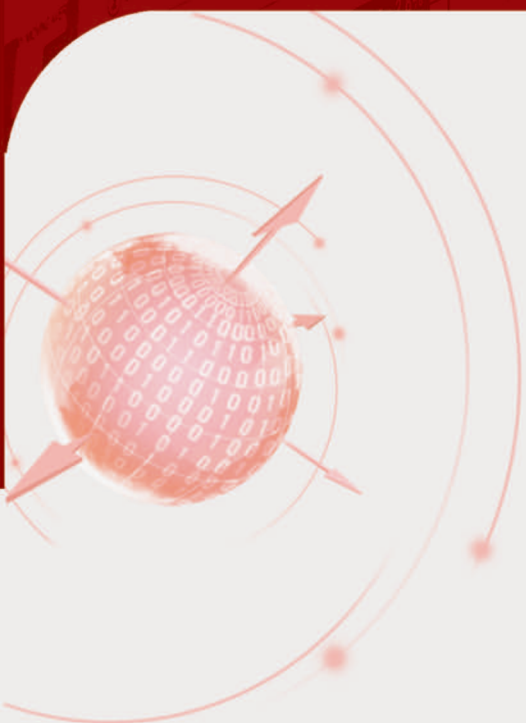
信息推荐系统

第二版

曾子明　著

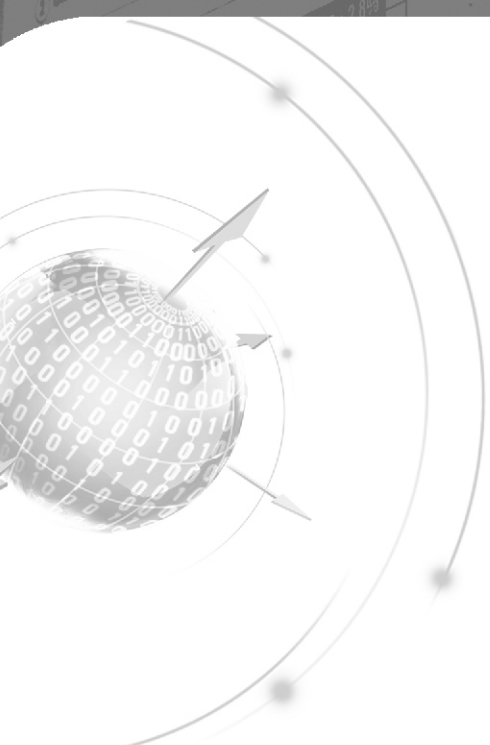


WUHAN UNIVERSITY PRESS
武汉大学出版社

信息推荐系统

第二版

曾子明　著



WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

信息推荐系统:第二版/曾子明著. —武汉:武汉大学出版社,2020.8
ISBN 978-7-307-21566-5

Ⅰ.信… Ⅱ.曾… Ⅲ.信息系统 Ⅳ.G202

中国版本图书馆 CIP 数据核字(2020)第 151506 号

责任编辑:詹 蜜 黄河清 责任校对:汪欣怡 版式设计:马 佳

出版发行:**武汉大学出版社** (430072 武昌 珞珈山)
(电子邮箱:cbs22@whu.edu.cn 网址:www.wdp.com.cn)
印刷:武汉中科兴业印务有限公司
开本:720×1000 1/16 印张:19.5 字数:329 千字 插页:2
版次:2020 年 8 月第 1 版 2020 年 8 月第 1 次印刷
ISBN 978-7-307-21566-5 定价:58.00 元

版权所有,不得翻印;凡购我社的图书,如有质量问题,请与当地图书销售部门联系调换。

作者简介

曾子明，男，1977年生，博士，博士后，教授，博士生导师，武汉大学信息资源研究中心研究员，入选为武汉大学“351人才计划”之珞珈青年学者。先后主持国家社会科学基金重大项目子课题、国家自然科学基金面上项目、国家自然科学基金青年项目、教育部人文社科重点研究基地重大项目、教育部人文社科研究项目、中国博士后科学基金、湖北省教育厅人文社科研究项目和企业委托项目等项目，参与教育部人文社会科学重点研究基地重大项目、香港特区自然科学基金研究项目、国家自然科学基金项目等。近年来，在国内外核心期刊和重要国际学术会议上发表论文80余篇，其中多篇论文被SCI、EI、CSSCI等重要检索机构收录，兼任国家自然科学基金通讯评审专家、教育部学位中心博士学位论文抽检通讯评议专家以及国内多家期刊审稿专家，同时兼任国内外多个学术期刊的审稿专家。

