网站信息组织优化

——基于网络日志的用户行为分析

李志义 等◎著

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

华南师范大学哲学社会科学优秀学术著作出版基金资助出版

网站信息组织优化

——基于网络日志的用户行为分析

李志义　沈之锐　义梅练　著

電子工業出版社
Publishing House of Electronics Industry
北京 • BEIJING

内 容 简 介

本书采用了众多流行的数据挖掘算法，如利用 K-means 算法进行信息聚类和网页自动抽取，利用贝叶斯分类器实现信息过滤与分类，将知识组织与网站优化有机地结合起来，使得主题、目录组织的思想融合贯通在智能网站设计当中。全书共分 6 章，主要介绍了网络日志的数据来源、类型及其预处理技术；用户信息行为，包括网络用户行为的构成因素、分类，信息行为模型；用户行为数据的提取和分析，用户个性化知识服务需求的影响因素；网站优化算法的设计；智能技术在网站开发中的应用；机器学习的实现原理与训练模型，利用贝叶斯分类算法对垃圾信息进行自动过滤；最后，还对网站导航优化效果进行了调试与展示，并给出了实现的核心代码。

本书涉及数据挖掘、计算机编程、知识组织等多门学科的知识，理论性强。全书内容深入浅出，既有较深的理论分析，也有适当的设计案例，具有理论学习和实用开发双重意义。本书可作为高等院校信息管理与信息系统、电子商务、计算机等专业的本科生或研究生的教学参考书和教材，也可供从事智能网站开发、Web 挖掘等应用程序开发的工作人员参阅。

图书在版编目（CIP）数据

网站信息组织优化：基于网络日志的用户行为分析/李志义，沈之锐，义梅练著.
北京：电子工业出版社，2015.1
ISBN 978-7-121-25078-1

Ⅰ. ①网… Ⅱ. ①李… ②沈… ③义… Ⅲ. ①网站—信息管理 Ⅳ. ①TP393.092

中国版本图书馆 CIP 数据核字（2014）第 288406 号

策划编辑：薄 宇
责任编辑：周宏敏
印 刷：北京季蜂印刷有限公司
装 订：北京季蜂印刷有限公司
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：720×1 000 1/16 印张：10.75 字数：193 千字
版 次：2015 年第 1 版
印 次：2015 年 1 月第 1 次印刷
定 价：38.00 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 zlts@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

Web 技术的发展与演绎，有力地证明了网站信息组织必须与智能技术相结合，并应建立在用户行为的数据挖掘之上，充分利用弥足珍贵的点击数据流，以优化导航，提供个性化服务。本书正是出于此种理念撰写的，在全面研究国内外关于用户信息行为与网站信息组织的基础上，提出了网站信息组织优化的思想；结合用户信息行为分析，探讨了经过网站智能优化之后的网站特点以及它与传统网站的不同。另外，对用户信息行为进行了梳理，包括信息寻求行为、需求行为、浏览行为、检索行为、选择和存储行为以及吸收和利用行为等，分析了影响用户信息行为的因素，并通过一个 B2C 网站用户信息行为示例来分析基于用户认知的信息检索行为模式。这实际上是对网站开发技术的极大改良，比如文中涉及的智能推荐引擎能够根据用户的使用情况向用户有的放矢地推送信息；网页信息自动抽取技术，能够根据用户来访情况，分析用户需求，进而丰富网站内容；智能预测技术，能够预测网站来客，分析出什么样的用户最可能成为网站的注册用户，从而对其进行重点关注。本书将这些较复杂的算法原理，通过简单易懂的例子进行介绍，使智能算法的功能和使用可以被迅速地应用于 Web 实际开发中。

此外，本书还分析了如何对数据进行预处理，如何根据共同协作的方式来构建推荐引擎；介绍了编辑距离的计算、聚类技术在网页信息抽取中的应用，以及决策树算法在网站日志挖掘中的应用。通过各种算法的组合，网站设计者可以根据自身需求设计出属于自己的推荐系统、信息抽取系统、智能预测系统。另外，本书通过分析垃圾信息给交互性网站带来的巨大困扰，探讨了贝叶斯分类算法在社交网站信息过滤中的应用，并给出了具体实现的核心代码，这些代码可以很好地应用于针对中文的垃圾信息识别中。最后，介绍了一个智能网站的开发，展示

