

# 元数据（META DATA）技术学术发展分析报告

Meta data technology academic development analysis report

方建勇<sup>1</sup>（余姚，浙江 315400）

**摘要：**元数据有很大机遇，设计一套元数据框架，基于大数据的存储形式，如果做到高效，安全，可维护是难点，关键现在正是 META DATA 发展的重要时期。元数据智能化，不知道算不算机器学习的一部分？通过超星发现系统，我们大致了解到元数据技术所涉及的相关领域，这些领域的研究课题属于哪些学科，哪些机构发表的学术成果较多，集中在哪些刊物发表等信息，还有指出了哪些元数据技术的相关学术成果被引用较多，为我们研究元数据技术这个课题做了比较好的指引。

**关键词：**元数据 知识管理 情报学 数据分析 分析报告

**Abstract:** Meta data has a great opportunity to design a set of metadata framework, based on large data storage form, if efficient, safe, maintainable is difficult, the key is now META DATA development of an important period. Metadata intelligent, do not know part of the machine learning? Through the superstar discovery system, we have a general understanding of the relevant areas of metadata technology, which disciplines of these disciplines, which institutions publish more academic outcomes, which publications are published and what information Data technology related academic results are cited more for our research metadata technology to do a better guide to this topic.

**Key words:** Meta data; Knowledge management; Information science; Data analysis; Analysis report

《中文元数据标准框架及其应用》，通过对北京大学数字图书馆中文元数据标准框架的主要内容及应用实例的介绍,阐述了中文元数据标准制定的原则、方法和工作流程。《国外常用元数据标准比较研究》，通过对目前国际上较流行的

---

<sup>1</sup> 方建勇，男，1978 年生，1998 年考入浙江大学数学系，主要从事信息技术领域，现为一家物流公司副总经理，曾陆续任职于某副部级央企省级分公司信息技术部，北京某 IT 咨询公司资深数据库顾问。美国电气电子工程师学会 IEEE 会员，美国计算机学会 ACM 会员，中国工业与应用数学学会会员，中国计算机学会会员，中国物流学会会员，浙江大学历史系研究生学历，浙江大学数学系本科毕业，理学学士学位。

七种元数据标准 CDWA、DC、EAD、FGDC、GILS、TEI、VRA 的介绍、比较和分析，总结出元数据标准在设计 and 实现过程中的几个关键问题。

《元数据管理策略的比较研究》，讨论了元数据及元数据管理的若干问题，并通过对几种元数据管理策略和元数据交换标准的比较，来分析元数据管理的发展方向，以及实现元数据管理的方式和具体步骤。

《商务智能系统中的元数据管理策略研究》一文指出，随着数据仓库的数据质量问题越来越引起重视，更进一步让人们认识到了元数据管理的重要性。元数据管理技术和元数据管理策略都是进行成功的元数据集成所必需要重视的研究内容。该文从元数据来源、元数据范围、元数据分类、元数据用户、元数据集成项目开发方法、元模型和元数据标准、元数据管理系统的实现途径等方面阐述了商务智能系统中的元数据管理策略。《电子文件元数据智能分析与可视化呈现》一文，提出将元数据智能分析的理念引入电子文件的开发利用工作中的创新思路，通过数据挖掘流程与技术创新的有机结合，自动化、智能化、高效率、低成本地将电子文件中的浅层信息变为深层信息，将隐性知识转化为显性知识，力求加速知识创新，为电子文件的深度开发与利用奠定基础。

本文谨对元数据（META DATA）技术的学术情况作一个基于大数据的分析，元数据大多应用在云计算、知识管理、情报学、数据标准化、数据分析、智能显示和可视化等领域，希望能对研究能有所帮助。

一、元数据（META DATA）技术学术发展趋势

超星发现系统收录的元数据（META DATA）技术历年发表的学术成果，见表 1，总量为 45219 条记录，包括中文 25989 条和外文 19230 条。其中包括图书 (287)、期刊 (15270)、报纸 (275)、学位论文(5803)、会议论文 (5249)、标准 (459)、专利 (13843)、音视频 (310)、科技成果 (834)、年鉴 (170)、法律法规 (54)、案例 (4)、信息资讯 (2641)、特色库 (20) 等。

