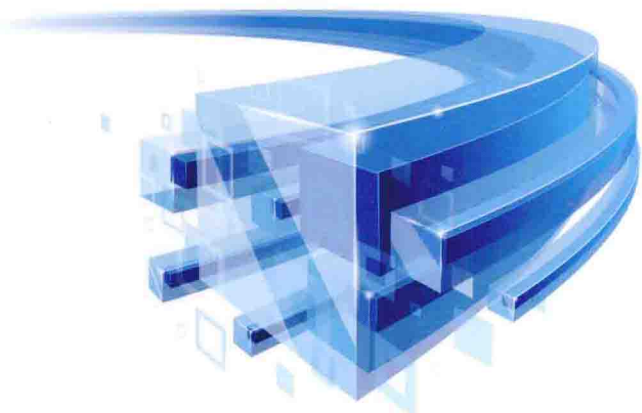


自然科学基金重大研究计划/重点支持项目 资助
青年科学基金项目



跨媒体

信息技术与应用

杨毅
王胜开
陈国顺 编著
徐为群
黄鹤云



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

跨媒体信息技术与应用

杨 毅 王胜开 陈国顺 徐为群 黄鹤云 编著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书首先对多媒体与跨媒体信息技术及互联网信息检索技术涉及的基本概念、基本理论、基本技术等进行了介绍,描述了跨媒体信息技术相对多媒体信息技术而言的突破及其发展状况;然后从文本、语音、图像、视频等媒体基本表现形式入手,介绍了主要的多媒体与跨媒体信息处理技术,以及国内外最新研究成果,勾勒出跨媒体信息处理的基本框架;接着对当前发展迅速的互联网信息检索与处理技术、模型、特点、应用等进行描述;最后以中国少数民族语言跨媒体共享平台、基于数字图书馆的跨媒体信息检索平台、跨媒体互动教育游戏平台系统等为例,对跨媒体信息检索方法、处理技术及其实际应用等进行了论述。

本书可供信息技术、模式识别、信号处理、多媒体与跨媒体信息处理等领域的工程技术人员、科研管理人员,以及相关专业的大专院校学生、研究生参考。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

跨媒体信息技术与应用 / 杨毅等编著. —北京: 电子工业出版社, 2014.9

ISBN 978-7-121-21834-7

I. ①跨… II. ①杨… III. ①多媒体—信息技术 IV. ①G202 ②TP37

中国版本图书馆 CIP 数据核字(2013)第 266762 号

责任编辑: 田宏峰

印 刷: 中国电影出版社印刷厂

装 订: 中国电影出版社印刷厂

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×980 1/16 印张: 15.75 字数: 350 千字

版 次: 2014 年 9 月第 1 版

印 次: 2014 年 9 月第 1 次印刷

定 价: 59.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zllts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

前 言

跨媒体是近年来伴随互联网和多媒体技术的迅猛发展而逐渐兴起的一种新兴信息媒体形式，正引起业界的广泛关注。跨媒体不只是指媒体的多样性、多样化，更涉及集成交互处理多种媒体信息的新理论、新技术、新手段和新方法及其新应用、新发展。

以更符合人的交互交流习惯、以更自然的方式方法表示和处理信息将是现代文明与技术进步的重要标志。跨媒体信息处理将以计算机、互联网、通信、音/视频处理技术等为支撑，结合最新的大数据、云计算等技术，把文本、语音、音频、图像、视频等复杂的多种媒体有机组合起来，建立不同媒体间的联系与关联关系，并对其进行加工处理，最终以丰富、自然、交互、高效的方式与形式呈现给观众、听众、受众。跨媒体信息技术的研究、发展与应用将极大拓展计算机技术与应用的领域，改变人们的工作、学习、生活和生产方式，给大众传播媒体等带来巨大影响和新鲜空气。

在当今的互联环境中，不断涌现的媒体对象之间的相互合作、相互补充、互动与协调，使得共生共存的多种、多个媒体对象之间已呈现出“跨媒体”的发展态势，这种态势既表现为文本、图像、音/视频等复杂媒体对象的混合并存，又表现为不同媒体对象之间构成复杂的关联关系和组织结构，还表现在具有不同模态的媒体对象跨越媒介或平台高度交互与融合。

