较	274
6.2.10	结论与建议	280
6.3	术语注册与术语服务系统架构设计	281
6.3.1	系统功能需求	281
6.3.2	系统设计原则	282
6.3.3	系统整体架构	283
6.3.4	任务服务层	285
6.3.5	工具服务层和组件层	289

6.3.6 数据层	293
6.4 术语注册与术语服务实例系统构建	295
6.4.1 实例系统架构	295
6.4.2 词表验证	297
6.4.3 术语存储与检索	305
6.4.4 基于 Jersey 框架的术语服务构建	310
6.4.5 基于 Clerezza 框架的术语服务构建	316
6.4.6 术语服务客户端构建	318
6.5 面向信息检索的术语服务应用研究	322
6.5.1 术语服务在信息检索中的应用	322
6.5.2 面向信息检索的术语服务有用性测评	324
6.5.3 面向信息检索的术语服务可用性测评	327
6.6 总结与结论	334
参考文献	337

第 1 章 绪 论

1.1 语义网的诞生

自 1991 年万维网(简称 Web)诞生以来,Web 已经发展成为一个拥有亿级页面的巨大分布式信息空间,为用户提供海量的信息服务。二十多年来,Web 虽然经历了一系列变化与发展,但无论是 1.0 时代的只读静态网页,还是 2.0 时代的交互式信息平台,基于超文本格式的非结构化文档之网(Web of Documents)的特性一直都没有改变,这使得当前 Web 还远远不能满足人们对信息共享和处理的需要,主要表现在^[1]:① 当前 Web 只能供人阅读和进行信息共享,计算机并不能“理解”Web 的内容,并在“理解”的前提下处理和利用这些信息;② 即使目前有大量网页的内容是由来自底层数据库的结构化数据自动生成,但是网页一经生成,信息反而失去了在数据库中的结构化特征,而这一特征对于机器理解和处理信息是非常有用的;③ 人们虽然能在 HTML 网页中建立超链接关系,但是无法在生成这些网页的底层数据间建立关联关系,导致 Web 底层的海量数据孤立而分散地存在着,无法进行集成和互操作,从而形成了一个信息孤岛。

鉴于 Web 页面的机器不可“理解”性,人类对当前 Web 的利用无法得到软件工具的很好支持。一直以来,伴随着 Web 诞生的搜索引擎是人们搜索和使用 Web 信息几乎唯一的、不可或缺的工具。但是以关键词匹配为核心的 Web 搜索引擎同样面临着一些严重问题,譬如:① 高匹配但低精度;② 低匹配或者无匹配;③ 查询结果对查询词高度敏感;④ 用户必须自行在检索结果中浏览和定位所需文档并从中抽取有用信息进行集成^[1]。虽然研究者们试图采用各种手段提高搜索引擎的智能性和精度,但是上述问题仍无法从根本上得以解决,这同样归因于当前 Web 非结构化、非语义化的信息表示形式,使得搜索引擎只能通过简单的关键词匹配而非语义匹配来搜索相关信息,并且关键词之间只能通过简单的布尔关系而非准确的概念关系来描述。

对于当前 Web 在信息表达、组织、检索中存在的严重缺陷与不足,有两种可

能的途径来解决上述难题:一种考虑是,采用人工智能和自然语言处理技术开发出更为复杂的程序来对网页上的非结构化信息进行处理,但是很遗憾,人工智能技术本身存在着很大局限性,再精妙的机器(计算机程序)也无法真正像人一样理解和思考,这一途径无疑遇到了无法突破的瓶颈;另一种考虑是,采用一种适于机器理解和访问的新方式来表达 Web 上的内容,从而方便机器的处理,这就是语义网想法的最初由来^[1]。1998 年,万维网的发明人伯纳斯·李在他的 Web 设计笔记里首次提出了对语义网的设想,即“一个在某种程度上类似于全局数据库的数据之网(Web of Data)”^[2]。2001 年 5 月,伯纳斯及其合作者在 *Scientific American* 杂志上发表了题为“The Semantic Web”的论文,系统论述了他对下一代万维网架构——语义网——的蓝图,这篇论文同时也被认为是语义网诞生的标志^[3]。

语义网的目标是通过给万维网上的文档添加能够被计算机所理解的语义(元数据,Metadata),让计算机能够“理解”分布在网上的信息和知识,并在“理解”的前提下更好地处理、利用这些信息和知识,从而使整个 Web 成为一个支持全球化知识共享的智能信息服务平台。由此看出,语义网相对于现有万维网的最大优势是“机器可理解”,它对 Web 进行扩展使得 Web 具有知识理解及一定的自动处理和推理能力,它的出现给 Web 带来了革命性的变化,使人和机器协同工作理解并处理 Web 上的信息成为可能。由于语义网的建立极多地涉及人工智能领域,与 Web 3.0 智能网络的理念不谋而合,因此语义网也被看作 Web 3.0 的重要特征之一。

