

●赵水森

基于因特网的个性化信息服务研究

摘要 个性化信息服务,是针对不同用户采用不同服务策略和方式提供不同信息内容的服务。它具有以用户为中心、对用户需求进行挖掘、灵活多样和主动将信息推送给用户的特点。其类型有:个性化内容定制服务;个性化信息检索定制服务;个性化界面定制服务。用户个性化需求可通过用户访问记录挖掘、Bookmark 和 Agent 获取。个性化信息服务模型,可采用信息 Agent 自主学习法、信息过滤法、基于多 Multi-Agent System 的合作法等方法构建。参考文献 14。

关键词 因特网 信息服务 个性化 服务特点 服务类型 需求获取 模型构建

分类号 G252.6

ABSTRACT Personalized information services are to provide different service strategies and patterns to different users. In this paper, the author summarizes some characteristics, and introduces some service patterns and the modeling of personalized information services. 14 refs.

KEY WORDS Internet. Information services. Personalization. Service characteristics. Service patterns. Requirements. Modelling.

CLASS NUMBER G252.6

随着因特网的飞速发展,网上信息资源迅猛增长,现有的搜索引擎如 YAHOO 等越来越难以满足用户获取信息的需要。这些搜索引擎在服务中,一方面忽略用户的个性化要求,另一方面不能提供主动信息服务,使因特网上的海量信息和用户对信息的有限获取能力形成了强烈的反差,使得因特网上信息的获取和利用面临着严重的效率和质量问题。有效解决这些问题的办法,就是利用新技术提高用户获取信息的能力,使信息服务更加符合用户自己的需要。概括地说,就是建立面向用户的个性化服务机制。

1 个性化信息服务的定义与特点

1.1 个性化信息服务的概念

个性化信息服务是网络信息服务发展的重要方向,已经引起国内外许多专家学者的研究兴趣,因而近两年这方面的研究成果较多。但何谓个性化信息服务?个性化信息服务应包括哪些内容?如何实现个性化信息服务等问题,现在还没有统一的科学认识。所以在一些文献里,又将个性化信息服务称之为个性化服务、个性化定制服务、个性化信息搜索服务、个性化信息推荐服务、个性化信息提醒服务和个性化信息代理服务等等。下面是一些具有一定代表性的观点。

(1) 个性化信息服务是根据客户的特性提供具有针对性的信息内容和系统功能^[1]。

(2) 个性化信息服务就是将用户感兴趣的信息主动提交给用户^[2]。

(3) 个性化信息服务是基于信息用户的信息使用行为、习惯、偏好和特点,来向用户提供满足其各种个性化需求的一种服务^[3]。

(4) 所谓个性化信息服务,是指能够满足用户的个人信息需求的一种服务,在某一特定的网上功能和服务方式中,自己设定网上信息来源方式、表现形式、特定网上功能及其它网上服务方式等,而主动向用户提供可能需要的信息服务^[4]。

(5) 所谓个性化信息服务,首先它应该是一种能满足用户的个体信息需求的服务,即根据用户提出的明确要求,或通过对用户个性、习惯的分析而主动向用户提供其可能需要的信息和服务;其次它应该是一种培养个性,引导信息需求的服务,以此促进信息业的多样化和多元化发展^[5]。

(6) 个性化信息代理就是一个能够理解用户真正需求的智能 Agent。所谓个性是指 Agent 行为中所表现的一致的、稳定的、可信的、典型的特征和态度集;个性可以表示为有序的目标集和对于目标的规划的偏好程度^[6]。

笔者认为,个性化信息服务,就是针对不同用户

采用不同的服务策略和方式,提供不同的信息内容的服务。

1.2 个性化信息服务的特点

综上所述,个性化信息服务的根本就是尊重用户(或者说以用户为中心),研究用户的行为、兴趣、爱好和习惯,为用户搜索、组织、选择、推荐更具针对性的信息服务。通过对个性化信息服务特点的了解和掌握,可以使我们的服务更具针对性,从而提高用户对信息服务的满意度。具体地说,个性化信息服务具有以下特点:

(1)以用户为中心。所有的服务必须以方便用户、满足用户需求为前提。

(2)允许用户充分表达个性化需求,能够对用户需求行为进行挖掘。信息服务的系统不仅要提供友好界面,而且要方便用户交互,方便用户描述自己的需求,方便用户反馈对服务结果的评价。要能够了解用户的个人需求、习惯、爱好和兴趣,为其提供“量身定制”的个性化信息服务。

(3)服务方式更加灵活、多样。不仅要为用户提供更加准确的信息,而且还要能够按照用户指定的方式进行服务,如满足用户对信息的显示方式、提供结果的方式(纸质、电子版、网络版、电子邮件等)的要求,对服务时间的要求,对服务地点的要求等。

(4)能够主动将用户所需信息推送给用户。

2 个性化信息服务的类型

在个性化信息服务系统中,用户可以根据自己的需要,选择不同的信息资源;依据自己的习惯,指定信息的显示方式。如对信息检索结果的不同的元数据排序。对于一般用户可能更关心需求的主题,希望将论文的题名放在显著的位置,而对于一个专家,他可能更关注学科内某些专家的研究成果,希望将著者姓名放在最显著的位置等等。就现阶段的环境条件而言,个性化信息服务至少应包括个性化内容定制服务、个性化信息检索定制服务、个性化界面定制服务和个性化信息推荐服务四个方面。

2.1 个性化内容定制服务

个性化定制服务是有效获取用户需求信息的方法之一。它通过用户定制获取用户的个性化信息,从而理解用户的需求,为用户提供更为准确的信息服务,提高用户的满意度。同时通过与用户的直接或间接的沟通,改善与用户的关系,增加用户的忠诚度^[7]。