跨媒体信息处理与应用涉及数据压缩与编码、数据传输与高速网络、数字图像与识别、数字音频与识别、数字视频与机器视觉、多媒体变换与模式识别、文本识别与超媒体、全息图像与传感网络、机器翻译与自然语言理解、机器学习与人工智能、数据库与多媒体索引、互联网、大数据与云计算等众多技术，目前正朝着互联网化、大数据化、智能化的方向发展，并逐步进入远程教育、远程医疗、文化娱乐、自动测控等领域，随着发展，关键技术得以不断突破、系统性能指标得以不断提高，人机交互的自然性、高效性越来越强，跨媒体信息处理与应用的重要性日益凸显，前景美好而广阔。

跨媒体发展与应用涉及各种信息检索、信息处理新技术，正引起学术界的广泛兴

趣，着力开展研究、积极寻求突破。跨媒体信息系统通过挖掘特征层面上多种媒体之间的相关性，试图找到底层视听特征与高层跨媒体语义之间的隐含关系，使得提交一种类型的多媒体查询例子，即可检索到语义相似、但类型不同的多种媒体对象，并以更直接、更自然的方式与形式提交给用户。跨媒体信息系统正是能够满足人们这种需求的一种新技术、新系统，有望实现从单一媒体检索与处理到多种其他媒体的检索与处理。

从全书结构与内容布局上，本书首先对多媒体与跨媒体信息技术及互联网信息检索技术涉及的基本概念、基本理论、基本技术等进行了介绍，描述了跨媒体信息技术相对多媒体信息技术而言的突破及其发展状况；其次，从文本、语音、图像、视频等媒体基本表现形式入手，介绍了主要的多媒体与跨媒体信息处理技术，以及国内外最新研究成果，勾勒出跨媒体信息处理的基本框架；再次，对当前发展迅速的互联网信息检索与处理技术、模型、特点、应用等进行描述；最后，以中国少数民族语言跨媒体共享平台、基于数字图书馆的跨媒体信息检索平台、跨媒体互动教育游戏平台系统等为例，对跨媒体信息检索方法、处理技术及其实际应用等进行了论述。

全书内容丰富，涉及多媒体信息检索与处理、互联网信息检索与处理、跨媒体信息检索与处理等众多概念与技术，阐述深入浅出，理论描述与实例分析相结合，可作为信息技术与信息系统相关专业本科生、研究生学习多媒体和跨媒体新技术、新理论的教材与教辅材料，也可供广大技术研究和工程管理人员参考。

在本书编写过程中，清华大学刘润生教授、中国科学院声学研究所颜永红教授、中国科学院声学研究所潘接林教授等专家和学者，给予了大力指导和支持，并得到自然科学基金重大研究计划/重点支持项目(NSFC:90920302)和青年科学基金项目(NSFC:61105017)等的资助，在此表示衷心的感谢！

因作者水平和经验有限，书中错误之处在所难免，敬请读者批评指正。

作者

2014年8月

目 录

第 1 章 绪论	1
1.0 引言	1
1.1 多媒体信息技术	1
1.2 跨媒体信息概念与定义	3
1.3 跨媒体信息表示	5
1.4 跨媒体信息处理	6
1.5 跨媒体信息技术应用	8
1.6 本章小结	8
参考文献	9
第 2 章 多媒体信息技术基础	10
2.0 引言	10
2.1 多媒体技术概述	10
2.1.1 多媒体定义与特征	10
2.1.2 多媒体系统基本结构	13
2.2 音频信息处理技术	15
2.2.1 音频信号数字化	16
2.2.2 音频压缩	19
2.2.3 音频合成	27
2.3 图像信息处理技术	29
2.3.1 图像信号数字化	29
2.3.2 图像压缩编码	31
2.4 视频信息处理技术	35
2.4.1 视频信号数字化	35
2.4.2 视频压缩编码	36