第二版前言

随着“互联网+”战略的推进以及信息技术的高速发展，大数据已成为我国经济和社会发展过程中重要的战略资源，被认为是未来引起社会变革和技术创新的“下一个前沿”。在互联网和不同行业和领域的深度融合过程中产生了大数据，其蕴含的丰富价值和巨大潜力，将给社会生产和生活带来深刻的变化，改变人们的信息获取和使用方法，以及社会生产、生活和学习方式。然而，基于大数据的互联网各种应用中出现了信息过载的问题，信息服务领域面临“信息丰富、但有用信息获取困难”的窘境，这使得人们从互联网中有效地获取信息变得日益困难。目前，搜索引擎是最普通的辅助人们获取信息的工具，但只能满足主流需求，没有考虑用户的个性化信息需求，仍然无法很好地解决信息过载的问题。以“信息推送”为服务模式的信息推荐系统，是当前解决信息过载问题的主要手段，它能够在分析预测用户需求的基础上主动推送其可能需要但又无法获取的有用信息，并能够以用户为中心，通过研究用户行为、兴趣和环境等，为用户推荐更具针对性的信息，即实现信息的“按需定制服务”。

信息推荐系统作为一种人机交互系统，主要应用信息检索、信息过滤、数据挖掘、深度学习等多种技术和方法为用户提供“信息推送”服务，帮助用户在互联网海量信息中筛选符合其个性化需求的信息资源，为用户带来全新的信息服务体验。随着电子商务、社交媒体的流行和发展，信息推荐系统作为信息服务科学一个重要的研究领域，已得到了国内外学者、研究机构和企业界的广泛关注。因此，系统地探讨信息推荐系统的基本原理、技术以及研究热点无疑将从理论和实践上推动信息推荐系统的进一步发展。

本书较为系统和全面地论述了信息推荐系统的相关原理、技术和应用，全书内容较新颖，反映了信息系统和信息服务领域的发展动态以及作者多年来的研究成果。相比于第一版，第二版在保留第一版基础知识和研究风格的基础上，做了较大幅度的调整和更新，使得第二版内容能够反映该领域最新研究热点和作者近年来的研究成果。例如，新版增加了三章内容，包括社会化信息推荐系统、基于用户画像的信息推荐系统、智慧图书馆信息推荐服务，与研究热点和应用领域紧密结合；同时，增加了相应章节内容，如基于深度学习的推荐系统、内容过滤推荐的用户兴趣建模、基于情境感知的个性化移动阅读推荐等，使得新版在内容上更加丰富完整。

全书共 9 章：

第 1 章是概述部分，在介绍两种服务模式的基础上介绍了基于“推送”模式的信息推荐系统，包括它的应用领域和研究热点。

第 2 章首先介绍了信息推荐的基本技术，包括信息检索、信息过滤、数据挖掘和信息推荐的基本算法。在此基础上，本章引入深度学习技术，在介绍深度学习技术的基础上，探讨了融合深度学习的信息推荐系统模型和方法，并对基于深度学习的推荐系统发展趋势进行了分析和展望。

第 3 章介绍信息内容过滤推荐的模型和方法，包括内容过滤推荐的相关技术、内容过滤推荐的系统模型和相关算法、内容过滤推荐的用户兴趣建模等。

第 4 章首先介绍两种基本的信息协同过滤推荐方法，在此基础上介绍基于模型的协同过滤推荐方法，并提出一种移动环境下基于隐式评分的博客信息推荐方法。

第 5 章针对顾客经常购买的商品，提出了一种基于 Web 挖掘技术的商品信息推荐系统。

第 6 章首先介绍社会化信息推荐系统的概念、研究现状和基本模型，在此基础上重点探索基于主题的科技文献社会化推荐模型和知识网络社区中基于声誉的社会化信息推荐。

第 7 章针对泛在环境下用户个性化信息需求具有情境敏感性，研究基于情境感知的信息资源推荐服务的理论、模型与方法，从不同研究视角提出相应基于情境感知的信息推荐方法。

