

● 周庆山, 赵雪, 赵需要, 周格非 (北京大学 信息管理系, 北京 100871)

我国数字内容产业研究的主题知识图谱分析*

摘要: 借助现代文献管理工具与软件, 对 2000—2010 年间我国数字内容产业方面已发表的相关文献及关键词进行搜索、整理、调研和分析, 并进行相关的词频、聚类分析。

关键词: 数字内容产业; 关键词; 知识图谱; 主题分析

Abstract: By means of modern document management tool and software, this article collects, arranges, investigates and analyzes the published literatures and relevant keywords in digital content industry in China from 2000 to 2010, and conducts relevant word frequency and clustering analysis.

Keywords: digital content industry; keywords; knowledge mapping; subject analysis

数字内容产业是以数字技术为依托、多个产业融合形成的新兴产业, 是社会信息化和经济全球化发展进程中文化产业变革的必然趋势。数字内容核心产业包括电影、电视、广播、音乐、动画、游戏、工艺、时尚设计、广告、建筑、视觉艺术、互动休闲软件、表演艺术、出版、软件、信息服务等。在我国 2006 年 3 月 16 日发布的《国民经济和社会发展第十一个五年规划纲要》中, 数字内容产业在国家文件中正式出现, 此后“数字内容产业”频繁出现在国家的各类发展规划中。在 2011 年出台的《国民经济和社会发展第十二个五年规划纲要》中, “数字内容产业”仍然是其中的重要内容。而在国外, 1995 年, 西方七国信息会议首次提出“内容产业”(Content Industry)概念, 作为产业统计的一个正式门类则于 1997 年首次出现在《北美产业分类系统》(NAICS)内。1999 年, 欧盟的“Info2000 计划”才真正对内容产业进行了界定: 那些制造、开发、包装、销售信息产品和服务的企业^[1]。

在学术研究领域, 2000 年, 《当代韩国》(2000 年夏季号)上发布一则消息, “韩国将制定新法以促进数字内容产业发展”, 此为国内较早明确提及数字内容产业^[2]。随后, 国内学者逐渐开始对数字内容产业的概念、内涵、分类等进行全方位地研究。鉴于相关研究基本始于 2000 年, 因此笔者对 2000—2010 年间我国数字内容产业已发表的相关论文及关键词进行统计分析, 并对我国数字内容产业学术研究的主题知识图谱予以绘制。

* 本文为国家社会科学基金重大项目“我国文化产业发展战略研究”子课题“数字内容产业发展战略研究”的阶段成果, 项目编号: 10zd&021。

1 材料来源与研究方法

1.1 材料来源

笔者用“数字* 内容”、“数字* 产业”、“内容* 产业”3 个关键词经布尔逻辑“与”组合后, 通过中国知网中的中国学术期刊网络出版总库、中国博士学位论文全文数据库、中国优秀硕士学位论文全文数据库、中国优秀硕士学位论文全文数据库(增刊)4 个库进行跨库题名(精确)检索, 年代为 2000—2010 年间共检索到文献 1 660 篇, 筛除重复文献 162 篇、无关文献 245 篇, 实际纳入统计的文献 1 253 篇。见表 1。

表 1 2000—2010 年间数字内容产业相关文献统计分布

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	小计
初始文献	27	52	52	79	121	188	199	228	175	244	295	1660
重复文献	2	3	1	2	5	21	26	11	18	30	43	162
不相关	6	15	12	19	11	37	34	42	10	33	26	245
纳入统计	19	34	39	58	105	130	139	175	147	181	226	1253

从检索情况上看, 2000 年至 2010 年间, 相关文献数量呈现快速增长趋势(见图 1)。这反映出进入 21 世纪以

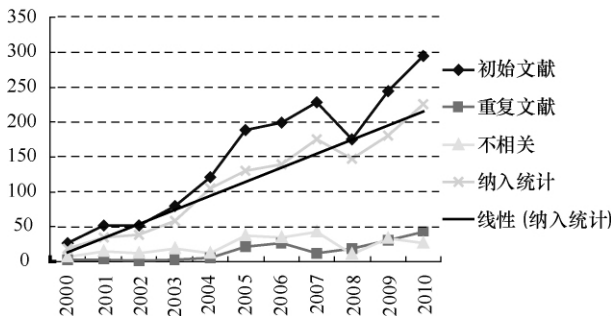


图 1 2000—2010 年间数字内容产业相关文献增长趋势

后,数字内容产业研究在国内学术界方兴未艾,正得到日益广泛的认可和关注。同时,根据图1中的文献增长曲线,预计未来文献数量还将进一步增长。

1.2 研究工具与方法

本文采用的研究方法主要包括文献计量法、数学统计法、共词分析法、聚类分析法和多维尺度分析法。Pajek是大型复杂网络分析工具,也是用于研究目前所存在的各种复杂非线性网络的有力工具。共词分析是统计两个关键词在同一篇文献中出现的次数,采用自编软件的方式对文献数据进行预处理与词频统计。同时,还借助 NoteExpress, Microsoft Excel 2003 和 SPSS19.0 进行数据统计和相关分析。