2.5 本章小结	41
参考文献	41
第3章 多媒体信息技术应用	42
3.0 引言	42
3.1 应用概述	42
3.1.1 媒体类型	43
3.1.2 关键技术	44
3.1.3 应用系统	45
3.2 多语种文语转换系统	46
3.2.1 概述	46
3.2.2 汉语音节结构	47
3.2.3 语音规则系统	48
3.2.4 声学参数	48
3.2.5 语调规则	48
3.2.6 声学模型	48
3.3 语音身份识别系统	50
3.3.1 概述	50
3.3.2 基本框架	52
3.3.3 特征参数	53
3.3.4 主要方法	55
3.3.5 系统评价	56
3.4 视频点播系统	57
3.4.1 概述	57
3.4.2 系统结构	58
3.4.3 视频点播应用与云平台的交互	60
3.4.4 应用展望	62
3.5 地理信息系统	63
3.5.1 概述	63
3.5.2 系统结构	64

3.5.3 应用展望	66
3.6 数字水印系统	67
3.6.1 概述	67
3.6.2 系统结构	67
3.6.3 基本要求	70
3.6.4 算法分类	71
3.6.5 关键技术	71
3.6.6 应用展望	74
3.7 本章小结	74
参考文献	75
第4章 互联网信息检索技术基础	76
4.0 引言	76
4.1 信息检索技术基础	76
4.1.1 定义与分类	76
4.1.2 基本过程	78
4.1.3 信息检索评估	79
4.2 信息检索建模	81
4.2.1 基本概念	81
4.2.2 模型抽象	82
4.2.3 模型分类	83
4.3 信息检索经典模型	85
4.3.1 布尔模型	85
4.3.2 向量模型	87
4.3.3 概率模型	89
4.3.4 经典模型比较	90
4.4 信息检索要素	91
4.4.1 基本概念	91
4.4.2 文档标记	92
4.4.3 文档组织	94

4.4.4 查询	95
4.5 互联网信息检索	103
4.5.1 互联网特点与结构	104
4.5.2 搜索引擎架构	107
4.5.3 搜索引擎排序	109
4.6 本章小结	111
参考文献	112
第5章 互联网信息检索技术应用	113
5.0 引言	113
5.1 互联网信息检索系统	114
5.1.1 主要特点	114
5.1.2 发展趋势	114
5.2 互联网信息系统应用	115
5.2.1 搜索引擎	115
5.2.2 搜索引擎与传统信息检索的比较	116
5.3 基于云计算的多媒体检索技术平台	117
5.3.1 原理与功能简介	117
5.3.2 基于云计算的多媒体信息服务	122
5.4 基于物联网的多媒体信息检索系统	128
5.4.1 原理与功能简介	128
5.4.2 基于物联网的多媒体信息检索系统平台	130
5.5 智慧城市与智慧地球	138
5.5.1 智慧城市	138
5.5.2 智慧地球	141
5.6 本章小结	146
参考文献	146
第6章 跨媒体信息表示	147
6.0 引言	147

6.1 跨媒体信息特征	147
6.2 跨媒体信息度量	148
6.2.1 跨媒体信息表示	148
6.2.2 跨媒体信息检索排序	160
6.2.3 跨媒体信息降维处理	165
6.3 跨媒体信息相关性度量	170
6.3.1 针对聚类特征的度量	170
6.3.2 针对 MMDSS 的度量	172
6.4 其他跨媒体信息表示方法	173
6.5 本章小结	174
参考文献	175
第 7 章 跨媒体信息处理	178
7.0 引言	178
7.1 音频信息处理	179
7.1.1 音频的处理与分析	179
7.1.2 语音的索引与检索	185
7.2 视觉信息处理	189
7.2.1 图像信息的检索	189
7.2.2 视频信息的检索	197
7.2.3 视觉信息的高层语义特征提取	198
7.3 跨媒体信息理解	201
7.3.1 跨媒体语义	201
7.3.2 跨媒体信息理解	204
7.4 跨媒体信息检索	206
7.4.1 跨媒体信息的索引与检索	206
7.4.2 跨媒体信息的检索框架	207
7.5 本章小结	215
参考文献	216

第 8 章 跨媒体信息技术应用	219
8.0 引言	219
8.1 跨媒体信息系统简介	219
8.2 中国少数民族语言跨媒体共享系统	220
8.2.1 简介	220
8.2.2 中国少数民族语言跨媒体共享平台	222
8.3 方正跨媒体阅读系统	226
8.3.1 简介	226
8.3.2 方正跨媒体阅读平台	227
8.4 基于数字图书馆的跨媒体信息检索系统	229
8.4.1 简介	229
8.4.2 基于数字图书馆的跨媒体信息检索平台	234
8.5 跨媒体互动教育游戏系统	235
8.5.1 简介	235
8.5.2 跨媒体互动教育游戏平台	236
8.6 本章小结	242
参考文献	242