第 8 章详细介绍用户画像的概念，探讨用户画像对信息推荐系统的推动作用，提出一种新颖的基于用户画像的信息推荐服务模型，并针对电影和移

动图书馆两个领域进行应用。

第 9 章从不同视角研究智慧图书馆个性化信息推荐服务，首先探索智慧图书馆个性化信息推荐服务体系；接着提出基于情境感知的智慧图书馆个性化推荐服务模型；而后提出基于 SoLoMo 模式的智慧图书馆嵌入式推荐服务模式，最后探索将推荐系统嵌入读者泛在学习过程，为读者提供泛在、全面、互联、高效的智慧型服务。

本书反映了作者及其领导的学术团队近年来的研究成果。在撰写、成稿的过程中，参考了国内外许多专家、学者的论著，他们的成果为本书提供了丰富的素材和理论支撑，书中在每章的章后参考文献中进行了标注。如果有不慎遗漏的，亦表示特别的歉意。

信息推荐系统可以无缝嵌入不同领域的应用场景中，为相关领域用户提供个性化的场景式服务体验。同时，信息推荐系统的理论和方法与时俱进、发展迅速，有许多新问题亟待进一步深入解决，因此本书出版希望能起到抛砖引玉的作用。尽管笔者在项目研究和本书撰写过程中付出了艰辛的努力，但由于该领域研究内容新颖，一些理论方法和技术还在不断发展之中，书中错漏之处难免，欢迎读者不吝赐教。

曾子明

2019 年 12 月

目　　录

第 1 章　信息推荐系统概述 …… 1
1.1　网络信息资源及获取服务模式 …… 1
1.1.1　网络信息资源 …… 2
1.1.2　信息获取服务模式 …… 3
1.2　基于“信息推送”模式的信息推荐系统 …… 9
1.2.1　信息推荐系统的概念与通用模型 …… 9
1.2.2　信息推荐系统与个性化信息服务 …… 12
1.2.3　信息推荐系统的研究内容 …… 15
1.2.4　信息推荐系统的分类 …… 17
1.2.5　信息推荐系统的发展现状和实例 …… 19
1.3　信息推荐系统的应用领域和研究热点 …… 22
1.3.1　信息推荐系统的应用领域 …… 22
1.3.2　信息推荐系统的研究热点 …… 25
参考文献 …… 27

第 2 章　信息推荐系统技术 …… 31
2.1　信息推荐系统的基本技术 …… 31
2.1.1　信息检索和信息过滤 …… 31
2.1.2　数据挖掘技术 …… 34
2.1.3　信息推荐算法概述 …… 36
2.2　基于深度学习的推荐系统 …… 43

2.2.1 问题的提出 …… 43
2.2.2 深度学习技术 …… 44
2.2.3 基于深度学习的推荐模型和方法 …… 50
本章小结 …… 61
参考文献 …… 62

第 3 章 信息内容过滤推荐系统 …… 68
3.1 引言 …… 68
3.2 内容过滤推荐系统的相关技术 …… 71
3.2.1 信息检索模型 …… 71
3.2.2 文本特征抽取 …… 73
3.3 内容过滤推荐系统的模型和算法 …… 75
3.3.1 内容过滤推荐系统的模型 …… 76
3.3.2 基于向量空间模型匹配的信息推荐算法 …… 78
3.3.3 基于朴素贝叶斯分类的信息推荐算法 …… 81
3.4 内容过滤推荐的用户兴趣建模 …… 83
3.4.1 问题的提出 …… 83
3.4.2 理论基础 …… 84
3.4.3 研究设计 …… 85
3.4.4 数据采集与处理 …… 87
3.4.5 实验 …… 89
本章小结 …… 93
参考文献 …… 94