表 1 元数据（META DATA）技术各类型学术发展趋势

“元数据” “meta data”-各类型学术发展趋势									
序号	年份	图书	期刊	学位论文	会议论	专利(数	标准	报纸	科技成果(数量)

		(数量)	(数量)	(数量)	文(数量)	量)	(数量)	(数量)	
1	1937	0	1	0	0	0	0	0	0
2	1938	0	0	0	0	0	0	0	0
3	1939	0	1	0	0	0	0	0	0
4	1940	0	0	0	0	0	0	0	0
5	1941	0	0	0	0	0	0	0	0
6	1942	0	0	0	0	0	0	0	0
7	1943	0	1	0	0	0	0	0	0
8	1944	0	0	0	0	0	0	0	0
9	1945	0	0	0	0	0	0	0	0
10	1946	0	0	0	0	0	0	0	0
11	1947	0	0	0	0	0	0	0	0
12	1948	0	0	0	0	0	0	0	0
13	1949	0	0	0	0	0	0	0	0
14	1950	0	0	0	0	0	0	0	0
15	1951	0	0	0	0	0	0	0	0
16	1952	0	0	0	0	0	0	0	0
17	1953	0	0	0	0	0	0	0	0
18	1954	0	0	0	0	0	0	0	0
19	1955	0	0	0	0	0	0	0	0
20	1956	0	0	0	0	0	0	0	0
21	1957	0	1	0	0	0	0	0	0
22	1958	0	0	0	0	0	0	0	0
23	1959	0	0	0	0	0	0	0	0
24	1960	0	1	0	0	0	0	0	0
25	1961	0	0	0	0	0	0	0	0
26	1962	0	0	0	0	0	0	0	0
27	1963	0	0	0	0	0	0	0	0
28	1964	0	0	0	0	0	0	0	0
29	1965	0	0	0	0	0	0	0	0
30	1966	0	0	0	0	0	0	0	0
31	1967	0	0	0	0	0	0	0	0
32	1968	0	2	0	0	0	0	0	0
33	1969	0	0	0	0	0	0	0	0
34	1970	0	1	0	0	0	0	0	0
35	1971	0	0	0	0	0	0	0	0
36	1972	0	0	0	0	0	0	0	0
37	1973	0	0	0	0	0	0	0	0

38	1974	0	0	0	0	0	0	0	0
39	1975	0	2	0	0	0	0	0	0
40	1976	0	1	0	0	0	0	0	0
41	1977	0	1	0	0	0	0	0	0
42	1978	0	6	0	1	0	0	0	0
43	1979	0	0	0	0	0	0	0	1
44	1980	0	2	0	1	0	0	0	0
45	1981	0	3	0	0	0	0	0	0
46	1982	0	2	0	0	0	0	0	1
47	1983	0	6	0	0	0	0	0	0
48	1984	0	14	0	1	2	0	0	0
49	1985	2	7	0	0	1	0	0	0
50	1986	0	13	0	1	0	1	0	0
51	1987	0	5	0	0	1	0	0	0
52	1988	0	6	0	1	1	0	0	1
53	1989	1	14	3	4	9	0	0	2
54	1990	1	26	0	4	2	0	0	1
55	1991	0	36	0	2	7	0	0	1
56	1992	0	34	2	6	5	0	0	1
57	1993	0	28	1	8	6	0	0	4
58	1994	0	46	0	8	28	1	0	0
59	1995	0	52	2	19	8	1	0	0
60	1996	0	73	3	29	25	0	0	3
61	1997	0	100	5	30	22	1	0	1
62	1998	0	142	18	56	71	1	0	3
63	1999	2	143	20	81	102	11	0	2
64	2000	17	216	48	111	326	10	2	5
65	2001	4	350	70	194	433	2	8	12
66	2002	11	554	101	266	437	9	8	15
67	2003	3	687	190	420	559	12	11	14
68	2004	14	805	242	683	592	9	25	16
69	2005	5	862	359	425	846	32	19	35
70	2006	11	912	491	359	860	19	34	33
71	2007	9	887	501	330	869	5	23	39
72	2008	16	872	456	319	797	11	1	42
73	2009	24	956	484	248	721	24	7	39
74	2010	20	922	412	248	924	17	10	33
75	2011	9	802	479	207	1080	12	6	30
76	2012	10	904	424	159	1213	41	5	13

