

网格环境下的个性化信息推荐服务模型研究¹⁾

苏新宁

(南京大学信息管理系, 南京 210093)

摘要 本文主要解析了个性化信息推送服务的形式,探讨了在网格环境下各种信息推送服务的模式,特别强调了在网格环境下推送服务中间件的作用,揭示了中间件在不同类型的推送服务中将如何工作。

关键词 网格环境 个性化服务 信息推荐 网格中间件

Model of Personalized Information Recommendation Based on Grid

Su Xinning

(Department of Information Management, Nanjing University, Nanjing 210093)

Abstract This paper analyses the forms of personalized information recommendation, discusses the patterns of information recommendation in grid, especially emphasizes the function of grid middleware, and indicates the principle of grid middleware in various recommendation systems.

Keywords grid environment, personalized service, information recommendation, grid middleware

1 引言

网络社会时代,人类既被淹没在信息的海洋中,同时又被一个个信息孤岛所隔离,造成人们在大量的信息中无法快速准确地得到所需要的信息。Web技术虽然实现了互联网上所有网页的连通,但并没有建立网络所有资源的全面连通。人们迫切希望有一种技术能够实现网络上所有资源的连通、共享和协同工作,网格技术无疑是实现人们愿望和建立网络资源全面连通的最佳方法。

计算机网格思想源于电力网格,它是对互联网上的所有资源(包括硬件、软件和数据等)实现连通,并能够共享资源和协同工作。国际知名网格专家 Ian Forster 把网格描述为“多个虚拟社团、机构间实现动态资源共享和协同解决问题的过程”^[1]。李国

杰院士把网格看作为是第三代互联网,他认为:“第一代互联网实现了计算机硬件的连通,第二代互联网实现了网页的连通,而作为第三代互联网的网格则试图实现网上所有资源的全面连通”^[2]。这里所说的所有资源应当包括:计算资源、存储资源、通信资源、软件资源、信息资源、知识资源等。

那么,网格究竟是什么?我们的理解是,网格是把整个网络构造成一台巨大的虚拟计算机,以实现网络上的计算资源、存储资源、信息资源、知识资源、专家资源等的全面共享,并促进分布在全球的硬件资源、软件资源和通信资源的全面协同工作。所以,我们既可以把网格看作是一种技术,也可以将其看作是设施。

可以看出,网格不仅仅将网络资源有机地联系在一起,而且还将实现网络服务的一体化。目前,虽然有许多图书馆提供了“一站式”检索的服务,但它

收稿日期: 2006年4月7日

作者简介: 苏新宁,男,1955年生,南京大学信息管理系教授,博士生导师,南京大学信息技术开发研究所所长,研究方向为信息自动化处理与检索,知识管理、引文索引及其分析研究。

1) 教育部人文社会科学重点研究基地重大项目成果(05JJD870004)。

还不是网络意义下的服务一体化和信息共享的模式,它的服务还是局限在一定的范围。因此,在网络环境下原有的信息服务模式必须要进行改革,以充分发挥网络的效能和满足人们对信息的需求。正是在这样一种需求下,我们仔细分析了传统网络下的个性化推荐服务模式,并结合网络的特点,来进行个性化推荐服务研究。目前,国内外发表有许多的基于网络的数字图书馆研究与实践方面的文章,例如文献[3]~[6]。但专门对网络环境下的个性化信息推荐服务的文章尚不多见。因此,本文将从解析个性化推荐服务功能出发,着重探讨了网络环境下各类个性化推荐服务系统的模式。

2 个性化推荐服务解析

个性化推荐服务是指信息服务系统为了满足个人特别需要的信息推荐服务。近几年来学术界和应用领域对个性化信息推荐的研究逐渐成为热点,发表有大量文章,我们在 CNKI(中国学术期刊全文数据库)的信息技术与应用类用“个性化”、“推荐服务”关键词去检索,命中有百余篇文章,有从算法研究和技术角度出发的文章,如文献[7]~[10],也有从信息服务角度出发的文章,如文献[11]、[12],这些研究为我们理解个性化推荐服务有很大的帮助。