第 4 章 信息协同过滤推荐系统 …… 97
4.1 引言 …… 97
4.2 基于内存的信息协同过滤推荐 …… 101
4.2.1 基于用户的信息协同过滤 …… 101
4.2.2 基于项目的信息协同过滤 …… 106
4.3 基于模型的信息协同过滤推荐 …… 109
4.3.1 基于降维技术的协同过滤推荐 …… 109
4.3.2 基于聚类的协同过滤推荐 …… 111

4.3.3　基于 Bayesian 的协同过滤推荐 …… 114
4.4　移动环境下基于隐式评分的博客推荐 …… 115
4.4.1　问题的提出 …… 115
4.4.2　相关工作 …… 116
4.4.3　隐式评分的计算方法 …… 117
4.4.4　基于隐式评分的协同过滤推荐算法 …… 119
4.4.5　实验及结果分析 …… 120
本章小结 …… 123
参考文献 …… 124

第 5 章　基于 Web 挖掘的商品信息推荐系统 …… 126
5.1　问题的提出 …… 126
5.2　点击流相关理论和技术 …… 128
5.2.1　点击流概述 …… 128
5.2.1　基于点击流的商品信息个性化推荐服务 …… 130
5.3　Web 挖掘技术 …… 131
5.3.1　Web 挖掘概述 …… 131
5.3.2　Web 挖掘与商品信息推荐系统 …… 136
5.4　基于 Web 挖掘的商品信息推荐模型 …… 137
5.4.1　商品信息推荐系统的体系结构 …… 137
5.4.2　商品分类树 …… 138
5.4.3　基于点击流的顾客偏好分析 …… 140
5.4.4　基于点击流的商品关联规则挖掘 …… 141
5.4.5　商品信息推荐算法 …… 143
5.5　商品信息推荐的实验及结果分析 …… 145
本章小结 …… 147
参考文献 …… 148

第 6 章　社会化信息推荐系统 …… 150
6.1　社会化信息推荐概述 …… 151
6.1.1　社会化信息推荐的概念 …… 151
6.1.2　社会化信息推荐的研究现状 …… 152

6.1.3 社会化信息推荐的基本模型 …… 154
6.2 基于主题的科技文献社会化信息推荐…… 155
6.2.1 问题的提出 …… 156
6.2.2 科技文献个性化推荐系统框架 …… 157
6.2.3 基于社会网络分析的个性化推荐建模 …… 159
6.2.4 基于社会网络的科技文献个性化推荐框架与流程 …… 165
6.3 知识网络社区中基于声誉的社会化信息推荐…… 167
6.3.1 问题的提出 …… 167
6.3.2 基于用户信任的评分调节 …… 168
6.3.3 用户与项目的声誉评估 …… 170
6.3.4 基于声誉的社会化协同过滤推荐原理及算法 …… 172
6.3.5 实验结果与分析 …… 174
本章小结 …… 176
参考文献 …… 176

第 7 章 基于情境感知的信息推荐系统…… 180
7.1 情境感知信息推荐的提出…… 181
7.1.1 情境感知推荐——个性化信息服务新模式 …… 181
7.1.2 情境感知推荐的研究现状 …… 182
7.2 融合多种情境的信息多维推荐服务模型…… 185
7.2.1 情境信息识别获取与语义描述方法研究 …… 185
7.2.2 信息资源多维推荐服务模型 …… 186
7.2.3 基于情境感知的信息资源推荐算法 …… 187
7.2.4 信息多维推荐服务的系统体系结构 …… 191
7.3 基于情境感知的个性化信息协同推荐…… 192
7.3.1 基于情境感知的协同过滤推荐 …… 193
7.3.2 实验及结果分析 …… 198
7.4 基于情境感知的个性化移动阅读推荐…… 199
7.4.1 基于情境条件熵的情境影响度 …… 200
7.4.2 情境感知的移动阅读个性化信息推荐 …… 203
7.4.3 实验及结果分析 …… 206
本章小结 …… 208