表2 2000—2010年数字内容产业领域高频关键词列表

序号	关键词	频次	序号	关键词	频次
1	产业发展	124	26	产业联盟	22
2	产业链	122	27	解决方案	22
3	数字电视产业	120	28	数字出版	22
4	数字电视	85	29	有线电视	22
5	出版产业	59	30	行业协会	21
6	数字化	54	31	三网融合	20
7	内容产业	51	32	数字家庭	20
8	数字内容	51	33	文化产业	20
9	机顶盒	46	34	运营商	19
10	数字	44	35	传统出版业	18
11	有线数字电视	42	36	数字机顶盒	18
12	数字技术	40	37	有线电视网络	18
13	信息产业部	37	38	数字图书馆	17
14	数字内容产业	34	39	产业集群	15
15	中国数字	34	40	出版业	15
16	产业化	30	41	传统出版	15
17	数字娱乐	29	42	论坛	15
18	产业	28	43	媒体产业	15
19	信息产业	28	44	数字娱乐产业	15
20	地面数字电视	27	45	文化创意	15
21	网络游戏	27	46	新闻出版	15
22	互联网	24	47	信息技术	15
23	数字媒体	24	48	增值业务	15
24	数字时代	24	49	中国电子	15
25	产业基地	23	50	北京市	13

2 研究及结果

研究主要通过4个步骤予以实现:①确定数字内容产业领域的高频关键词。②建立共词矩阵、相似矩阵和相异矩阵。③基于共词矩阵选取聚类分析、多维尺度分析绘制数字内容产业领域知识图谱。④对所得结果予以深度分析。

2.1 关键词词频分析

通过人工及文献管理工具 NoteExpress 提取所有数字内容产业相关文献中的关键词,并按关键词出现的频次进行汇总、统计和排序。在纳入统计的1253篇文献中,共包含关键词9699个,出现10次及以上的有115个;出现20次及以上的有33个,为了计算方便,抽取前50个词汇为高频关键词,见表2。

2.2 建立高频关键词共词矩阵、相似矩阵、相异矩阵

对全体文献进行关键词的统计,通过编程统计它们在论文中同时出现的频率,将总频次前50位关键词的共现关系总结成为一个50×50的关键词共词矩阵(见表3)。

表3 表征数字内容产业研究方向的
关键词共词矩阵(部分)

	产业发展	产业链	数字电视产业	数字电视	出版产业	数字化	内容产业
产业发展	124	26	20	12	22	11	12
产业链	26	122	29	11	13	10	6
数字电视产业	20	29	120	1	0	4	2
数字电视	12	11	1	85	1	3	7
出版产业	22	13	0	1	59	16	1
数字化	11	10	4	3	16	54	8
内容产业	12	6	2	7	1	8	51

共词矩阵是一个相关矩阵,对角线上的数据为该词出现的频次,如关键词“产业发展”共出现了124次,它与“产业链”同时在26篇论文中出现,也就是说,有26篇论文的关键词中同时含有上述两个关键词。

由于将要运用的多元统计方法对矩阵的数据结构有不同的要求,为了统计分析方便,笔者将共词矩阵转化为相似和相异矩阵,以适应不同的多元统计方法对数据的要求,因此用 Ochiai 系数^[3]将共词矩阵转换成相似矩阵(见表4)。

相似矩阵中的数字为相似数据,数字的大小表明相应两个关键词之间的距离远近,由于相似矩阵中,0值过多,统计时容易造成误差过大,为了进一步进行系统聚类与多维尺度分析,用1与全部相似矩阵上的数据相减,得到表示两词间相异程度的相异矩阵(见表5)。