总结个性化推荐服务,其特征主要表现在两个方面:其一,个人化的专门服务,即根据个人的要求和需要提供的个性化服务;其二,主动的信息推荐服务,即计算机系统主动将信息推送给用户的服务。个性化信息推荐服务的形式多种多样,我们大致将它们分为:定题信息推送、使用挖掘推送、热点信息推送和定制信息推送。图1给出了个性化推荐服

务的概貌。

3 网络环境下个性化推荐服务模型

当互联网进入了网络时代以后,整个网络就好像一个巨大的虚拟计算机,网络资源实现了共享,所有资源能够协调一致地工作,资源一体化,服务一体化,使信息服务部门的工作环境发生了又一次变革,过去在传统网络(互联网和 Web)下的信息服务也将得到全面的改造,同样我们正在研究的个性化信息推荐的服务模式也将会发生巨大的改变,它必然从原来孤立的系统,拓展到整个网络。我们将分别讨论网络环境下不同类型的个性化推荐服务模式。

3.1 网络环境下定题信息推送模式

定题信息推送是传统情报服务中的定题情报提供(SDI, Selective Dissemination of Information Service)发展而来,它是由用户参与(用户提交自己对信息的需求),ICP(内容服务提供商)根据用户提交的需求提问,定期(阶段或长期)的向用户提供与用户需求相关信息的个性化信息推荐服务。这项个性化推送服务所面对的主要是一些较为成熟的用户、或专业人员、或对自己的需求能够精确表达的用户,这些用户有一个共同的特点就是,他们将长期或未来的一段时间需要这类信息,而且需要的是最新信息。

整个定题信息推送的工作流程如下:用户在 ICP 提供的用户需求采集的交互界面上输入需求提问以及个人信息和网络通信方式(Email),ICP 将对用户给出的需求提问进行规范并存档(用户提问档),同时关联用户的 Email,以后每当数据库增加了新的信息,系统将自动激活推送软件,该软件将根据

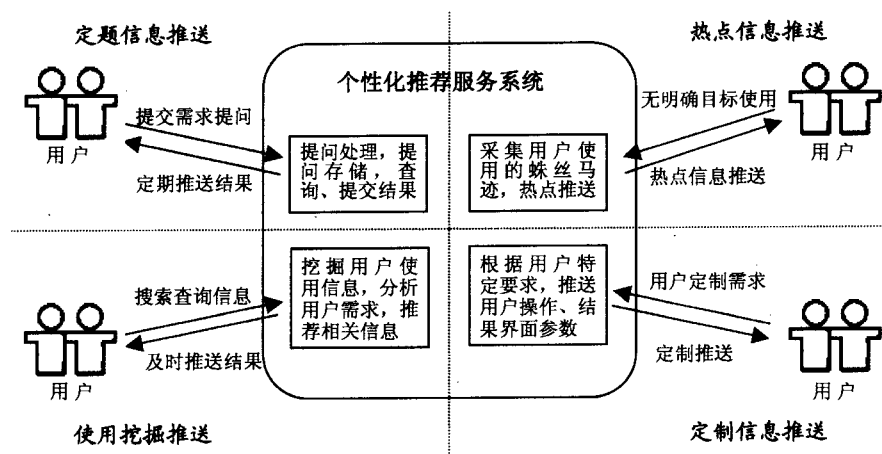


图1 个性化推荐服务概貌图

需求提问档信息对新入库的数据进行检索匹配,各提问的命中结果被分发到与提问相关的用户电子邮箱中。这样就保证了用户始终能够及时地了解它所需要的最新信息。

由于网络强调的是资源共享、分布计算和协同工作,所以,在网格环境下的定题信息推送应当是一个分布计算、统一推送的过程,整个系统的协调工作由一个叫定题推送服务中间件完成,具体的流程和运行模式参见图2。

由图2我们可以看到,在网格环境下,当用户把需求信息(提问)提交某一个定题信息推荐系统时,该定题信息推送系统除了对自己拥有的信息进行检索推送外,还把该需求提问转交给推荐服务中间件,中间件将分析提问主题,根据提问主题选择相应的信息服务系统,并将与其提问相关联系统的系统信息、检索程序名及交互规范回传给定题信息推送系统。当定题服务系统需进行新的信息推送时,把提问分别传送给相应的信息服务系统,实现分布计算(检索),并各自将检索结果返回定题信息推送系统,系统将结果按用户提问号归类汇总,然后发送给用户。