参考文献 …… 209

第 8 章　基于用户画像的信息推荐系统 …… 212
8.1　用户画像 …… 212
8.1.1　用户画像的概述 …… 212
8.1.2　用户画像的构成 …… 213
8.1.3　用户画像的构建流程 …… 216
8.1.4　用户画像的分类 …… 218
8.2　用户画像的研究现状 …… 219
8.2.1　基于主题的用户画像方法 …… 219
8.2.2　基于本体的用户画像方法 …… 220
8.2.3　基于兴趣的用户画像方法 …… 221
8.2.4　基于行为的用户画像方法 …… 222
8.3　基于用户画像的信息推荐服务 …… 224
8.3.1　基于用户画像的信息推荐——个性化推荐研究的新视角 …… 224
8.3.2　基于用户画像的信息推荐服务模型 …… 225
8.3.3　融合用户画像的推荐系统体系结构 …… 230
8.4　基于用户画像的电影个性化推荐 …… 232
8.4.1　基于用户画像的电影推荐系统模型 …… 232
8.4.2　电影用户画像建模 …… 233
8.4.3　基于用户画像的协同过滤推荐 …… 235
8.4.4　实验及结果分析 …… 240
8.5　基于用户画像的移动数字图书馆多情境信息推荐 …… 242
8.5.1　移动数字图书馆用户画像建模 …… 243
8.5.2　移动数字图书馆用户画像构建 …… 246
8.5.3　融合画像的协同过滤推荐方法 …… 248
8.5.4　实验及结果分析 …… 252
本章小结 …… 254
参考文献 …… 255

第 9 章　智慧图书馆信息推荐服务 …… 258
9.1　智慧图书馆与智慧服务 …… 258
9.1.1　智慧图书馆 …… 259
9.1.2　智慧服务 …… 261
9.2　智慧图书馆个性化信息推荐服务体系及保障模式 …… 261
9.2.1　智慧图书馆个性化推荐智慧服务需求 …… 261
9.2.2　基于智慧知识服务的智慧图书馆个性化推荐模型体系构建 …… 263
9.2.3　图书馆个性化推荐系统智慧服务模式保障 …… 266
9.3　基于情境感知的智慧图书馆个性化推荐服务 …… 268
9.3.1　情境感知和智慧图书馆 …… 269
9.3.2　基于情境感知的智慧图书馆个性化推荐模型 …… 271
9.3.3　基于情境感知的智慧图书馆个性化推荐体系结构 …… 274
9.3.4　智慧图书馆推荐服务应解决的关键问题 …… 276
9.4　基于 SoLoMo 模式的智慧图书馆嵌入式信息推荐 …… 278
9.4.1　嵌入式服务 …… 278
9.4.2　SoLoMo 模式 …… 279
9.4.3　基于 SoLoMo 模式的嵌入式信息推荐服务需求 …… 279
9.4.4　基于 SoLoMo 模式的嵌入式信息推荐服务模型 …… 280
9.4.5　基于 SoLoMo 模式的嵌入式信息推荐体系结构 …… 282
9.4.6　应解决的关键问题 …… 285
9.5　面向泛在学习的嵌入式信息推荐 …… 287
9.5.1　面向泛在学习的嵌入式推荐服务的定位 …… 287
9.5.2　面向泛在学习的嵌入式推荐服务模式 …… 289
9.5.3　面向泛在学习的嵌入式推荐服务的保障机制 …… 294
本章小结 …… 299
参考文献 …… 299

