

Web 网络用户兴趣关联系统的设计

陈 冈,朱 茜

(武汉纺织大学 管理学院,湖北 武汉 430073)

摘 要:在信息过载的环境下,Web 用户要想获取感兴趣的信息,往往比较困难。协同过滤推荐的核心点在于基于相似性算法为同类兴趣的用户推送相似资源。文中设计了一种基于协同过滤推荐原理的用户兴趣关联系统,为 Web 用户主动推荐兴趣信息,这有利于解决电子商务环境中的信息过载问题。该系统基于 Java MVC 和 Flex 结合的总体架构,设计了兴趣关联处理流程和业务处理流程,研究了一种能够将业务逻辑自动装配的核心架构,提出了兴趣数据的切面和侧面关系数据库表示模型。最后简要介绍了实现技术。

关键词:兴趣关联;用户概貌;Web;相似性算法;切面;侧面

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2014)11-0199-04

doi:10.3969/j.issn.1673-629X.2014.11.050

Design of Web User Interest Association System

CHEN Gang, ZHU Qian

(School of Management, Wuhan Textile University, Wuhan 430073, China)

Abstract: In the information overload environment, it is more difficult for Web users to obtain the interest information. The core of collaborative filtering recommendation is to push similar resources for similar interested users based on similarity algorithm. In this paper, present a user interest correlation system based on collaborative filtering recommendation principle to advice interest information for Web users, which helps to solve information overload problems in the electronic commerce environment. The system, based on Java MVC and Flex, designs the interest association processing flowing and business processing flowing, researches an automatic assembly core which builds business logics, and proposes a relationship database representation model based on aspects and sections. Finally briefly introduce some implementation technologies.

Key words: interest association; user profile; Web; similarity algorithm; aspect; section

0 引 言

在信息过载的大环境下,Web 网络用户往往需要耗费大量的时间和精力,才能获得自身感兴趣的内容。这是大量 Web 网站系统存在的弊端:缺乏一种基于用户兴趣偏好的角度展现关联信息的服务。

一般认为兴趣关联^[1-3]信息服务就是挖掘并自动提供给用户符合他们个人需求和偏好的信息。兴趣关联信息服务系统与传统 Web 站点的定向推送服务不同,兴趣关联信息服务是由计算机系统主动进行、基于用户需求模型导向的 Web 系统。因此,传统的用户被动接收或定制信息的方式,将向更加智能化的主动服务模式演进。

1 系统原理说明

1.1 设计理论

网络用户兴趣关联系统基于协同过滤^[4]推荐技术来分析网络用户概貌^[5-6]。所谓用户概貌,实质上是用户个性化特征数据的集合,是兴趣关联系统个性化信息服务的基础和核心。而协同过滤推荐技术为电子商务中的“信息过载”问题提供了一条有效解决途径。

协同过滤推荐是一种基于海量数据挖掘平台的主动信息推荐方法,目前获得了广泛、深入的关注和研究。协同过滤推荐从用户角度出发,通过对 Web 用户的浏览、购买、评价等网络行为,发现其关联兴趣以及潜在的兴趣所在。协同过滤推荐技术的核心是相似性算法^[7-8],该算法的经典处理思路可概括为:基于同类兴趣用户对资源的评分数据,将其量化成推荐模型数

据。然后通过兴趣数据模型库,或者为用户寻找其兴趣相投的邻居用户,或者根据若干邻居用户对某项目的评分做出兴趣预测,也就是说重点关注于不同用户之间的兴趣关联性,以及可能或潜在的用户兴趣点,并形成一种协作推荐的关系,为电子商务活动中的信息服务组合、商品关联推荐、潜在兴趣发掘等提供决策支持。

1.2 兴趣关联模型

兴趣关联系统有两个着眼点:协同和关联。协同的目的是找到与当前用户有相似兴趣的邻居用户,关联的目的是将他们共同的兴趣信息主动推送给此用户。利用协同过滤技术来从用户的角度进行自动的关联推荐主要包括两方面内容:用户相似性和访问内容相似性^[9-12]。