从以上的过程描述来看,推送服务中间件主要完成的是提问和有关信息系统的关联,整个定题检索结果是分布计算得到的,推送是由定题信息推送服务系统统一激活和汇集推送的,整个过程完全体现了网格的资源共享、分布计算和协同工作的特点。

3.2 网格环境下使用挖掘推送模式

使用挖掘推送是系统根据用户搜索网络、查询信息的过程中产生的使用信息,对其进行分析挖掘,获取用户的意图,然后将系统中与用户意图有关的信息发送给用户个性化推送服务。使用挖掘推送是用户在完全不知的情况下,由系统主动分析用户行为,挖掘用户需求主题,最后根据挖掘出来的主题

检索数据库或相关网站,得到的信息主动推送给用户。使用挖掘推送服务的关键在于捕获用户使用的关键词,跟踪用户的网上游历轨迹,以发现用户的兴趣。

使用挖掘实现信息的推送所面对的用户通常对网络的使用有一定程度的熟悉,能够在网络上搜索自己所需信息,但对网络资源的了解还不够全面、对信息检索技术的掌握还不是非常精通,因此,它需要网络能够帮助他快速、准确获取自己最需要的信息,使用挖掘推送信息就是完成这样的任务。使用挖掘推送和定题信息推送不同的是:它完成的是即时推送,而不是长期的推送;它是由系统主动获取用户需求信息的主题,而不是由用户向系统提交自己的需求;它是直接把信息推送到用户桌面,而不是把信息发送到用户邮箱。

在网格环境下通过使用挖掘向用户推送信息,其实现主要依赖于推送服务中间件,该中间件完成的主要任务就是抓取用户使用信息,发现用户的意图和挖掘用户关心的主题,还负责把信息直接推送到用户的桌面。整个过程参见图3。

图3反映出,当用户在使用网络和搜索信息时,推送服务中间件就将其使用网络的信息抓取下来,然后分析用户的使用行为和访问过程,根据用户使用的关键词和访问的网站或栏目等,挖掘用户关心的主题,用这些主题词去搜索相关的信息服务系统,最后,中间件将搜索得到的信息即时传送到用户的桌面上。整个推送过程是在用户不知觉的情况下完成的,更为重要的是,得到的信息可能是来自许多网站,中间件在用户和网络信息之间起到了穿针引线的作用。