1.2 语义网发展历程

在语义网出现的最初大约十年,研究者们主要致力于语义网技术本身的建设和完善。首先是制定和完善语义网的各项语言规范,如 RDF、RDFS、OWL、SPARQL 等,其次各种语义网实现工具也相继被开发出来,譬如,RDF 三元组存储器 3Store、Virtuoso 和 AllegroGraph,RDF 数据转换工具 RDFizers,语义网开发工具包 Jena,本体编辑器 Protégé 和 OntoEdit,本体推理机 Pellet、RacerPro 和 FaCT++,从而使语义网技术有了在实践中进行应用的可能。但是,相比传统 Web 的最初发展,语义网的发展速度还显得比较迟缓,对语义网的研究也主要集中在高校和研究机构,缺乏企业界的普遍参与,除了在少数专业领域(如生物医学),几乎没有对广大 Web 用户有价值的语义网数据集存在。原因主要在于:在语义网研究的早期,过分强调推理的必要性,大量依靠本体进行知识建模和语义标注,使得语义网的应用只能局限在特定领域的狭小范

围,而无法扩展至 Web 级的海量数据,因此严重制约了语义网在整个 Web 上的推广与应用。

2006 年,语义网的发明者伯纳斯·李进一步提出了关联数据的概念^[4]。关联数据是指在网络上发布、共享、连接各类数据、信息和知识的一种方式,是推荐的语义网最佳实践^[4]。关联数据从技术上来说虽然很简单,然而却使 Web 发生了深刻的改变,它摒弃了语义网早期过度依赖本体进行知识建模和数据集成的做法,强调 RDF 数据的互联与 Web 访问,促进了数据之网(Web of Data)的创建,为语义网的大规模应用奠定了基础。严格说来,数据之网并不能算是真正的语义网,因为它主要强调数据的结构化和关联,还远未达到伯纳斯所设想的语义与智能的程度,因此可将其看作语义网的一个子集或初级阶段。近年来,许多机构纷纷采用关联数据作为发布结构化数据的一种途径,从而构成了一个全球的数据空间。该数据空间的出现源自语义网研究社区的努力,特别是得益于万维网联盟(World Wide Web Consortium,简称 W3C)“Semantic Web Education & Outreach”工作组支持下的“关联开放数据(Linking Open Data,简称 LOD)”项目。截至 2014 年 8 月,在 LOD 云中关联的数据集已达到 1014 个,描述了约 800 万个资源^[5]。整个 LOD 云以 DBPedia^①为核心,囊括了地理、政府、媒体、生命科学、图书馆、用户生成内容等领域的数据以及一部分跨领域数据,其中图书馆及其相关领域(如教育、出版)的关联数据集有 96 个,约占整个 LOD 云的 9.47%^[5]。目前,LOD 云中的数据几乎以每 3 年 2 个数量级的速度在增长,以致到了几乎无法计量的地步。

在注重将现有结构化数据以关联数据方式在 Web 上进行发布的同时,如何实现海量传统网页向语义网的自然过渡更是值得考虑的问题。通过基于本体的语义标注将非结构化的 HTML 网页完全转换为结构化的 RDF 数据,意味着完全放弃传统 Web 及背后的成熟技术,这种做法不仅复杂,而且事实证明也不被广大 Web 用户所接受。将 RDF 数据以 RDF/XML 格式直接嵌入 XHTML 网页中也是不可行的,因为这是目的和结构完全不同的两种表示格式,互不兼容。W3C 和 Web 民间社区殊途同归地分别提出了三种功能相同的 Web 语义标注格式:来自 W3C 的 RDFa 格式^[6]、来自民间的微格式^[7]和来自 HTML5 规范的数据^[8]。这三种格式虽然具体规则和表示不同,但本质上都是通过通过在原有 HTML/XHTML 网页中嵌入语义标签,从而将传统网页在人类可读的基础上提高到机器可读、可理解的状态。2011 年 6 月,微软、谷歌和雅虎三大搜索引擎(后期包括俄罗斯搜索引擎 Yandex)联手发起了 Schema.org 项目^[9],以促使网

① DBpedia 是从 Wikipedia 词条中抽取结构化数据并将其以关联数据形式在网络上发布的项目,见 <http://wiki.dbpedia.org/Datasets>。

站管理员在网页里使用结构化数据标记来帮助搜索引擎更好地理解网页内容,那些符合 Schema.org 标注格式的网站,不仅能提高用户检索效率,还能增加网页被检索到的可能性,这一措施无疑进一步促进了语义网技术的普及与应用。