第 1 章 绪 论

➡ 1.0 引言

在 Web 2.0 时代，网络多媒体是以服务的形式兴起的。通过提供多元化的媒体服务，多媒体计算已成为一项受人瞩目的技术，它可以生成、编辑、处理、搜索各种媒体内容，如图像、视频、音频、图形。对基于互联网和移动无线网络的多媒体应用和服务而言，由于同时服务数以百万计的用户，需要大量的计算资源，因此对多媒体云计算的需求也非常大。在基于云的新型多媒体计算模式中，用户可以在云中分布式地存储和处理多媒体应用数据，不需要在计算机或终端设备上安装媒体应用软件，从而减轻用户对多媒体软件维护和升级的负担。

跨媒体技术的提出，正是基于多媒体云计算处理技术的蓬勃发展，它将更加符合人脑对视觉、听觉等多种认知和感知信息的综合处理模式，能够更好地完成对人机交互中多种不同类型媒体数据的管理、处理、分析和使用。

本章从多媒体信息技术定义入手，通过对跨媒体信息概念与定义等的界定，引出跨媒体信息的表示与处理方法，最后介绍跨媒体信息技术的应用。

➡ 1.1 多媒体信息技术

根据 2005 年清华大学出版社出版的《多媒体技术及应用》一书中的定义，媒体 (Medium) 是指信息表示和传输的载体。媒体在计算机领域有两种含义：一是指媒介，即存储信息的实体，如磁盘、光盘、磁带、半导体存储器等；二是指传递信息的载体，如数字、文字、声音、图形、图像等。

媒体有以下主要类别：

(1) 感觉媒体 (Perception Medium)。感觉媒体是指能直接作用于人的感官、使人直接产生感觉的媒体，如人类的语言、音乐、声音、图形、图像、计算机系统上的文字、数据和文件等，都是感觉媒体的例子。

(2) 表示媒体 (Representation Medium)。表示媒体是指为加工处理和传输感觉媒体而人为研究构造出来的一种媒体，其目的是为更有效地加工、处理和传送感觉媒体，包括各种编码方式，如语言编码、文本编码、图像编码等。

(3) 表现媒体 (Presentation Medium)。表现媒体是指在感觉媒体与用于通信的电信号之间用于转换的一类媒体，可分为两种类型：一是输入表现媒体，如键盘、摄像机、光笔、话筒等；二是输出表现媒体，如显示器、音箱、打印机等。

(4) 存储媒体 (Storage Medium)。存储媒体是指表示媒体的存储介质，如计算机的硬盘、软盘、磁带、光盘等。

(5) 传输媒体 (Transmission Medium)。传输媒体是指用于将媒体从一处传送到另一处的物理载体，是通信中的信息载体，如双绞线、同轴电缆、光纤等。

2008 年清华大学出版社出版的《计算机与信息技术基础教程》指出，多媒体技术是基于计算机技术的综合技术，包括数字信号处理技术、音频与视频技术、计算机硬件与软件技术、人工智能与模式识别技术、通信与图像技术等。

多媒体技术具有以下基本特征：

(1) 综合性。综合性是指将计算机、声像、通信技术等合为一体，将多种媒体有机地组织在一起，以共同表达一个完整的多媒体信息。

(2) 交互性。交互性是指人与计算机能“对话”，以便能够实现人工干预、控制等。

(3) 数字化。数字化是指多媒体中的各个单媒体都是以数字形式存放在计算机中。

(4) 实时性。实时性是指在多种媒体集成过程中，有些媒体（如声音、图像等）是与时间密切相关的，这就决定了多媒体技术必须支持实时处理。

在 2005 年清华大学出版社出版的《信息技术基础教程》一书中也提到，多媒体

(Multimedia)是多种媒体的综合,而多媒体技术则是对多种媒体进行综合的技术,确切地说,多媒体技术是把数字、文字、声音、图形、动画等各种媒体有机组合起来,利用计算机、通信、广播电视等技术,使之建立起逻辑联系,并能进行加工处理的技术。多媒体系统是一个可组织、存储、操纵和控制多媒体信息的集成环境和交互系统,综合、集成、多样、交互是其关键特性,尤其是多媒体系统的交互性,采取人机对话方式,实现了人与计算机之间的有效沟通。