了基于网络日志优化后的网站效果。实验表明这种智能网站能对用户的各种信息行为进行分析，并将最优的导航效果和最佳的用户体验带给网络访问者。

本书为广东省哲学社会科学规划项目（项目编号：GD11CTS02）结项结果。毋庸置疑，书中的研究成果将对有关学科的发展与推动产生积极的影响。首先，将网络日志与信息组织优化、数据挖掘结合在一起研究，在学术上是一种较新的尝试；其次，目前能指导网站信息组织优化、进行网站设计的理论成果并不多，少有突破性的文献从用户行为的角度来探讨网站信息组织优化。所以，本书的出版将丰富用户信息行为的应用研究，拓展图书馆学、情报学有关用户服务的理论体系；另外，在一定程度上也将完善网站设计的理论与方法，进一步深化电子商务的研究内容，推动其学科发展。

本书的研究成果还具有较好的实际价值，突出表现为可推进人本计算、人本服务的理念，促进用户的网络体验并有效地利用网络数据和资源，解决网站导航可能出现的迷失问题，进而推动网站个性化推荐产品的研发，将传统互联网挖掘技术推广至移动互联网平台上。

全书共 6 章。第 1 章是绪论，主要介绍本书的写作背景、相关研究成果述评、研究的技术路径和方法，以及研究的主要内容。第 2 章对网站信息组织优化与网络日志挖掘进行概述，介绍了二者之间的关联，网络信息组织优化的基本内容，网络日志数据来源、类型及其预处理技术。第 3 章对网络信息用户行为进行了分析，包括网络用户行为的构成因素、分类，信息行为模型，用户个性化知识服务需求的影响因素，基于网络日志的用户行为数据的提取和分析。第 4 章探讨了网站信息组织优化算法的设计与实现。第 5 章对智能技术在网站开发中的应用进行了实例分析，介绍了机器学习的实现原理与训练模型，以及对垃圾信息进行自动过滤的贝叶斯分类算法。第 6 章给出了网站导航优化效果的展示，对导航优化前后的调试效果进行了比较，并给出实现的核心代码。“后记”给出了本书的结论和不足，并对下一步研究进行了展望。

本书是集体创作的结晶。首先感谢我的硕士研究生沈之锐、义梅练、杨雄威同学，他们在本书的资料收集、整理和章节的编写上付出了辛勤的劳动和汗水，本书正是与他们一起创作的结晶，在此谨致谢忱。

本书得到华南师范大学哲学社会科学优秀学术著作出版基金资助出版，我由衷地高兴并深感谢意！

还要感谢对本书的编写给予支持与帮助的所有朋友、领导和同事：感谢华南师范大学经管学院彭碧玉院长、王鸣老师对本书出版给予的鼎力支持和极大鼓励；感谢电子工业出版社策划编辑薄宇为本书的编辑与出版所付出的努力、支持和帮助；感谢本书中所引用和参考的前沿性研究成果的所有作者，在此特致深深的敬意；还要感谢电子工业出版社编审的辛勤工作和对本书出版的大力支持。最后，还要感谢我的爱人董志云，是她给予了我无限的动力，并不时督促我努力完成本书的撰写，深深感谢她在背后所付出的辛劳和莫大的支持，我将牢记在心；感谢我所有的家人，他们的支持令我力量倍增，在此我特向他们鞠躬并致谢意。

由于时间仓促，加之作者水平所限，书中难免存在错误和不当之处，恳请读者不吝赐教。我的电子邮箱是 leeds@scnu.edu.cn，欢迎读者与我联系，以使本书更臻完善。

李志义
于华南师范大学
2014 年 7 月 3 日

目　录

第1章 绪 论

1.1 背景介绍

1.1.1 目前网站建设存在的主要问题

随着电子商务以及基于 Web 的信息系统的不断发展和壮大，网站的日常运行积累了大量用户数据和点击数据流，这些数据都记录在网站日志里。对这些数据进行分析，可以帮助网站对其信息进行全面的架构，以优化 Web 应用程序的功能，并帮助网站分析客户的生命周期、设计产品、进行个性化服务和交叉销售等。基于网络日志（Web Log）的用户信息行为分析与网站信息组织优化研究，能够为访问者提供更加个性化的内容，并为网站设计更加有效的逻辑结构。

网络日志有两种理解，一种是指博客（Blog，即 Web Log 的合成词），也就是由个人管理、不定期张贴新的文章的网站。人们通过网络日志与人分享自己的一些喜怒哀乐、心情和生活点滴，在这个过程中增添更多见识。另一种是指服务器日志。本书关注的正是后者，即服务器日志。