(1) 用户相似性。

该相似性指标通过不同用户对指定分类内容爱好的相似程度来标识,指示出了不同用户兴趣爱好的相似程度,计算式如下:

$$\text{Sim}(a, u) = \frac{\sum_{i \in I(a) \cap I(u)} (R_{a,i} - R'_a)(R_{u,i} - R'_u)}{\sqrt{\sum_{i \in I(a) \cap I(u)} (R_{a,i} - R'_a)^2} \sqrt{\sum_{i \in I(a) \cap I(u)} (R_{u,i} - R'_u)^2}} \quad (1)$$

其中, $\text{Sim}(a, u)$ 代表 a 用户、 u 用户的相似性,这里 a, u 代表不同用户; I 表示评分项目的集合, R 表示具体的评分分值,例如 $I(a)$ 代表 a 用户评分的项目集合, $R_{a,i}$ 表示用户 a 对 i 项目的评分分值; $R_{a,i}, R_{u,i}$ 中的 i 项目,也就是 a 用户、 u 用户兴趣的交集项目; $R_{u,i}$ 表示用户 u 对项目 i 的评分; R'_a 和 R'_u 分别表示在共同评分数据中用户 a 和 u 的评分均值,计算式如下:

$$R'_a = \frac{\sum_{i \in I(a) \cap I(u)} R_{a,i}}{k} \quad (2)$$

$$R'_u = \frac{\sum_{i \in I(a) \cap I(u)} R_{u,i}}{k} \quad (3)$$

其中, $|I(a) \cap I(u)|$ 是不同用户共同评分项目的交集。

为了更准确地度量用户兴趣爱好的相似程度,可以对 $\text{Sim}(a, u)$ 进行相似性调整,也就是考虑共同评分项目数与预设期望经验值参数的比值来进行调整,即:

$$\text{Sim}'(a, u) = \frac{\text{Min}(k, \gamma)}{\gamma} \bullet \text{Sim}(a, u) \quad (4)$$

其中, γ 为调整参数,该参数根据实际项目的不同期望经验值进行预设; k 表示 a 用户、 u 用户共同评分的项目数; $\text{Min}(k, \gamma)$ 表示取共同评分项目数和期望经验值中较小的数值。也就是说,只有共同评分项目

数小于期望经验值参数时,才需要调整用户的相似性。

(2) 访问内容相似性。

访问内容相似性代表用户对某种内容访问兴趣的趋同程度,用 $\text{Sim}(i, j)$ 来表示,这里的 i 和 j 表示需要考察相似性的两类内容。具体处理时,主要根据用户对内容的评分数据来计算,具体计算公式如下:

$$\text{Sim}(i, j) = \frac{\sum_{u \in U(i) \cap U(j)} (R_{u,i} - R'_i)(R_{u,j} - R'_j)}{\sqrt{\sum_{u \in U(i) \cap U(j)} (R_{u,i} - R'_i)^2} \sqrt{\sum_{u \in U(i) \cap U(j)} (R_{u,j} - R'_j)^2}} \quad (5)$$

其中, $U(i)$ 是对 i 内容评分的用户集合, $U(j)$ 则是对 j 内容评分的用户集合,二者的交集自然代表对 i 内容和 j 内容都评过分的用户集合,在公式中用 u 表示; $R_{u,i}$ 表示用户 u 对内容 i 的单项评分, $R_{u,j}$ 表示用户 u 对内容 j 的单项评分; R'_i 表示用户群体对 i 内容共同评分数据的均值, R'_j 表示用户群体对 j 内容共同评分数据的均值。其计算式如下:

$$R'_i = \frac{\sum_{u \in U(i) \cap U(j)} R_{u,i}}{l} \quad (6)$$

$$R'_j = \frac{\sum_{u \in U(i) \cap U(j)} R_{u,j}}{l} \quad (7)$$

其中, $l = |U(i) \cap U(j)|$ 是对 i 内容评分用户和对 j 内容评分用户的交集数,即共同评分个数。与对用户相似性度量调整原理类似,为了更准确地度量用户对内容兴趣趋同的相似程度,可以对 $\text{Sim}(i, j)$ 进行相似性调整,调整式如下:

$$\text{Sim}'(i, j) = \frac{\text{Min}(l, \delta)}{\delta} \bullet \text{Sim}(i, j) \quad (8)$$

其中, δ 为调整参数,该参数根据实际项目的不同期望经验值进行预设; l 表示对内容 i 和内容 j 共同评分的用户数; $\text{Min}(l, \delta)$ 表示取共同评分用户数和期望经验值中较小的数值。也就是说,只有共同评分用户数小于期望经验值参数时,才需要调整访问内容的相似性。