77	2013	21	806	441	112	1202	22	72	5
78	2014	26	975	350	135	1018	36	16	8
79	2015	22	901	303	99	882	9	14	0
80	2016	7	923	186	38	602	30	12	1
81	2017	9	1077	139	28	451	16	9	0

## 二、元数据（META DATA）技术学术成果统计<sup>2</sup>

### 1、关键词

关键词涉及元数据(8170)、数字图书馆(1101)、XML(1015)、数据仓库(839)、本体(479)、数据管理(468)、数据库(441)、信息资源(435)、电子文件(314)、数据共享(306)、数据挖掘(289)、设计实现(278)、数据集成(275)、Marc(267)、管理系统(259)、网络信息资源(256)、互操作(249)、系统设计(245)、数据模型(242)、信息组织(234)、地理信息系统(224)、RDF(223)、数据标准(223)、internet(221)、信息共享(209)、图书馆(200)、GIs(197)、电子政务(194)、搜索引擎(190)、data mining(188)、WEB 服务(182)、信息检索(179)、信息系统(177)、ETL(173)、技术研究(168)、资源整合(165)、空间数据(165)、数字资源(164)、文件系统(162)、D C (160)、标准(156)、网格(154)、网络信息(148)、标准化(147)、数据交换(141)、语义网(140)、分布式文件系统(137)、体系结构(134)、应用研究(132)等。