第 1 章　信息推荐系统概述

近年来，随着云计算、大数据、物联网和智能计算等信息技术的发展，互联网各种应用(电子商务、社交媒体和信息搜索等)层出不穷，数据资源的规模和数量呈几何级数激增，信息服务领域面临“数据丰富、但有用信息获取困难”的窘境。一方面，互联网各种应用产生了大数据 ，其蕴含着丰富的价值和巨大的潜力，给社会生产和生活带来深刻的变化，但同时出现了严重的“信息过载”(Information Overload)问题；另一方面，用户被这些数据所包围着，却无法从中有效获取自己所需的信息资源。针对信息过载的问题，网络信息获取以现代信息技术为手段，向用户提供所需的信息资源，其服务模式包括信息拉取和信息推送两种类型[1]。信息拉取包括门户网站、信息检索和搜索引擎等，但这些工具只满足主流需求，没有个性化的考虑，仍然无法很好的解决信息过载的问题[2]。推荐系统(Recommendation System)作为一种“信息推送”模式的重要方法，是当前解决信息过载问题的主要手段，它能够在分析预测用户需求的基础上主动推送其可能需要但又无法获取的有用信息，并能够以用户为中心，通过研究用户行为、兴趣和环境等，为用户推荐更具针对性的信息，即实现信息的“按需定制服务”。本章首先介绍网络信息获取及其提供的两种主要服务模式；在此基础上重点介绍基于“推送”模式的信息推荐系统；最后介绍信息推荐系统的应用领域和研究热点。

1.1　网络信息资源及获取服务模式

我们生活在一个越来越依靠信息的时代，并在向数字化时代迈进。数字

化时代来临时，各种信息资源的电子化传递都将成为数字化经济的标志[3]。信息社会化、社会信息化，信息生产与消费促进了信息产业和信息技术的飞速发展。当前，互联网已经成为人们获取信息的重要来源，是人们获取信息、改变生活方式、赢得商机的重要媒介。然而，互联网规模和信息资源的迅猛增长使得人们从互联网中有效获取信息日益困难。因此，针对用户的信息需求，如何高效地为其提供高质量的信息获取服务就成为当前互联网可持续发展的关键。

1.1.1 网络信息资源

网络信息资源是指以数字化形式记录的，以多媒体形式表达的，存储在网络计算磁介质、光介质以及各类通信介质上的，并通过计算机网络通信方式进行传递的信息内容的集合[4]。随着 Internet 的普及和发展，网络信息资源对日常生活及商务活动起到了越来越重要的作用。人们可以足不出户，方便地接触到大量的信息，信息资源不足的问题已不复存在。但是，人们也深刻感受到，目前最大的问题不是信息资源的缺乏和不足，而是网络信息资源严重膨胀，出现“信息过载”的问题；另一方面，用户被这些信息所包围着，却无法从中有效获取自己所需的信息资源，即用户面对的信息资源远远超出其处理的能力。现代信息服务领域已经进入了一个“信息丰富、但有用信息获取困难”的两难窘境。因此，针对用户的信息需求，如何高效地为其提供高质量的信息获取服务已成为当前网络信息资源健康、可持续发展过程中亟待解决的重要问题。

目前，Internet 网络上的信息资源主要有以下几个特点[5]：

第一，存储结构无序性。网络信息资源分散存储在 Internet 网络中不同的 Web 服务器中，采用不同的系统平台、数据管理平台、存取方式及人机接口，并缺乏集中统一的管理控制和质量控制，使得网络信息资源处于无序的分布式存储状态。

第二，形式多样性。网络信息资源形式广泛，涉及人们生活的各个领域。传统信息资源主要是以文字或数字形式表现出来的信息。而网络信息资源则以文本、图像、音频、视频等多种形式存在，涉及科学研究、商务、教育、娱乐、体育和艺术等不同领域。例如雅虎网站，其主页就是按字母排列的 15 个大类，464 个二级类目，每个二级类目又分为若干个三级类目。搜狐网站与雅虎网站相似，它将所提供的信息资源分为 18 个大类，各个大类

又细分为二、三、四级类目，各级子类目总数达 5 万多个，其主题几乎涉及各个行业和领域。

第三，网络信息资源的实时性。网络信息发布者根据各自的需要和迎合用户的需求，经常更新信息，甚至变换发布信息的网页形式和内容。在增加网络信息的同时，也增加了其动态性，这些都给信息获取增加了难度。CNNIC 于 2012 年 1 月发布的《第 29 次中国互联网络发展状况统计报告》指出，2011 年中国网页数量已达到 866 亿，年增长率为 44.3%。由于网络信息更新容易，传播迅捷，其实时性也是任何传统载体所无法比拟的。