1.3 系统总体架构

兴趣关联系统采用 Java 和 Flex 结合、并基于 Java 的 MVC 模式^[13]。Flex 与 MVC 可实现无缝组合; Flex 采用异步方式和服务器端通信,并对服务器端的数据进行处理,传递给表现层。用户与这些应用交互时,并不要求刷新页面,而且在通信中只传输已更改的那部分数据,减少了数据的信息量,有效地利用了网络资源。Flex 主要利用功能组件将界面切分开来,实现了代码的封装,符合 MVC 所倡导的处理原则,减少了程序代码的耦合度。系统总体结构如图 1 所示。

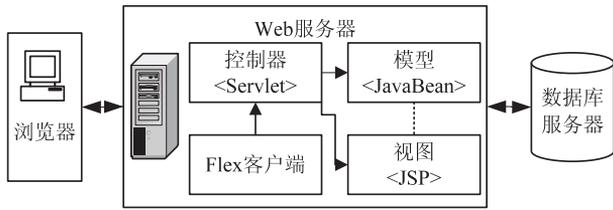


图 1 系统总体架构

2 系统关键设计

2.1 设计要求

网络用户兴趣关联系统设计的基本思想:相似性算法是系统构建的核心基础,在实现个性化信息服务时起着关键的作用;系统以该模型为依托,通过记录用户所访问的信息、搜集其提交的评分信息,来向用户主动推送其感兴趣的若干内容,以此改善用户对感兴趣内容的被动搜索,达到强化访问交互性的目的。

设计的基本原则:能很好地反映用户的兴趣偏好;能够关注用户兴趣信息的新领域,并向主动用户提供推送相应的信息服务内容;当用户兴趣偏好发生改变时,系统能够对用户概貌做适应性的改变。

2.2 兴趣关联处理流程

在用户首次浏览站点内容时,评分信息搜集机制就开始进行初始化运作。具体流程如图 2 所示。

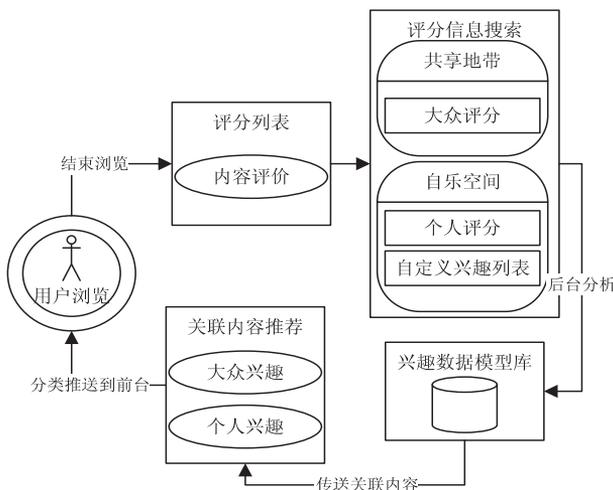


图 2 兴趣关联处理流程

对于个人用户(即已登录系统的用户),其访问权限及类别更高,同时,兴趣关联的实现更具个性。兴趣关联系统会向登录用户提供定制的分类访问信息。当用户自主选择某种类别的信息点浏览后,系统会根据列表信息来定制用户感兴趣的内容。对于公共用户(即未注册或已注册而未登录的用户),系统首先会随机地提供一些内容供用户访问,如新闻浏览、音乐播放等等。而这里的“随机”实际上是有目的的潜在导向:随机显示的都是用户评分较高、符合大众化的兴趣标准。而对于兴趣并非大众化的用户,系统则以友好的

互动方式向用户请求评价。

2.3 业务处理流程设计

兴趣关联系统核心部分包括共享地带和自乐空间。这两个部分的组合,以达到兴趣关联这个共同的目标。前者主要面向公共用户,后者则主要面向已登录的用户。自乐空间用户具有两个关联内容分类:不但可以获得基于大众共同需求特征的兴趣关联内容,还可以获得基于个性特征的私有兴趣关联内容。兴趣关联处理流程如图 3 所示。