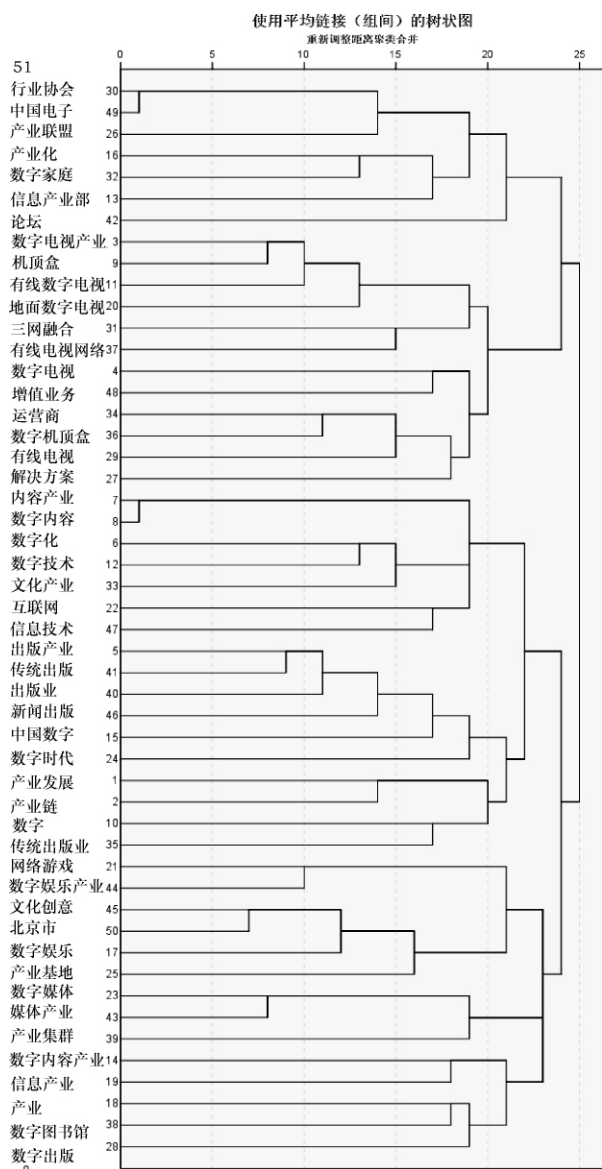


图3 2000—2010 数字内容产业研究高频关键词聚类象之间的相似性，有高度相似性的对象聚集在一起，形成一个类别^[4]。本文的做法是通过 SPSS19.0 统计软件 Scale 功能中的多维尺度分析功能（ALSCAL），选取平面对称的图形描述数字内容产业关键词的数据结构，用序数数值作为数据测度水平的指标，对数字内容产业关键词的相异矩阵进行二维尺度分析，见图4。

2.5.1 主题域1——参与机构研究 在主题域1中，信息产业部、产业联盟、行业协会、中国电子、论坛等高频关键词包含在该子领域当中，共有5个关键词，说明该子领域的研究主要围绕数字内容产业相关参与机构进行。

其中，信息产业部（现已变更为“工业和信息化部”）作为政府监管部门，肩负着维护行业秩序，制定相关政策的重任，先后出台了电子行业标准^[5]、数字群体体制标准^[6]、通信行业标准^[7]、数字音频行业标准^[8]等一

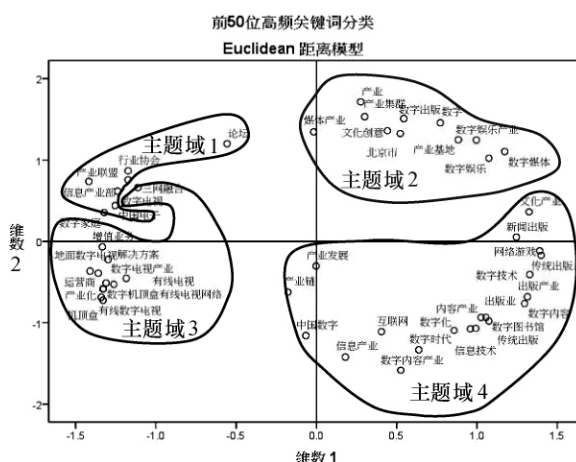


图4 2000—2010 年数字内容产业学术研究知识图谱

系列标准，为数字内容产业的长远发展奠定基础。“行业协会”作为一种民间性组织，它是政府与企业的桥梁和纽带，如中国电子视像行业协会^[9]，中国电子商会（CECC）^[10]等，并且这类行业协会常冠以“中国电子”的前缀，因此导致后者也成为热门关键词。产业联盟（Industry Alliance）是企业间结成的互相协作和资源整合的一种合作模式。数字内容产业作为一个新兴产业，要快速和良性地发展必须遵照一个科学的发展模式，而以产业联盟为基础的发展模式^[11]受到较多关注。与此同时，作为思想交流和理论宣导的平台，各种数字内容产业“论坛”相继出现，如数字技术应用高峰论坛^[12]，数字电视产业高峰论坛^[13]，数字知识产权论坛^[14]等。