3.3 网格环境下热点信息推送模式

热点信息推送是指将当前人们所关心的热点问题或某个研究领域内的热点问题及相关信息推送给

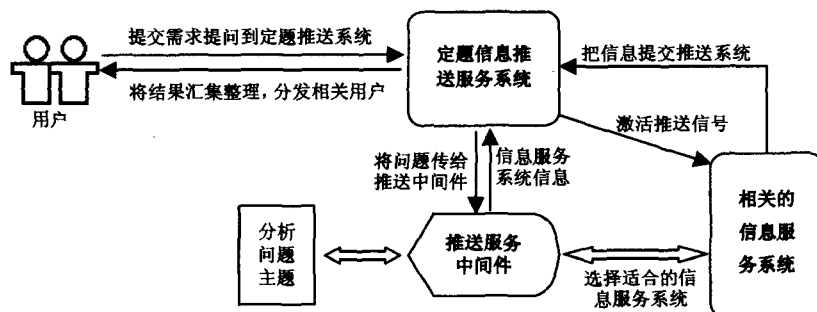


图2 网格环境下定题信息推送模式

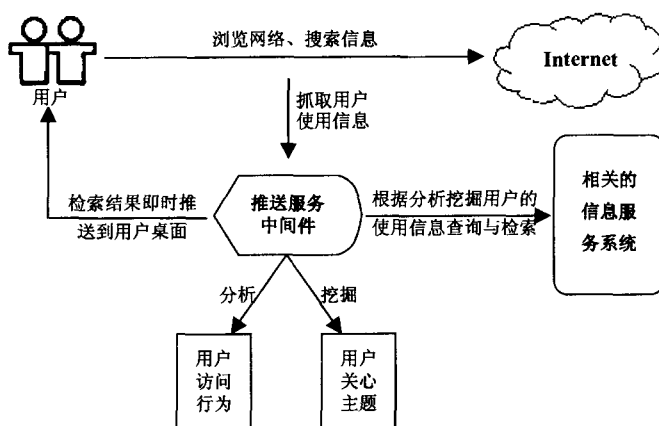


图3 网络环境下使用挖掘信息推送模式

有关用户的服务。热点推送主要是为对网络和网络资源不熟悉的用户提供的信息推送服务,它是根据用户在网络上游历过程和访问过程的蛛丝马迹,尝试性地推荐用户可能需要的热点信息。也可能是向任何一个用户推送当前大家所关心的热点问题和新闻事件,如我国神舟六号载人航天飞行期间,及时向每个用户提供有关神舟六号载人航天飞行的最新新闻。

热点信息推送和上面两种个性化信息推荐服务不同的是,用户可能并没有对信息有明确的需求,也可能对网络的资源和使用不甚熟悉,他们需要系统能够主动地推送一些当前热点信息,而这些热点信息的确定是系统预先根据对用户访问统计获得的。热点信息的推荐服务主要用于新闻热点和社会热点事件,但随着个性化推荐服务的深入,向学者推荐本学科研究热点的服务也将会兴起,这对大多数青年学者来说,是非常需要的,这种学科研究热点的获取,可以通过用户对专业数据库检索中使用的检索词分析来得到。

网络环境下的热点推送与挖掘使用的推送服务模式是相近的,运行流程也很相似,其差别主要在于:推送服务中间件按照设定的参数不断地从互联网中采集热点信息,针对那些无明确目标和主题的用户,推荐他们可能需要或关心的热点信息。其流程和模式参见图4。从图4可以看出,热点信息推送服务的中间件更注重对各类门户网站信息的采集,热点信息的采集主要是通过访问量得到的。

3.4 网络环境下个性化定制推送模式

定制信息推送是系统为了向用户提供更加符合用户本人个性化要求的网络交互界面和检索结果输出格式而提供的定制信息推送服务。这种定制推送主要包括这样几个方面:用户门户的栏目设定、个人访问页面的内容编排的设置、检索系统交互界面的检索项目和格式的设置、检索结果输出格式和编排要求的设置等。各种定制完全可以按照用户的要求自行定义,这些设置参数可以存放在客户端,也可以在服务端进行注册和定义。不论存放在哪一端,当