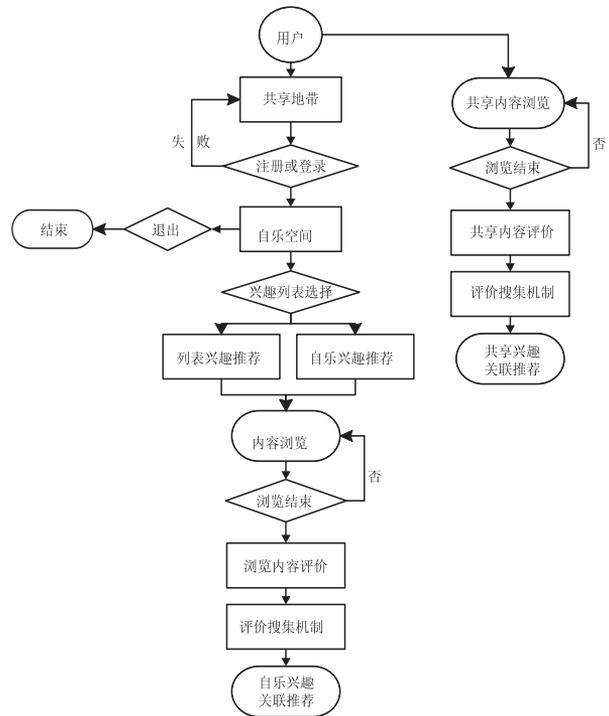


图 3 业务处理流程

3 实现技术简介

系统主要涉及开发平台有:Java MVC、MySQL、ActionScript3.0、RSS、Flex Builder3 等等。

Java MVC 是基于著名的模型-视图-控制器开发模式,实现了业务逻辑、数据、页面的良好解耦。MySQL 是一个开源关系型数据库管理系统,其突出优势在于体积小、速度快、开发成本低,目前已成为主流数据库产品之一。基于安全和隐私考虑,兴趣数据采用 AES^[14] 加密算法进行加密处理。ActionScript3.0 是 Flash Player Runtime 演化过程中的一个重要标志,用于创建一种适合快速地构建用户体验效果丰富的 Web 应用程序。RSS 则是 Web2.0 的组成部分之一。RSS 通过包含头部信息和主体内容的 XML 文件,能够实现用户兴趣数据或关联信息的表述、索引、解析和查询。Flex Builder3 是基于 Eclipse 的 IDE,将桌面应用程序的丰富性和 Adobe 业务平台的跨平台性相结合,构建一种将 HTML、XML、Java 等集成的 RIA 系统,更容易

提升用户浏览页面的兴趣。

Java MVC 用于构造兴趣关联系统的整体架构基础,MySQL 和 RSS 构建了兴趣关联系统的数据表示,ActionScript 和 Flex Builder 则用来组建有别于传统链接式 Web 的兴趣关联系统的人机交互接口。

3.1 业务逻辑自动装配

对于系统的业务逻辑处理,研究了一种能够将业务逻辑处理自动装配的处理核心。该处理核心能够自动装配常规数据处理操作,例如用户内容评价的 SQL 提交时,关联系统并不需要显式地编写 SQL 语句。对于兴趣点的关联查询,当视图页面增加或删除查询条件时,无需修改后台逻辑处理代码。业务逻辑处理核心提供了一个标准化、高质量的业务逻辑处理基础平台,提高了软件复用程度。业务逻辑自动装配架构如图 4 所示。

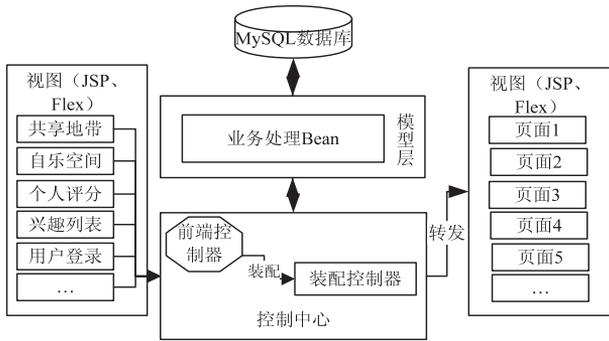


图 4 业务逻辑装配流程

来自于页面的业务处理请求,由处理核心的装配控制器自动装配,然后调用模型层相应的业务处理 Bean,并与数据库进行交互。最后,控制中心将结果数据自动转发给相应的结果页面。