第四，信息冗余度高。由于大量的网络信息是免费的，而且相关信息网站没有统一的规划，所以各个网络站点之间存在着大量冗余的信息，这就造成了信息检索过程中对相同信息的重复检索，这既增加了网络信息资源建设的成本和开销，同时也增加了用户获取资源的困难，浪费了用户的时间。

第五，信息资源的异构性。Internet 网上包含了海量的 Web 站点，这些 Web 站点的信息格式各异，有结构化的数据(如关系型的 Web 数据库)、半结构的数据(如 HTML 网页等)和非结构化的数据(如图片、广告视频等多媒体信息)。由于 Interent 开放式网络架构和网际互联协议的成功，如今人们可以轻易地在 Interent 网络许可的范围内访问任何信息资源。但网络信息资源在数据结构定义、语法格式、语义描述、不同信息系统的异构性以及访问语言上的异构，使得信息资源并不能得到有效的开发和利用，即存在“信息过载”现象。

Internet 网络的发展带来了信息资源的极大丰富，成为人们获取信息的重要来源，但是网络信息资源的无序多样、异构以及冗余度高等特点也给用户从 Internet 网络中获取信息带来了困难。人们通常借助各种网络信息获取工具，如门户网站、搜索引擎、专业数据索引、RSS 等，从网络获取信息。网络信息获取服务是指在互联网上，针对用户的信息需求，以现代信息技术为手段，向用户提供所需的信息资源及服务，其服务模式包括信息拉取(Information Pull)和信息推送(Information Push)[6]。

1.1.2　信息获取服务模式

信息拉取和信息推送是人们从网上获取信息的两种主要手段。信息拉取和推送的区别在于用户获取信息的模式不同。所谓信息拉取是用户根据自身的需求，通过信息服务系统搜索或浏览信息，经过不断筛选和重定位，找到

所需信息资源的过程；信息推送是指信息服务系统无须用户表明自身的信息需求，就能根据用户的历史访问记录以及所处环境等，主动在网上搜索信息，将符合用户需求的信息以合适的方式主动推送给用户的过程。可以看出，从信息系统服务的视角上，信息拉取是系统根据用户的请求，被动地提供信息服务；而信息推送则是系统主动地提供信息服务。两种服务模式之间的关系如图 1-1 所示。

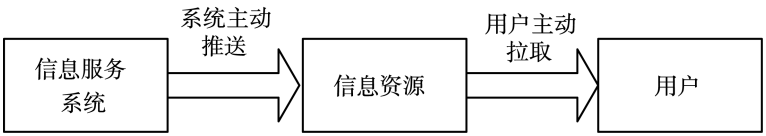


图 1-1 “信息推送”和“信息拉取”示意图

(1)信息拉取服务

信息拉取是用户获取网络信息资源的传统方式，它是指用户有目的地在网络上主动查询信息，其一般过程为：用户从浏览器给 Web 服务器发出请求，由 Web 服务器处理用户的请求，并将处理结果(即满足用户请求的信息或无法满足的信息)返给用户[7]。信息拉取服务的方式主要有两种：基于关键词检索的信息拉取和基于分类目录的信息拉取服务[8]。

①基于关键词检索的信息拉取。

这种信息拉取方式是搜索出符合检索条件的网页信息，信息服务系统服务器端通常采用基于机器人的技术，即使用一个被称为机器人(Robot，也叫做 Spider、Web Crawler)的机器人程序自动访问 Web 站点。机器人程序以某种策略自动地在 Internet 中搜索和发现信息，由索引器为搜索到的信息建立索引，根据用户的查询输入检索索引库，并将查询结果返回给用户。这种拉取服务方式是基于网页的全文检索服务，其优点是信息量大、更新及时、不需要人工干预；缺点是返回信息过多，没有个性化的考虑，用户仍需要从检索出的信息中手工选择自己所需要的信息。另外，基于关键词检索的信息拉取仍会出现两个问题：一个是用户有时很难简单地用关键词来准确地表达需要检索的内容，用户需求表达困难将导致检索困难；另一个是检索词的语义问题，同一概念可以用不同的检索形式来表达，如“计算机”和“电脑”，同时同一检索词在不同的上下文环境中可能语义不同，如“苹果”可以指一种