多媒体是伴随着互联网和多媒体技术而快速发展起来的一种信息媒体形式,它以文本、语音、音频、图像、视频、三维模型等为基本内容,随着多媒体信息内容的爆炸性增长,多媒体检索、处理的必要性、重要性日益凸显。文本检索技术目前已相对成熟并得到广泛应用,而多媒体检索技术则相对复杂,目前主流的实用多媒体检索系统从根本上来说还都是传统的、基于关键词的文本搜索技术的扩展,但关键词很难充分和完整地表示相关的多媒体信息,因此,基于内容的多媒体检索和处理技术成为多媒体检索和处理领域的研究热点。

1.2 跨媒体信息概念与定义

在当今的互联网环境中,不断涌现的媒体对象之间的相互合作、补充、互动与协调,使得共生的媒体对象呈现出“跨媒体”的发展态势。这种态势既表现为包括网络文本、图像、音频、视频等复杂媒体对象混合并存,又表现为各类媒体对象形成复杂的关联关系和组织结构,还表现在具有不同模态的媒体对象跨越媒介或平台高度交互融合。通过“跨媒体”能从各自的侧面表达相同的语义信息,能比单一的媒体对象及其特定的模态更加全面地反映特定的内容信息。相同的内容信息跨越各类媒体对象交叉传播与整合,只有对这些多模态媒体进行融合分析,才能尽可能全面、正确地理解这种跨媒体综合体所蕴涵的内容信息。

跨媒体信息是指多媒体提取出的、能够跨越媒体类型的信息描述,一般来讲就是高层的语义信息描述,由于其与媒体类型无关,因此不同于传统多媒体检索方法中使用的底层物理特征,它更接近于人们的感知与认知习惯,符合计算机识别技术和系统

的最终目的，即最终使计算机能够有人类一样的识别、检索、处理和分析能力。如何填补底层特征和高层语义特征（即跨媒体信息）的差异是多媒体检索和处理技术研究和应用中最具挑战性的课题。

跨媒体是一个比较广义的概念，主要包括以下几个研究范畴。

1) 跨媒体检索

用户向计算机提交一种类型的多媒体对象作为查询例子，系统可以自动找到其他不同类型、在语义上相似的多媒体对象。虽然不同类型的多媒体对象之间没有直接的可比性，如一幅山水画和一段描述小河流水声的音频在底层内容特征上彼此异构，但却可以用机器学习、统计分析等方法学习二者在统计意义上潜在的相关性，并以此为依据进行跨媒体检索。

2) 跨媒体推理

推理是指从一个命题合理演绎到另一个命题，跨媒体推理就是从一种类型的多媒体数据，经过问题求解，转向另一种类型的多媒体数据。例如，OCR（Optical Character Recognition）技术是从图像到文本的推理、基于内容的图像检索是从图像到图像的推理、视频动画技术是从视频数据到动画序列的演绎等。跨媒体推理囊括了对这些不同类型的多媒体数据之间的转换的研究。

3) 跨媒体存储

现有处理海量数据的检索技术主要针对的是文本信息，如谷歌和百度等搜索引擎，针对多媒体检索的研究工作的出发点并不是针对跨媒体海量数据；跨媒体存储研究高效压缩、索引和分片等方法，以及对用户行为的个性化索引等技术，用于提高海量环境下的跨媒体检索效率，以更好地支持上层应用。

上述三点，从底层数据存储到上层应用技术的不同方面，描述了跨媒体思想对多媒体研究领域的技术涵盖和突破性要求，是一个整体性的研究框架设计。要实现上述研究思路，需要在海量数据库、多媒体索引、并行计算、机器学习和统计分析、计算机视觉，以及信息检索等诸多领域取得突破性进展。