服务器日志，是记录 Web 服务器接收处理请求以及运行时错误等各种原始信息的文件，它包括 Web 服务器访问日志和应用服务器日志。网络日志可以记录请求对象的关键点，例如，什么时候接收到请求、什么时候服务器响应、什么时候服务器完成对象的发送等。网络日志大概分为四部分：第一，使用记录数据。它是被 Web 和应用服务器自动收集的日志数据，体现了访问者的导航行为。它也是网站智能优化中使用的首要数据来源。第二，内容数据。包括传送给用户的对象和关系的集合，主要由文字材料和图片组成。第三，结构数据。它展示了以设计者的角度所看到的网站的内容组织结构，这种结构通过页面间的超链接来反映。第四，用户数据。包括注册用户的信息，用户对产品和广告的访问率，用户的购

买记录或者访问记录。

网络日志对于管理员来说是必不可少的，法律事件可能要求用日志来发现欺诈或者重现事件，但是本书则是利用它来监测用户的行为，并改进网站的结构。当前大部分网站存在的问题主要体现在以下方面：

第一，信息量过多，用户寻找需求的信息非常困难。琳琅满目的信息会增加用户选择的难度，造成极大负担。用户对此也会有下意识的抵制心理，甚至会放弃浏览。基于网络日志的智能网站的开发任务之一就是使网站具备智能推荐的功能。

第二，网站内容一成不变，与个性化需求相背。目前网站所提供的界面很多是主观的，不会随着顾客的需求而改变界面。网站为不同的顾客提供相同的信息和服务内容，这与个性化信息需求的趋势是完全相违背的。所以，网站智能优化的任务之二是使网站具有能够不断根据用户需求而扩展的能力，能够分析用户来源链接，改变网站的内容，将符合用户兴趣的页面抽取出来，提供给用户更多更丰富的内容。

第三，顾客对网站内的信息利用不足。很多网站是为了注册而注册，用户填写的注册信息常常被存储在数据库中，却得不到利用。网站服务器上存在大量用户点击、评价和购买的信息，这些信息实际上是非常宝贵的，但是目前有不少互联网上的网站还未使用相应的技术对这些信息进行挖掘和预测。对用户行为数据进行分析，可以分析出客户的行为模式，进而辅助决策和进行客户行为的预测。智能网站的开发将改变这种被动局面，使大部分用户数据得到利用。

本书分析了在互动性越来越频繁的今天，网站开发面临的各种问题；分析了在交互性越来越频繁的环境下网站开发必须充分利用网络日志的功能，并对网站内的信息进行分类和组织；在给出各种智能技术的理论基础上，进一步建模和实现智能化网站的各种功能，并把它应用于网站客户聚类、网站信息聚类、链接自适应以及用户行为预测等各个领域，解决网站中存在的个性化服务不足的问题和难题。

1.1.2 基于网络日志的用户信息行为研究具有重要价值

一般认为，人类信息行为主要是指以信息为对象所进行的各种收集和分析活动，具体包括信息收集、寻找、检索、处理和利用等一系列行为。

可以说，在人们日常的生活环境中，信息需求和信息行为是普遍的，随着社会信息化的不断升级，普遍的日常事务处理也变得越来越先进和复杂化。在这样

的情况下，除了人们的信息需求日益迫切外，更希望更方便地获取所需信息，且所接收的信息质量能够不断提高。通过信息需求的研究，有利于促进学术领域关于用户信息行为和信息服务的关注度。信息服务产业以客户满意度为导向，努力加强自己的形象，不断提升自身的服务能力，使用户的信息需求得到满足。就目前而言，信息服务能力有所提升，但许多研究都显示，很多信息工作和信息服务仍不能达到令人满意的程度，还有很多方面需要改善和提高。

所以，从科学的角度看，我们需要重新审视和继续探索用户的信息行为，解决信息工作和信息需求的满意度，为信息服务提供理论指导。

用户信息行为的研究主要集中在“信息需求”、“信息检索”、“信息使用”以及用户在网络环境下的各种信息交流活动等方面。近年来，虽然一些学者不断地吸收其他学科的理论和方法，将该研究推向新的高度，但并没有获得很大的突破。用户信息行为有着复杂的内部机制。首先，它涉及人的情感领域；其次，在信息显示方面，它表现出了用户的感觉、知觉、态度等信息。因此，需要使用多学科的知识，才能全面分析这个复杂的行为过程。除了使用情报学、计算机科学来研究用户信息行为，同时也应广泛借鉴其他相关学科，如心理学、社会科学、认知科学以及其他行为科学理论，并继续开展该领域的研究，这样才能使研究更加深入。