2.5.2 主题域2——区域内容产业实践研究 在主题域2中，产业集群、数字出版、文化创意、产业基地、数字娱乐、数字媒体、北京市等高频关键词包含在这个领域中，共有11个关键词，从这些关键词可以看出，该子领域的研究主要围绕区域内容产业实践主题。随着数字技术的不断升级与发展，数字内容与创新文化、创智生活的结合愈加紧密，再加上国家宏观政策的导向和支持，数字创业实践得到了愈加快速的发展，所以从产业基地、文化创意、文化产业、产业集群等关键词可以充分表现数字创业方面的学术研究日趋得到重视。在数字内容产业研究的关键词中，北京、上海、深圳等作为地区性名词屡次被提出来，其中“北京市”更是成为了高频词，表明北京在区域产业实践中的特殊地位和优势，其中最突出的就是大众媒体产业特别是数字出版和数字娱乐产业成为国家重要的区域数字内容产业集聚区，这些区域依托北京雄厚的人才、文化及中关村技术优势，在引领全国现代产业发展方面起到很好的先导作用。据统计，2006年北京市互联网科学技术出版市场规模大约为3.4亿元，占全国90%以上的市场份额。北京市也拥有众多的数字内容产业集聚区，这些也

为研究产业集群和产业基地提供了实践基础。

2.5.3 主题域3——数字电视研究 在主题域3中,数字电视产业、数字电视、机顶盒、有线数字电视、产业化、地面数字电视、解决方案、有线电视、数字家庭、三网融合等高频关键词包含在该领域中,共有11个关键词,从该子领域包含的高频关键词中可以看出,研究主要围绕数字电视及其产业进行。各高频关键词聚集现象明显,说明这个领域中研究主题关系密切,是我国数字内容产业研究中相对成熟且重要的研究领域。

其中,数字电视产业、数字电视、增值业务、产业化是研究的一个核心主题。2001年,我国政府发布《广播影视技术科技十五计划和2010年远景规划》,对数字电视发展作了明确规划:2010年全面实现数字广播电视;2015年停止模拟广播电视的播出^[15]。目前数字电视已成为我国普及程度最高的数字内容产业形式,国内相关的研究包括数字内容产业的特征与优势分析、产业模式研究、运营商产业角色研究、政府规制分析等^[16]。数字机顶盒是该子领域中另一个重要的研究主题。国内学术界对数字机顶盒的研究分为两个方面,一方面,众多通信、计算机领域的学者从技术角度研究数字机顶盒的研发、制造等;另一方面,主要研究媒体、数字产业的专家学者则从数字机顶盒的应用模式、产业形态、市场前景等方向研究,着重突出了数字机顶盒在整个数字电视产业中相对独立存在这一特征,主要应用产业经济学、传播学等理论与方法展开研究^[17]。相关的研究主题还包括数字机顶盒解决方案、应用模式、市场等。其中还包括有线电视网络、地面数字电视等方面的研究,这个主题与一般有关媒体的学术研究最为接近。国内有许多学者运用传播学理论及方法重点研究数字电视的传播学特征、媒介性质、在大众传播中的应用等。研究热点包括数字电视网络的传播学性质分析、数字电视网络的建设与应用等,着重强调数字电视网络的具体应用及出现的问题。

随着IPTV试点推广力度的加强、数字电视整体平移提速、宽带日益普及、无线技术等快速发展,特别是互联互通的电子产品进入消费领域,数字家庭的理念被越来越多的消费者接受。数字家庭离不开三张网——通信网、电视网和互联网,三网融合则是完整数字家庭的基础^[18],能为用户提供语音,数据和广播电视等多种服务。

2.5.4 主题域4——产业核心问题研究 在主题域4中,主要关键词大约有20个左右,基本围绕产业为核心展开的,主要包括数字技术(数字化、数字时代、信息技术、数字图书馆、互联网、中国数字等)、数字出版(新闻出版、出版业、出版产业、传统出版、传统出版业等)、文化产业(文化创意、内容产业、数字内容产业、信息产