业务逻辑自动装配核心吸收了 Spring MVC、Hibernate 二者各自的长处,例如 Spring MVC 的 MVC 架构优势,Hibernate 的数据与对象的逻辑映射。同时,避开了二者的一些不足,例如 Spring MVC 的繁琐配置,Hibernate 无法较好地适应视图搜索条件的改变,等等。

业务逻辑自动装配原型系统已经申请软件著作权,限于篇幅,不再赘述具体实现代码。

3.2 兴趣数据的切面和侧面

如何更好地描述用户的兴趣数据,并与关系数据库的表示方式结合,是一个需要考虑的问题。兴趣关联系统提出了切面、侧面的数据划分模式。侧面,用于描述用户兴趣关联的横向归纳点,例如新闻浏览、音乐播放、网络游戏等;侧面,则是描述切面细节性特征的纵向关键点。通过切面和侧面,与时间维度相结合(因为不同时间点,用户兴趣可能发生变化),就可以较好地勾勒出用户兴趣数据的立体模型。表 1 是兴趣

数据切面、侧面的示范性展示。

表 1 兴趣数据模型

侧面数据	切面数据				
	用户特征	新闻浏览	音乐播放	网络游戏	网络购物
关键点 1	性别	娱乐性	歌手	国内	网站类别
关键点 2	年龄	文艺性	作者	国外	物品类别
关键点 3	职业	体育	类型	类别	购买金额
关键点 4	地域	...	音质	模式	折扣情况
关键点 5

显然,上述模型非常切合目前广泛使用的关系数据库模式,这为模型的实现提供了的数据库基础,因为关系数据库就是使用二维表来表示数据之间的联系。

切面还可以进一步进行延展,例如对不同人群的音乐播放兴趣点建立切面数据,再如对不同地域(属于用户特征的侧面兴趣点)用户的网络购物兴趣点建立单独的切面。关于如何更合理的进行延展,还需要进一步研究。

3.3 关联信息解析和推送

信息解析主要是基于 XML 的 RSS 内容解析^[15],并将解析内容以 HTML 连接形式推送给用户。部分关键代码如下:

```

internal function parseRSS(xml:XML):void{//解析
var info:XML = xml;
var items:ArrayCollection = new ArrayCollection();//集合数据
for(var i:String in info..item){
var obj:Object = new Object();
var node:XML = info..item[i];
obj.title = node.title;//索引节点
obj.link = node.link;//索引标题
obj.pubDate = node.pubDate;
obj.description = node.description;//描述
items.addItem(obj);
}
s_n.newsList.labelField = "title";//标题
s_n.newsList.dataProvider = items;//项目点
}
internal function showNews():void{//推送
var item:Object = s_n.newsList.selectedItem;
s_n.news_txt.htmlText = "<a href = '"+item.link+"'>"+item.title+"</a>";
s_n.news_txt.htmlText += "信息描述:";
s_n.news_txt.htmlText += item.description;
}

```

4 结束语

兴趣关联系统的设计,较好地解决了提高用户兴

(下转第 206 页)

价,并设计了评价系统。该系统已通过运行测试,应用于实际的油田产能建设项目后评价中,能够对项目进行准确、高效的评价,为产能建设项目后评价的研究提供了有意义的参考。

参考文献:

- [1] 冯红霞,牛连峰,王敏.后评价在油田产能建设投资项目管理中的应用[J].内蒙古石油化工,2010,36(4):25-27.
- [2] 王纯光.油田建设项目的后评价[J].统计与决策,2002(12):67-67.
- [3] 杨文升.油气田产能建设项目后评价研究[D].哈尔滨:哈尔滨工程大学,2009.
- [4] Heston S L, Nandi S. A closed-form GARCH option valuation model[J]. Review of Financial Studies, 2000, 13(3):585-625.
- [5] 司训练,党俊.基于模糊综合评价的油田产能建设项目影响后评价[J].西安石油大学学报(社会科学版),2013,22(3):1-5.
- [6] 滑东武,丁文国.大庆油田产能工程投入效用定量研究[J].油气田地面工程,1998,17(5):66-68.
- [7] 赵启双.辽河油区难采储量综合评价[D].北京:中国地质大学,2003.
- [8] 唐丹.产能建设项目的经济后评价[J].油气田地面工程,2012,31(6):72-72.