用户信息行为不只是对客观的用户行为规律进行分析，还要对用户心理进行分析。从主观上了解用户的行为目的、用户判断和选择信息的过程，并了解用户信息行为和目标之间的因果关系，才能更好地说明该用户的信息行为。因此，我们需要强调信息的主体作用，强调认知模型。知识结构和外部环境是用户信息行为的重要影响因素，加强对这些领域的用户信息行为研究，有利于开发用户的多样性信息行为分析框架。对用户信息行为进行分析通常有三个基本方向，即行为主体、行为对象和行为内容。

根据信息用户的信息行为和内部机制，可以从以下两个方面来解决所遇到的主要问题。

（1）指导和优化用户信息行为

从微观层面上说，用户信息行为的影响因素包括认知、情感和价值。用户是有目的的行动者，具有信息行为能力和先验知识。用户行为的心理偏见和行为能力大致与用户的信息素养和先验知识等相匹配，从这个意义上来说，用户的信息行为具有很大的可塑性。因此，本书的目的之一就是从用户的角度，根据用户的信

息素养、用户的个人基本特征和情感来引导优化用户的信息行为，提高用户信息行为能力，从而更有效地获得用户所需信息。

（2）研究用户信息活动和信息本身，以提高服务效率

信息活动和信息服务的成功，既取决于对信息的获取方式，也取决于双方对信息活动的认识和理解。在现实中，人们通过各种媒体来获取信息，主要是凭个人经验进行判断，而不是用科学的方法进行信息分析，因此有时获得的结果是不完整和有偏见的甚至得出错误的结论。而信息是从提供者的角度出发通过各种介质传递给用户的，有些信息服务提供者为了提高工作效率，在信息宣传方面做了很多努力，却不能让用户满意。究其原因，主要在于并未从用户的角度来考虑用户的需求。因此，我们需要转换思维方式，重新审视和理解用户所需要的信息，了解他们的评价和判断标准、他们的情感和认知。只有这样，才能解决信息活动深层次的问题。

1.1.3 网站设计与用户体验的最佳组合

基于网络日志的用户信息行为分析与信息组织优化具有以下优点。

（1）确定企业网站的客户基类

通过市场销售数据的分析来识别顾客的基类，确定网站上电商产品的布局显示，并给用户推荐产品，扩大产品销售。这有助于访问该网站，找到客户生命周期，制定相应的营销策略。

（2）提供个性化服务

根据用户的访问记录，向用户动态地推荐供应商产品。亚马逊和淘宝网等是个性化营销领域的经典网站，有巨大的商业价值。

（3）提高网站的设计水平

随着网站和 Web 服务的快速发展，网络结构的复杂性设计越来越多，维护的难度也越来越大，挖掘使用由用户提供的信息，通过网络，可以帮助网络网站设计者确定如何最好地修改和完善网站的结构。

（4）网站的评价

Web 使用挖掘可以得到用户到网站的使用情况的第一手资料，为网站评估提供证据。

（5）提高系统效率

利用数据挖掘技术，Web网站可以提供全方位的信息服务并提高信息效率。

通过Web日志等用户行为数据的分析和挖掘能帮助我们找到一个平衡服务器负载、优化传输、提高系统效率和服务质量的方案。基于以上论述可以看出，网络使用挖掘具有重大现实意义。实现的关键是准确有效地分析用户的网络行为，精准地分析和描述用户的兴趣。只有准确地把握用户的浏览行为，才能进行个性化的信息推送或协作推荐。准确描述用户的兴趣主要包括两个方面：①从用户的浏览器中获取信息；②挖掘出隐藏用户的兴趣点，使之准确地表示用户兴趣。

在预先不明用户的情况下，通常聚类算法可用来分析用户浏览信息的兴趣。聚类算法在用户信息挖掘中占有重要地位，通过研究网络用户的浏览行为模式，可以帮助网站运营商在网络经济时代获得更大的发展。由于网站运营商有关用户信息的数据在急剧增加，传统的数据分析方法已经很难满足大数据的需要。类似于Hadoop系统的解决方案能够处理海量（PB级或TB级）的各种结构的数据，能应对多个大数据挑战，并利用云计算平台，帮助用户开发应用程序快速摄取、分析和关联来自数千个实时源的信息。这种强大的数据挖掘的数据处理功能，可从海量数据中进行数据发现，挖掘出潜在的、有价值的信息，以协助网站运营。