---

<sup>2</sup> 数据来源于超星发现系统。

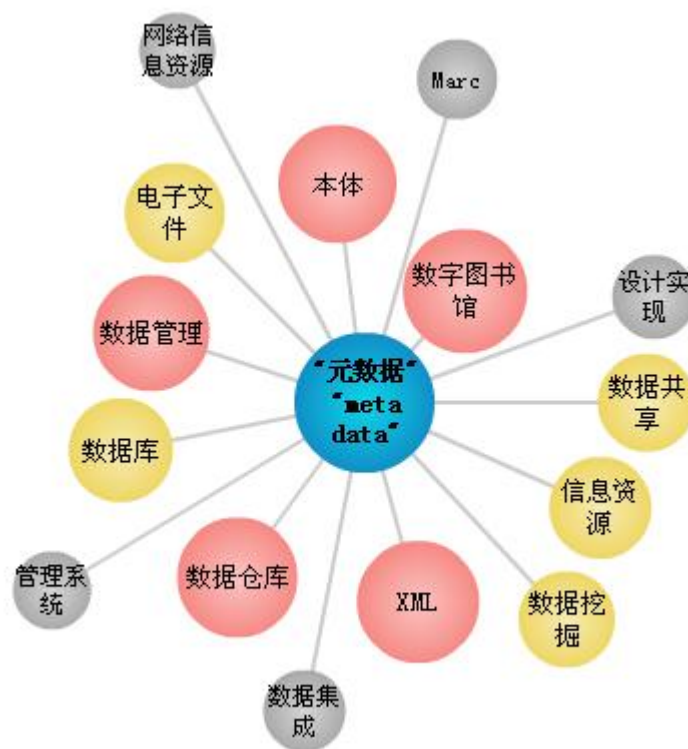


图 1 元数据 (META DATA) 技术关键词频次泡型图

## 2、机构、刊种与地区分布

发表机构分布依次为中国科学院(772)、武汉大学(543)、华中科技大学(514)、北京大学(332)、微软公司(283)、北京航空航天大学(282)、清华大学(278)、北京邮电大学(273)、上海交通大学(252)、国防科学技术大学(239)、电子科技大学(227)、浙江大学(215)、华为技术有限公司(210)、吉林大学(173)、华东师范大学(165)、中国人民大学(161)、南京大学(160)、复旦大学(149)、中山大学(136)、华南理工大学(135)、大连理工大学(127)、西安电子科技大学(122)、山东大学(121)、南京政治学院(120)、中南大学(112)、中兴通讯股份有限公司(111)、重庆大学(110)、东北大学(109)、同济大学(109)、北京师范大学(107)、北京理工大学(105)、东南大学(104)、西安交通大学(104)、北京工业大学(96)、哈尔滨工业大学(94)、中国地质大学(武汉)(93)、中国科学院大学(93)、四川大学(91)、天津大学(85)、山东科技大学(85)、华中师范大学(82)、成都理工大学(79)、中国科学技术信息研究所(77)、西北工业大学(74)、北京交通大学(73)、武汉理工大学(71)、湖南大学(71)、南京师范大学(70)、郑州大学(70) 等。

发表的刊物分布依次为现代图书情报技术(225)、DM Review(201)、图书情报工作(180)、计算机工程(174)、现代情报(143)、情报杂志(110)、计算机工程与应用(109)、计算机工程与设计(108)、数字图书馆论坛(104)、图书馆杂志(101)、情报科学(99)、图书馆学研究(93)、计算机科学(87)、计算机应用研究(86)、计算机研究与发展(84)、计算机应用与软件(73)、Journal of Clinical epidemiology(67)、情报理论与实践(65)、Statistics in medicine(64)、科技情报开发与经济(64)、计算机应用(63)、微计算机信息(62)、中国图书馆学报(61)、小型微型计算机系统(60)、情报学报(59)、计算机技术与发展(58)、情报探索(57)、测绘科学(56)、大学图书馆学报(54)、计算机工程与科学(52)、电脑知识与技术(52)、地球信息科学学报(52)、复印报刊资料(图书馆学、信息学、资料工作)(51)、图书馆理论与实践(51)、图书馆学刊(51)、测绘与空间地理信息(50)、图书馆论坛(45)、The Journal of the American Medical Association(43)、计算机系统应用(42)、图书情报工作动态(42)、计算机与现代化(42)、农业图书情报学刊(41)、InfoWorld(40)、图书馆建设(40)、浙江档案(40)、新世纪图书馆(39)、档案学研究(38)、测绘通报(38)、现代电视技术(37)等。

发表机构所属的地区分布依次是北京市(2451)、湖北省(1392)、江苏省(973)、上海市(954)、广东省(896)、四川省(570)、陕西省(511)、山东省(465)、辽宁省(444)、浙江省(404)、湖南省(329)、河南省(280)、黑龙江省(270)、吉林省(268)、天津市(236)、重庆市(191)、安徽省(187)、河北省(171)、福建省(170)、云南省(128)、山西省(112)、江西省(104)、甘肃省(81)、广西壮族自治区(72)、内蒙古自治区(59)、贵州省(36)、新疆维吾尔自治区(23)、海南省(16)、宁夏回族自治区(16)、青海省(4)、西藏自治区(1) 等。

### **3、元数据（META DATA）技术专项研究学位论文**

超星发现系统收录的元数据（META DATA）技术学位论文 5,803 篇，总被引 11611 次，可以说是代表了当前元数据（META DATA）技术水平最高的群体，元数据（META DATA）技术高引前 30 篇学位论文见表 2。

作者	学术成果	发表年份	学位	所在机构	总被引频次
崔巍	用本体实现地理信息系统语义集成和互操作	2004	博士	武汉大学	103
纪俊	一种基于云计算的数据挖掘平台架构设计与实现	2009	硕士	青岛大学	88
王春	无线传感器网络路由协议的设计与仿真	2004	硕士	电子科技大学	81
张校乾	基于 Lucene 的全文检索系统的研究与应用	2005	硕士	大连理工大学	78
何婷婷	语料库研究	2003	博士	华中师范大学	67
姚鹤岭	基于 REST 的 GIS Web 服务研究	2005	博士	中国人民解放军 信息工程大学	59
沈利峰	数字图书馆网络信息资源整合研究	2005	硕士	中国农业科学院	54
江绵康	“数字城市”的理论与实践	2006	博士	华东师范大学	51
陈永华	WebGIS 三维可视化的研究	2000	硕士	解放军信息工程 大学	50
孙坦	数字图书馆理论与发展模式研究		博士	中国科学院文献 情报中心	50
李冰	云计算环境下动态资源管理关键技术研究	2012	博士	北京邮电大学	46
刘立龙	动态对动态 GPS 高精度定位理论及其应用研究	2005	博士	武汉大学	46
曹风兵	基于 Hadoop 的云计算模型研究	2011	硕士	重庆大学	43