(上接第 202 页)

趣度、优化评分搜集机制、完善关联内容推荐的处理。兴趣关联系统尚需要对数据的切面侧面划分、关联规则定义、匹配搜索与处理、关联信息触发机制等方面进一步优化进行研究。

参考文献:

- [1] 邓智龙,张海粟,黄立威.一种基于社区结构的用户兴趣关联规则发现方法[J].计算机应用研究,2012,29(5):1799-1801.
- [2] 叶鑫,刘宏志,安思.NAT环境下基于兴趣关联规则的重定向技术[J].计算机与数字工程,2011,39(7):113-116.
- [3] 张素智,苏龙飞.基于用户兴趣的协同过滤推荐算法研究[J].郑州轻工业学院学报(自然科学版),2013,28(5):47-49.
- [4] 冯永,陈显勇.基于评分信息量的协同过滤算法研究[J].计算机工程与应用,2013,49(20):198-201.
- [5] 张付志,魏莎.基于局部密度的用户概貌攻击检测算法[J].小型微型计算机系统,2013,34(4):850-855.
- [6] 杨鹤标,刘志然.基于语义事务信息聚类的用户概貌构建[J].计算机工程与设计,2010,31(20):4497-4499.

- [9] 任成峰.油田产能建设项目项目优化评价模型与控制管理研究[D].哈尔滨:哈尔滨工程大学,2006.
- [10] Yang Zengling, Chu Tianshu, Han Lujia, et al. Regional applicability evaluation of technical integration for straw feed utilization[J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(23):186-193.
- [11] Dai Li, Zhu Aihua, Zhao Yun. Using AHP to calculate optimization objective weights of transplanting mechanism[J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(2):60-65.
- [12] Shuiabi E, Thomson V, Bhuiyan N. Entropy as a measure of operational flexibility[J]. European Journal of Operational Research, 2005, 165(3):696-707.
- [13] Saaty T L, Tran L T. Fuzzy judgments and fuzzy sets[J]. International Journal of Strategic Decision Sciences, 2010, 1(1):23-40.
- [14] 杨帆,苏木标,李青宁.采用层次分析法的不同标度计算铁路混凝土梁桥的部件权重[J].西安建筑科技大学学报(自然科学版),2013,45(2):222-227.
- [15] 花拥军.项目社会评价指标体系及其方法研究[D].重庆:重庆大学,2004.
- [16] 陈剑红.层次分析法在移民安置后评价中的应用[J].水利技术监督,2007,15(4):44-45.

- [7] 袁书寒,陈维斌,傅顺开.位置服务社交网络用户行为相似性分析[J].计算机应用,2012,32(2):322-325.
- [8] 王琳,冯时,徐伟丽,等.一种面向微博客文本流的噪音判别与内容相似性双重检测的过滤方法[J].计算机应用与软件,2012,29(8):25-29.
- [9] Hinneburg A, Aggarwal C C, Keim D. What is the nearest neighbor in high dimensional spaces? [C]//Proceedings of the 26th international conference on very large data bases. Cairo, Egypt: UKPMC Press, 2000:506-515.
- [10] 李其申,屈喜琴,管俊.关联规则的相似性度量与聚类研究[J].计算机工程与设计,2012,33(2):745-749.
- [11] 王晓阳,张洪渊,沈良忠,等.基于相似性度量的高维数据聚类算法研究[J].计算机技术与发展,2013,23(5):30-33.
- [12] 张新霞,王耀青.基于统计相关性的兴趣关联规则的挖掘[J].计算机工程与科学,2003,25(3):60-62.
- [13] 陈冈.Java开发入门真功夫[M].北京:电子工业出版社,2009.
- [14] Wikipedia. Advanced encryption standard[EB/OL]. 2013-10-10. <http://en.wikipedia.org/wiki/MVC>.
- [15] Li Xin, Guo Lei, Zhao Yihong. Tag-based social interest discovery[M]. Beijing:[s.n.], 2008.

Web网络用户兴趣关联系统的设计

作者: [陈冈](#), [朱茜](#), [CHEN Gang](#), [ZHU Qian](#)
作者单位: [武汉纺织大学 管理学院, 湖北 武汉, 430073](#)
刊名: [计算机技术与发展](#) 
英文刊名: [Computer Technology and Development](#)
年, 卷(期): 2014(11)

引用本文格式: [陈冈](#). [朱茜](#). [CHEN Gang](#). [ZHU Qian](#) Web网络用户兴趣关联系统的设计[期刊论文]-[计算机技术与发展](#)
2014(11)