1.3 语义网研究状况分析

本节通过对有关语义网的国内外研究文献进行调研分析,得出语义网当前的研究状况。

1.3.1 国际研究状况

国外文献调研主要设定在 2008 年至 2013 年范围内,之所以以 2008 年为调研起点,是因为 2004 年 RDF、RDFS 和 OWL 等语义网基础技术规范发布,2008 年 RDF 数据存取技术规范 SPARQL 发布,自此之后,语义网技术基础基本奠定,逐步进入了应用阶段。调研文献的数据采集自 Thomson Reuters 公司出版的引文数据库 Web of Science,通过检索获取该数据库中 2008 年至 2013 年间发表的与语义网相关的文献。根据语义网研究所属的学科范围,将检索范围限定为 Web of Science 核心合集下的五个数据库,分别是:

- *Science Citation Index Expanded* (SCI-EXPANDED)
- *Social Sciences Citation Index* (SSCI)
- *Arts & Humanities Citation Index* (A&HCI)
- *Conference Proceedings Citation Index-Science* (CPCI-S)
- *Conference Proceedings Citation Index-Social Science & Humanities* (CPCI-SSH)

在构建检索式时,首先将“Semantic Web”和语义网最佳实践“Linked Data”作为主题词。但由于“Linked Data”在非语义网领域有另外的含义,在作为主题词检索时会检出许多与语义网不相关的文献,因此进一步限定在文献主题中至少含有 RDF、OWL、SKOS、SPARQL、semantic、logic、ontolog* (表示 ontology、ontologies 等)、reason* (表示 reasoning、reasoner 等)、inferenc* (表示 inference、inferencing 等)这些与语义网紧密相关的词汇之一,以提高文献的查准率。最终构建的检索式如下所示:

主题=(*Semantic Web* OR *Linked Data*)AND(RDF OR OWL OR SKOS OR
SPARQL OR *semantic* OR *logic* OR *ontolog* * OR *reason* * OR *inferenc* *)

利用上述检索式在 Web of Science 核心合集下的五个数据库中进行检索,共得到 6119 篇相关文献。这些文献包括会议论文(proceedings paper)、期刊论文(article)、评论(review)、编辑文献(editorial material)、书评(book review)、会议摘要(meeting abstract)等多种形式,其中会议论文(4056 篇,占 66.3%)和期刊论文(1971 篇,占 32.2%)是最主要的两种文献类型,共占文献总量的 98.5%。

(1) 基本统计分析

将检索到的 6119 条文献记录从 Web of Science 数据库中导出,首先对上述文献的年度分布、国家/地区分布、学科领域分布进行基本的统计分析,了解世界范围内语义网研究分布。

① 文献年度分布

自 Web of Science 数据库可直接得到每年出版的文献数量,将文献数量按照出版年份进行统计,得出图 1-1 所示的文献年度分布图。该图表明,2009 年语义网文献数量达到最多,2010 年之后进入平稳期。这说明,在 2008 年语义网基本技术规范完成后,语义网研究掀起了一个小小高潮。

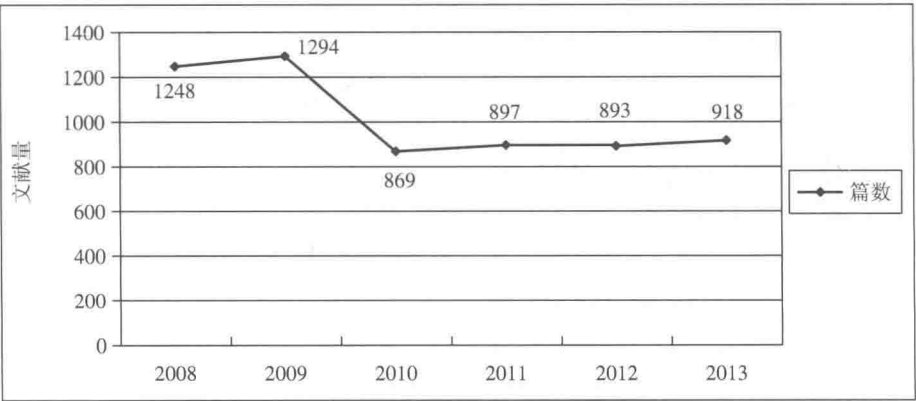


图 1-1 国际语义网文献的年度分布

② 文献国家/地区分布

对文献数量在国家/地区上的分布进行统计,能够反映出世界范围内语义网研究的热点区域。限于篇幅,这里仅列出发文量前十的国家/地区,如图 1-2 所示。该图表明,欧美国家是语义网研究的主力,在亚洲则主要是中国和韩国,从数量上来讲,我国的语义网研究已经在世界各国中居于领先地位。

将欧盟诸国的发文量综合起来与中、美两国进行对比,三者的发文走势基本类似,即 2009 年发文量达到一个顶峰,2010 年之后趋于平稳,如图 1-3 所示。观察每年的发文量,欧盟占绝对优势,超出中美之和,几乎占全球每年发文量的一半,这说明欧洲是世界上语义网研究最活跃的地区。