	与应用				
李宽	基于 HDFS 的分布式 Namenode 节点模型的研究	2011	硕士	华南理工大学	39
彭兵	基于学习对象的教学设计模型研究	2003	博士	华东师范大学	37
辜寄蓉	基于元数据的综合数据管理与信息共享：“数字成都”综合信息资源管理系统平台框架设计	2003	博士	成都理工大学	36
于雷易	基于空间数据网格的空间数据共享技术研究	2004	硕士	武汉大学	36
童艳	基于 ArcGIS 的基础地理空间数据管理系统设计与实现	2006	硕士	浙江大学	35
孔兰菊	SaaS 应用交付平台中多租房云数据管理关键技术研究	2011	博士	山东大学	34
黄俊华	面向服务的数字城市共享平台框架	2008	博士	武汉大学	34
黄东	基于 SQLite 的移动嵌入式数据库同步系统的研究和开发	2009	硕士	华中师范大学	32
龚卫华	数据库集群系统的关键技术研究	2006	博士	华中科技大学	32
翟巍	三维 GIS 中大规模场景数据获取、组织及调度方法的研究与实现	2003	博士	大连理工大学	32
陈建宏	可视化集成采矿 CAD 系统研究	2002	博士	中南大学	32
娄秀明	用关联数据技术实现网络知识组织系统的研究	2010	硕士	华东师范大学	31
吴永和	学习资源服务生态环境构建的研究	2009	博士	华东师范大学	31
曹蓟光	联机分析挖掘处理技术（OLAM）的研究	2001	博士	浙江大学	31

陈虎	基于 HDFS 的云存储平台的优化与实现	2012	硕士	华南理工大学	30
郝兴伟	基于知识本体的 E-learning 系统研究	2007	硕士	山东大学	30
周明建	基于本体的开放式知识管理研究	2004	博士	浙江大学	29

表 2 元数据（META DATA）技术高引前 30 篇学位论文列表

### 三、结语

通过超星发现系统，我们大致了解到元数据（META DATA）技术所涉及的相关领域，元数据大多应用在云计算、知识管理、情报学、数据标准化、数据分析、智能显示和可视化等领域，这些领域的研究课题属于哪些学科，哪些机构发表的学术成果较多，集中在哪些刊物发表等信息，还有指出了哪些元数据（META DATA）技术的相关学术成果被引用较多，为我们研究元数据（META DATA）技术这个课题做了比较好的指引。

### 四、参考文献

- [1] 超星发现系统[EB/OL].<http://www.chaoxing.com/>
- [2] 肖珑，陈凌，冯项云，冯英（北京大学数字图书馆研究所），中文元数据标准框架及其应用[J]，《大学图书馆学报》，2001 第 19 卷 第 5 期 P29-35，91。
- [3] 冯项云，肖珑，廖三三，庄纪林（北京大学数字图书馆研究所），国外常用元数据标准比较研究[J]，《大学图书馆学报》，2001 第 19 卷 第 4 期 P15-21。
- [4] 曹蓟光，王申康（浙江大学人工智能研究所），元数据管理策略的比较研究[J]，《计算机应用》，2001 第 21 卷 第 2 期 P3-6。
- [5] 于千城（北方民族大学计算机学院），商务智能系统中的元数据管理策略研究[J]，《电脑知识与技术》，2008 第 28 期 P178-180。
- [6] 姜伟，王健，班翊坤，马春茂，冯晓佳（北京市人民政府外事办公室信息中心；中国人民大学信息资源管理学院；北京市档案局），元数据管理策略的比较研究[J]，《北京档案》，2015 第 7 期 P24-26。