1.3 跨媒体信息表示

单对某一种媒体类别，可以通过底层特征来表示语义信息，获得相关度判断，并可进一步获得检索结果。但若涉及多种媒体类别之间的信息检索，单靠这种方法已无法有效地表示跨媒体信息。其主要原因在于音频和图像的底层特征完全不同，表示同一个语义特征的音频和图像在底层特征空间上可能完全不同；另外，从维度上讲，不同类别媒体的底层特征向量也不同，因此必须寻求一种统一的方法来表示。

多媒体的数据结构性不高，一般地，基于内容的多媒体内容分析都是用底层特征来表示的。不同类别媒体的特性完全不同，没有可比性，媒体与媒体内容之间存在“鸿沟”。“跨媒体内容鸿沟”可定义为：因不同类别的媒体数据分别使用不同维数、不同属性的底层特征进行表示，使不同类别的媒体之间无法直接根据特征来计算其相关性，而造成的彼此之间的异构性和不可比性。

数据表达是基于内容的多媒体检索实现相似度匹配的基本依据，在图像检索、音频检索等研究中，通常采用底层内容特征来表示多媒体对象。而对跨媒体检索而言，需要研究一种符合跨媒体特性的数据表达机制，建立统一的数据模型，以表达不同的多媒体对象；同时，构建的模型应能反映数据集之间的潜在关系。如果不同类别媒体的数据集被映射到一个统一的表达框架中，但破坏了数据集之间的相互关系，那么在此基础上得到的跨媒体检索结果将受到影响，因此，在用统一的模型来表达不同类别的媒体数据时，应尽可能地发现和保持数据之间潜在的相关性。

不同类型的数据集之间实际上还存在一种共生关系，挖掘这种数据共生关系可以帮助发现潜在的语义结构。视觉特征和听觉特征结构完全不同，然而在其底层特征中潜在共生关系，可以从中找到隐含的语义结构。共生关系分析最初应用于语言分析，不过语言分析与跨媒体分析有很大的差别，主要在于跨媒体没有包含直接的语义；另外，跨媒体中不包含文档和词汇（语义）的包含关系，因此，其对应的处理方法也不同。

跨媒体信息表示的目标是通过最大限度地分析和获取多媒体信息之间的相关性，

对多媒体信息进行整合和排序,以便进一步对其进行识别、检索、重构等处理。跨媒体信息表示涉及范围很广,主要包括智能信息处理、数据挖掘、模式识别、数据库等多个交叉学科。但到目前为止,跨媒体信息表示中的知识发现和表达,以及跨媒体的统一表示结构等问题仍未得到根本解决,需要技术人员进一步开展研究。

➡ 1.4 跨媒体信息处理

互联网的发展带来了更多的多媒体应用,随着多媒体网站的日益发展,针对跨媒体信息的检索和提取技术越来越引起人们的关注。但到目前为止,这方面的研究尚处于发展时期,大多数的信息提取和检索方法都采用文字检索或人工标注来进行,将较为复杂的跨媒体信息转移到较为成熟的文本信息上。这种方法在多媒体个数较少时可以得到较好应用,但当多媒体的个数不断膨胀时,人工标注已很难胜任;而文字检索多采用上下文检索,也很难取得令人满意的效果。

基于内容的多媒体检索一般只关注媒体低层特征,且通常针对的是单一模态媒体对象,无法实现跨越各种媒体类型的综合检索。例如,当用户提供一幅图像给多媒体检索系统时,通常希望看到与此图像相关的其他图像、音频和视频等。具体来说,就是希望通过仅提供多媒体中的某一种类型媒体,就能得到各种类型媒体的相关检索结果。

为了实现多模态媒体间的无缝连接,研究人员提出了跨媒体检索问题,并积极开展相关技术研究。事实上,早在1976年,McGurk就已经揭示了人脑对外界信息的认知需要跨越和综合不同的感官信息以形成整体性的理解;同时,认知神经心理学方面的研究也进一步验证了人脑的认知过程呈现出跨媒体的特性,即对来自视觉、听觉等不同感官的信息相互刺激、共同作用而产生综合的认知结果。这种类型的检索就称为跨媒体检索。跨媒体也可以看做一种由各种基于相同语义媒体对象构成的复杂媒体类型。作为一种新兴的多媒体检索方式,跨媒体检索越来越受到国内外学术界的关注和重视。

一般媒体的检索过程通常为:用户提供一个多媒体实例,系统根据实例来对数据