对网站运作而言，用户数据的缺乏或不理解用户的需求，将无法把握业务发展趋势和市场规则。建立和扩大商家的业务市场，必须建立自己固定和有效的客户群。为了实现这一目标，应采取一切方法和技术来吸引目标客户。但是网站运营商的资源是有限的，它不可能为每个用户提供专门服务。因此，通过对Web日志进行分析和挖掘，进行用户聚类和信息组织优化就显得尤为重要了。

1.2 国内外相关研究综述

1.2.1 基于网络日志的网站信息组织研究

一般来说，信息组织是按照一定的标准、规则和方法，对信息的内部特征和外部特征进行表征和排序，将无序、杂乱的信息转换成有序、方便使用和流通的模式，也就是将信息有序化与优质化。网站信息组织既要将网站上的海量信息进行组织，让网站信息有序化、结构优化、界面整洁，又要方便用户浏览和使用。面对不断增长的海量信息，网站主要采用栏目分级和专题两种方式，再结合各级

导航栏和不同页面与页面元素间的超链接来完成。栏目反映的是网站纵向信息组织，而专题反映的是网站横向信息组织。

无论新建立的网站还是已经建设好的网站都要关注网站的信息组织问题。在构建一个新的网站时，应该预先想好如何设置栏目等问题；而网站建设好后，应该根据具体情况、网站用户的行为等信息对网站进行优化。

从目前该领域的研究现状看，关于网络日志方面的研究可分为两种类型（涂承胜，2003）：第一种类型是访问模式跟踪，其目标在于利用智能算法来分析用户访问日志，进而理解网站用户访问模式和访问倾向，进一步还可以改进网站的布局和结构。Kohavi 就把电子商务网站的日志信息应用于商务智能的规则发现中，如应用于客户转换率和寿命价值等方面。网络日志挖掘的另一种类型是定制个性化的用户服务，其目标是分析单个用户的使用偏好，然后根据不同的访问模式来提高个性化的网页页面，满足用户个性化的需求。Mobasher 等人最先提出了将网络日志作为个性化服务工具的观点，而学者 Pirrakos 也在这个领域做出了很多贡献。在网页个性化推荐方面的研究，国内学者易明（2010）提出了一种基于语义层次的 Web 个性化推荐方法，取得了很好的效果。基于网络日志的挖掘，学者陈新中总结出了两种方法，分别是基于 Web 事务的方法和基于数据立方体的方法。在基于 Web 事务的方法中，Chen 等（1996）首次将数据挖掘应用于网站日志分析中，提出最大向前引用算法 MFR 的概念。Zaiane et al.首次提出数据立方体的网络日志分析方法，这种方法将 Web 服务器日志转变为结构化的数据立方体，能够从多角度、全面地进行挖掘和分析。对于网络日志分析的预处理，学者也提出了很多观点，目前比较认同的是将预处理分为数据过滤、用户识别、会话识别以及路径补充等环节（Cooley et al，1999）。另外，比较新的观点还有根据用户之间的行为相关性或时间相关性进行数据归类。在网络日志的分析阶段，学者们还经常借鉴统计分析的方法，关联规则挖掘技术、分类技术、聚类技术、路径分析技术等。学者王有为（2009）通过分析网站内的链接，实现了网页内链接的自适应优化。汪定伟（2008）是高端电子商务网站设计方面的集大成者，他提出的很多模型都得到了很好的应用。

从另一个技术角度讲，智能网站需要分析来访用户是从什么网页链接过来的，然后据此抽取相应的页面进行分析，以便根据用户的需要丰富网页的内容，获取更多相关信息，并把最适合的内容推荐给用户。在这个过程中会涉及网页信息自