业、产业发展、产业链)等。

首先,和传统的内容产业相比,现代意义上的内容产业一个突出的特征就是数字化,因此,在这个领域中,围绕数字技术、信息技术和数字时代等关键词成为研究频次较高的内容,并成为分析数字内容产业的基本限定语,同时,相关主题的密集度反映了这一产业的技术与内容双重属性。新兴的数字技术是文化产业繁荣发展的物质支撑,构成了文化产业核心竞争力的基础。打造以数字技术为核心的文化产品是文化产业迈向高端产业的必然趋势。

其次,从这些关键主题词中可以发现,数字内容产业是一个大的产业集群,这其中除了主题域3中数字电视等媒体的数字内容建设外,数字出版及相关主题词也较为集中和突出,这反映了新闻出版领域面对网络数字媒体激烈竞争及融合转化背景下的数字出版产业成为数字内容产业的一个重要领域,新闻出版产业不仅包括图书、报纸、期刊等纸介质媒体产业,也包括音像、电子、动漫、游戏以及数字出版等非纸介质战略性新兴产业,印刷、复制产业,新闻出版流通、物流产业,新闻出版产品对外贸易、版权输出、合作出版等领域。据统计,2009年,我国数字出版总产出已达到799.4亿元,总体经济规模超过图书出版,初步形成了北京、上海、广东等数字出版产业集聚区,在数字出版内部,手机出版的营业收入已超过传统的网络游戏,占数字出版全部营业收入的24.2%,位居首位。这表明像数字出版产业这一大众传媒产业作为内容的核心来源,越来越具有重要的地位,

第三,在数字内容产业发展过程中,从注重信息产业的全面发展到信息资源如数字图书馆产业的开发利用。进入21世纪,文化内容与创意内容成为研究主题新热点,反映了内容产业发展的一个特有轨迹与趋势,其中起决定性的三要素是数字技术、数字设备和数字内容,在内容产业发展中,原创内容是核心,信息内容向文化内容的转化反映了文化产业借助信息技术的媒介和平台迎来了一个真正的内容为王的新世纪,原创作为数字内容产品的核心价值,创造力是将内容和信息区分的重要标准,也是数字内容产品实现商业价值的关键。

数字内容产业是文化产业中内容创造和生产环节规模化后形成的新产业,其发展面临的瓶颈之一是缺乏数字内容产业发展的良性机制,产业结构发展不均衡,产业链也没有完全形成,未来数字内容产业应更强调内容,内容部分从媒介中独立出来,形成大规模的内容产品生产,并且通过更为广泛的传输渠道和信息终端,进行更大规模的交易,数字化的大众传媒内容产品将会融入数字内容产业。因此,在大量研究文献中,集中探讨了如何发展数字内容产业,如何构建成熟有效的产业链问题。

很多关键词涉及文化产业、文化创意、内容产业、数字内容产业、信息产业等主题词,在文献中也有很多学者分析这些概念的含义与相互关系,反映目前数字内容产业战略周期尚处于从概念向产业转化的萌芽时期,尽管概念表述有些差异,如在澳大利亚就将其称为“创意性内容产业”(Creative Content Industry),也有学者认为“信息内容”概念提法不妥,很容易让人理解成广义上的信息资源及信息内容产业^[19]。无论如何,在实质上都认同其为以内容为核心,信息或数字为纽带,强调内容与其他行业的交叉和融合,是一个新兴的文化产业领域,从文献不断增加的趋势来看,它是一个文化产业未来5~10年发展战略中应引起高度重视的战略产业。

3 结束语

此研究成果鸟瞰了我国数字内容产业学术研究的嬗变轨迹和研究范畴,勾勒出学者们10年来相关研究的重点和热点,用知识图谱法较为形象地揭示了数字内容产业的整体框架和内部结构,并对未来整个文化产业的发展趋向起着重要的导向作用,结果表明,我国数字内容产业日渐成为学术界最为关注的课题。