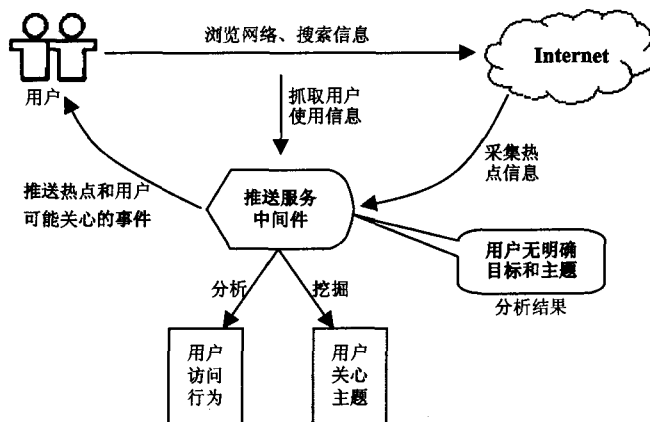


图4 网络环境下热点信息推送模式

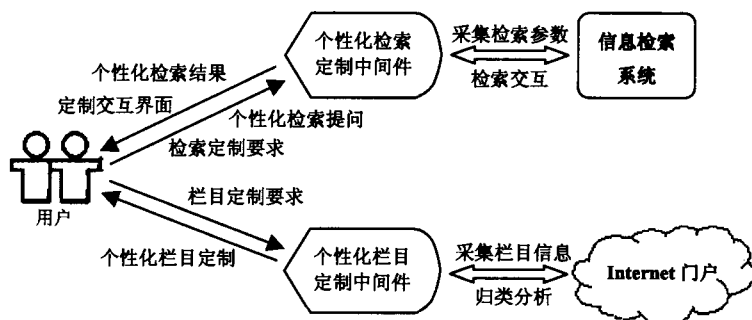


图5 网络环境下个性化定制推送模式

你访问与其关联的、提供个性化定制的网站时,这些参数将发生作用,为你构造一个自己定制的个性化页面。

个性化定制近几年已出现许多应用实例,一种是用户经常访问的门户网站的个性化定制,例如,著名的搜索引擎 Google 就提供了个性化定制服务,你可以根据自己的爱好和兴趣设定页面的栏目^[13]。许多检索系统很早就拥有了个性化交互界面的雏形,如简单检索界面和高级检索界面,目前已逐渐发展到根据你经常或习惯检索的字段来定义检索界面,还可以给出检索特定要求(如,精确检索或模糊检索),检索结果也可以根据用户的浏览习惯选定显示项数、顺序和排序栏等。

传统网络下的个性化定制推送服务主要来自于一个网站的定制或用户自己设置相关栏目,在网格环境下,他的定制推送将扩展到整个互联网,由个性化定制中间件采集检索系统的参数、采集互联网中重要和热点栏目信息,然后发送给用户,由用户自己选择、定制。整个流程由图5所示。我们可以看到,定制的要求来自于用户提出,中间件根据用户的要求采集相关信息提供用户进行个性化定制,以后用户就可以按照定制要求发送和接收信息了。

4 结束语

在信息社会中,在网络环境下,信息服务部门的服务将逐渐从被动的服务转向主动服务,从等待服务转向推送服务。个性化的信息推荐服务已得到信息服务部门(图书馆等)的高度重视。然而,网格世界的到来,个性化信息推荐服务也一定会发生变革,当我们接受了教育部人文社会科学重点研究基地重大项目《数字图书馆网格应用模型研究》以后,我们着重研究了网格环境下的个性化信息推荐模型,对不同类型的信息推荐服务进行了解析,探讨了在未

来网格环境下,推荐服务中间件的作用以及它的工作方式,并且给出了网格环境下的不同推荐服务系统的运行模式,希望我们的研究能对从事信息推荐应用研究的工作者有所帮助和指导意义。

参 考 文 献

- [1] Foster I, Kesselman C, Tuecke S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. International Journal of Supercomputer Applications, 2001, 15 (3):200-222.
- [2] 网格:机会在敲门. <http://www2.ccw.com.cn/02/0212/a/0212a05-2.asp>. Accessed: 2005, 11.
- [3] Shiri A. Schemas and Ontologies: Building a Semantic Infrastructure for the Grid and Digital Libraries. Workshop Report from E-Science Institute, Edinburgh, 2003.
- [4] Larson R, Sanderson R. Grid-Based Digital Libraries: Cheshire3 and Distributed Retrieval. Joint Conference on Digital Libraries, Denver, Colorado, USA, June, 2005.
- [5] 金海. 数字图书馆及其网格应用的发展. 现代图书情报技术. 2005, (9):1-5.
- [6] 李明禄, 薛广涛, 刘飞, 等. 基于多域和层次结构的网格数字图书馆. 现代图书情报技术, 2005, (10):1-4.
- [7] 曾春, 邢春晓, 周立柱. 个性化服务技术综述. 软件学报, 2002, 13(10):1952-1961.
- [8] 曾春, 邢春晓, 周立柱. 基于内容过滤的个性化搜索算法. 软件学报, 2003, 14(5):999-1004.
- [9] 崔林, 宋瀚涛, 龚永翌, 等. 基于 Web 使用挖掘的个性化服务技术研究. 计算机系统应用, 2005(3):23-26.
- [10] 杨焱, 孙铁利, 邱春艳. 个性化推荐技术的研究. 信息工程大学学报, 2005, 6(2):84-87.
- [11] 罗琳. 个性化服务与数字图书馆的发展. 图书情报知识, 2000(4):55-57.
- [12] 周青. 试论图书馆个性化服务与未来展望. 图书馆, 2004(1):73-75.
- [13] <http://www.google.com/ig>. Accessed: 2005, 11.

(责任编辑 芮国章)