动抽取的问题。Web 信息自动抽取的关键是实现页面中模板的检测。在早期的研究中，网页信息抽取经常是以人工方式收集具有代表性的网页，并通过归纳学习，生成抽取模板。包装器就是一种专门用于 Web 信息抽取的技术，它首先对相同结构的网页进行聚类，从聚类后的网页簇中自动泛化生成高效准确的抽取模板。因此，网页聚类是实现数据精确抽取的前提，它是 Web 信息抽取流程中极为重要的环节。参考文献[1]利用网页 URL 规则对网页进行聚类，但随着 Ajax 技术的发展，动态 URL 不断流行，这种方法的准确性正在下降。RoadRunner 等[2]选取了包括主页距离及 URL 相似度等在内的指标来量化网页间的相似度，并采用聚类算法汇聚相似网页，提炼抽取模板。Gupta 等人设计的 Crunch 系统利用区域中链接文本和普通文本之间的比值与某个既定阈值的大小关系来确定网页的正文区域。他认为在正文区域中，普通文本所占比例较大。相反，在广告区域或友情链接区域中，信息大部分以链接文本的形式出现。参考文献[3]应用一阶 HMM 抽取计算机科学研究论文的标题、作者等头部信息；一阶 HMM 虽然也能很好地进行信息抽取，但是在参数获取过程中捕获上下文的信息有限。参考文献[4]使用二阶 HMM 对论文头部信息进行抽取，发现比一阶 HMM 具有更好的识别性能，但该模型是基于前向依赖性的，而未考虑可能存在的后向依赖。参考文献[5]面向结构简单的新闻网页，用树编辑距离量化网页的相似度，并用层次聚类法自动获取模板，该方法复杂度高，而且效率较低。微软亚洲研究院的 Cai 等人最早提出利用网页的视觉特征来抽取信息的 VIPS（Vision-Based Page Segmentation）算法。Liu 等人在 2003 年提出了 MDR（Mining Data Records）算法，并在 2005 年对该算法进行了改进。MDR 算法提出了一种独特的判断 Data Region 的方法，不但适合单正文体网页，对于多正文体网页的处理也收到了很好的效果。国内很多相关研究都是面向网页 DOM 树的，经过结构的泛化，生成一定的模板，从而寻找最可能是正文的节点，以实现网页的噪音过滤及主题内容提取。而在 4.2 节中提出的“基于重复模式识别的网页信息自动抽取”和“基于自然标注的网页信息抽取”方法是 Web 信息抽取的新方法。

1.2.2　基于网络日志的用户信息行为研究

用户的行为研究涉及的学科较多，无论是从社会科学还是自然科学的角度来进行研究分析都具有巨大的价值，其中心理学和社会学的研究较早。随着网络的

发展，网络用户行为的研究延续和扩大了用户行为的研究范畴，也得到了相关学科的极大关注，其中计算机学科的研究在实践中取得了极大的突破。

目前基于用户体验的网站优化研究，无论在国内还是国外，每年都有好几十篇的发文量，每篇文献所论述的均有所不同，或有交叉，或相互补充。其大体上可概括为以下三个方向。

1．基于网站设计的用户体验优化

传统的UED（用户体验设计）基本上都适用于网站的用户体验优化，它涉及美学、心理学等多个学科，也是目前研究的一个重要方向。网站的用户体验优化设计包括网站页面设计的优化、网站结构的优化和网站性能的优化等内容。

（1）网站页面设计的优化

网站页面设计的优化涉及图片、Flash、HTML代码、Web技术等方面的内容。综合所收集的文献，网站页面设计优化主要包括以下内容：

① 网站页面内容的优化。包括两方面内容，一方面，网站内容要进行高质量的及时更新。另一方面，优化网站内容使之适用于搜索引擎检索，如网页中关键字的密度设置、网页中关键字的重点分布位置等。网站内容要经常更新，对于缺乏经验的用户，认知易用性是决定网站重游的一个重要因素；而对于有经验的用户来说，认知有效性对其影响更多[6]。

② 参考文献[7～10]等对网页编码格式统一采用UTF-8编码格式，用W3C标准来设计网站；网页布局统一采用DIV加CSS；此外，还有Flash的使用，图片的使用，动态网页静态化，框架的使用；对404错误页面设置等都有学者做了相应的研究和探讨。张良瑛、张宇红在参考文献[11]中还专门探讨了404错误页面设计的用户体验。

③ 网站的交互性设计。参考文献[12]对网站交互设计的概念、目标和原则做了简要的概述。参考文献[13]介绍了交互式遗传算法。

（2）网站结构优化

网站结构优化就是对网站页面的存储方式以及内部链接关系进行合理的调整，以减小页面的目录深度与重要页面的链接深度[14]。结构优化分为物理结构优化和逻辑结构优化。网站物理结构是指网站真实目录及文件存储位置所决定的结