此研究成果应用共词分析法,借助聚类分析和多维尺度分析勾画了我国数字内容产业学术研究的主题知识图谱,但目前只是一次初步的尝试性分析,其结果也难免有所不足。比如中国知网(CNKI)对文献的标注上仍存在着一定的主观性,有些期刊上发表的会议信息,本无关键词,但也被人工添加标注了关键词,这些难免会对研究结果产生一定的影响。比如《首届“中国北京数字内容创新发展高峰论坛”成功举行》一文^[20],被人工添加了10个关键词,分别为数字媒体、创新发展、北京市、内容产业、科学技术委员会、技术产业化、信息服务业、数字内容、产业发展、论坛。同时,对图4所示的知识图谱主题域的范畴确定也有一定的主观性,并且也存在一些与研究常理出现偏差的节点,如主题域2中“产业”和“产业集群”本应该和主题域4中的“产业链”关系更为紧密,但图谱上却相离甚远,所以此研究只是从较为宏观的层面,较为详细直观地反映出目前的学术研究主题知识图谱,但仍存在需要逐步改进和完善的地方。另外,本文只统计了前50个高频关键词,还不能全面反映关键词的一个全面系统的图谱。鉴于还有其他和数字内容产业类似或相近的概念,如信息内容产业、信息文化产业、数字文化产业、数字娱乐业、新媒体产业、网络文化产业,等等,更有很多涉及数字内容产业的具体产业内容,如文化产业新业态、手机内容、动漫、网络社区SNS等主题词等,这些都需要在今后的研究中进一步加以深入分析。□

参考文献

- [1] 贺德方. 中外信息内容产业的对比分析 [J]. 中国软科学, 2005 (11).
- [2] 新馨. 韩国将制定新法以促进数字内容产业发展 [J]. 当代韩国, 2000 (2): 17.
- [3] 张勤, 马费成. 国外知识管理研究范式——以共词分析方法 [J]. 管理科学学报, 2007 (12): 66-74.
- [4] 卢纹岱. SPSSforWindows [M]. 3版. 北京: 电子工业出版社, 2006.
- [5] 信息产业部批准发布数字电视相关25项电子行业标准 [J]. 安全与电磁兼容, 2006 (2): 9.
- [6] 信息产业部发布数字集群体制标准 [J]. 无线电工程, 2001 (3): 63.
- [7] 信息产业部发布“800MHz CDMA 数字蜂窝移动通信网设备总测试规范(移动台部分)”等四项通信行业标准 [J]. 通信世界, 2000 (S1): 6.
- [8] 国家数字音频行业标准颁布大力推进数字音视频产业发展 [J]. 数码世界, 2007 (2): 3.
- [9] 《中国数字电视》编辑部. 2009中国数字电视年度盛典(第四届)中国数字电视终端产业大检阅 [J]. 中国数字电视, 2009 (4): 76-77.
- [10] 翟文. 2009年中国数字电视产业发展高峰论坛 [J]. 记录媒体技术, 2009 (3): 62-64.
- [11] 李焱. 海淀园成立中关村数字电视增值业务产业联盟 [J]. 投资北京, 2008 (2): 76.
- [12] 中国标签市场发展令人瞩目——第二届国际标签产业及数字技术应用高峰论坛取得圆满成功 [J]. 印刷工业, 2010 (4): 61-62.
- [13] 晓波. 数字奥运, 数字电视新商机——第四届中国数字电视产业高峰论坛8月8日在京召开 [J]. 卫星电视与宽带多媒体, 2007 (16): 15.
- [14] 2010年数字电视产业标准与知识产权论坛召开 [J]. 信息技术与标准化, 2010 (10): 7.
- [15] 徐惠敏, 葛进平. 数字电视产业发展的政府规制 [J]. 科技信息, 2008 (23): 24-25.
- [16] 胡汉辉, 万兴, 周慧. 网络融合下中国数字电视产业的规制与发展 [J]. 产业经济研究, 2010 (4): 1-8.
- [17] 孙志行, 宋玉炎, 何道刚, 等. 数字机顶盒技术及其市场前景 [J]. 电视技术, 2002 (1): 46-49.
- [18] 刘清涛. 中国数字家庭产业推进数字生活新体验 [J]. 世界电子元器件, 2007 (9): 24-25.
- [19] 姜锡山. 发展我国的数字内容产业 [J]. 中国创业投资与高科技, 2005 (8): 68-69.
- [20] 陈治光. 首届“中国北京数字内容创新发展高峰论坛”成功举行 [J]. 科技潮, 2007 (10).

作者简介: 周庆山, 男, 1965年生, 博士, 教授。

赵雪, 男, 1981年生, 博士生。研究方向: 数字媒体与网络传播。

赵需要, 男, 1978年生, 博士生, 助理馆员。研究方向: 信息政策与信息法。

周格非, 男, 1988年生, 硕士生。研究方向: 传播学。

收稿日期: 2011-10-10