在个性化内容定制服务中,用户可以根据自己的兴趣和需求定制信息。定制的内容包括信息资源、界面和服务等等。个性化信息定制服务系统通常提供一些信息资源,让用户选择(定制)。例如在数字图书馆中,人们可以选择数据库资源、电子期刊、电子图书、相关网站、学科导航、搜索引擎等等。也有一些系统让用户指定资源地址,系统提供有关的更新通知等服务(这是一种简单的主动信息服务方式,系统能够及时主动将有关资源的更新信息推送给用户,提醒用户及时利用)。

另外,如安全与隐私保护也是个性化信息定制服务中需要注意的重要问题。安全包括用户私人信息的安全和系统管理的安全。隐私保护也是用户关心的问题,一旦用户的个人信息被泄露,用户就会失去对我们的信任,而远离我们的服务。隐私保护需要制定完善的隐私保护政策,提供隐私政策公示,并提供用户隐私公开的工具和运用保证隐私不外泄的保护技术等。

2.2 个性化信息检索定制服务

在数据库检索或网上信息的查询中,不同的用户由于其拥有的检索知识和所处的领域不同,往往其习惯也不同。初学者一般习惯用简单检索,专业人员习惯用高级检索。另外,不同用户可能用不同的词汇表达同一专业概念,不同用户对检索结果的选取原则和排序方法也可能不同。诸如此类,正是用户个性化的显著表现。因此,检索定制需要充分支持用户在检索策略、检索方法和检索结果处理上的个性化。根据初步分析,检索定制可包括^[8]:

(1)个人检索模板定制。用户可根据个人需要对所需信息的专业(用分类号或主题词、关键词确认)、应用目的(用文献类型区别,例如期刊论文、专利、会议文献等)、深度、语言、地域、数量、时域等进行定制,从而内在表达自己相对稳定的检索限制条件。用户在每次检索时也可以选择“屏蔽”(Overwrite)模板条件,灵活表达特殊要求。

(2)检索工具定制(选择数据库检索系统和搜索引擎)。已纳入资源定制。

(3)检索式表示方式定制。用户可自定义默认检索方式是简单或复杂方式,自定义布尔检索式中“与”、“或”、“非”的表示符号。系统还可为用户记录最近的特定数量的检索式,供用户选择使用。

(4)个人词表定制。可写入用户自定义的常用词汇的自定义规范词、通用限制词(检索时对所有检

索式均加上此检索词)、特殊限制词(对某一个或几个词使用的限制词)、以及同义词和近义词;当该用户使用这些常用词汇时,系统可根据个人词表调用相应词汇优化检索,或弹出相应词汇供用户选择使用。

(5) 检索结果处理定制。可对检索结果的相关度计算标准、输出格式、排序方式、重复记录整合方式、分类组织、下载格式、传送地址等进行定制。

(6) 检索历史分析定制。用户可要求系统动态追踪其检索过程,提取最常用的检索词和检索式,最经常选择的数据库、期刊、网站等资源,最经常排除的资源,相近用户最常用的检索词、检索式、数据库、期刊及网站等(但必须掩蔽“相近用户”实际身份)。当然,这种追踪必须事先得到用户的授权(包括是否同意作为“相近用户”被分析的授权),同时必须安全地保护这些历史数据。这些历史数据可用来自动修改用户的有关定制信息(例如定题选报记录)。

2.3 个性化界面定制服务

个性化界面定制服务是让用户根据自己的爱好选择桌面的显示方式,它包括界面的结构布局,显示颜色和显示内容的排列方式等。例如在一些网站中,让用户输入自己的个人信息,然后生成“我的主页”、“我的搜索引擎”、“我的图书馆”等等这类服务都属于个性化界面定制服务。

2.4 个性化信息推荐服务

通过网络信息的挖掘,了解用户的需求和兴趣,为用户提供个性化的实时信息推荐服务,也是因特网服务的一个发展方向。面向单个用户的实时个性化推荐服务有4种方法^[9]:

- (1) 在 Cache 内的协作推荐。
- (2) 利用聚类方法实现实时个性化推荐。
- (3) 利用分类方法实现实时个性化推荐。
- (4) 利用协作推荐的方法实现实时个性化推荐。

3 用户个性化需求的获取方法

对用户需求、兴趣、爱好的了解和获取是个性化信息服务具备针对性的关键,而对用户网络信息需求的分析、获取和管理是保证个性化信息服务实施的基础。对用户需求的获取,目前主要有两大类方法:一类是通过人-机交互模式获取用户的信息需求。这类方法的优点是可以准确地获得用户的需求信息,也是目前绝大多数检索系统所采用的方法。缺点是主动性差。对于以信息拉取技术为核心的信息

服务系统,系统一般提供一个与用户进行实时交互的检索界面,让用户表达需求。对于以信息推送技术为核心的信息服务系统,系统会提供给用户一个表达信息需求的定制表单,用户可以十分详细地填写自己的个性化需求信息。另一类是通过对用户访问记录的挖掘,获取用户的需求、兴趣和爱好等。这种方法的主动性强,一般用于个性化主动信息服务系统中。对用户需求行为的挖掘,目前有三种方法:一是通过 Cookie logs 挖掘;二是通过 Bookmark 进行网络信息过滤;三是利用智能 Agent 跟踪用户信息行为。

3.1 用户访问记录的挖掘

网络用户访问记录的挖掘主要是从 Web 的访问记录中抽取具有意义的模式。Web 中的每个服务器都保留了用户的访问日志(Web access log)等记录,保存了有关用户访问和交互的信息。分析这些数据可以用来了解用户的需求行为,为用户提供个性化的服务。目前 Web 使用记录挖掘技术和工具可以分为两大类:访问模式的追踪和个性化的使用记录的追踪。一般的访问模式追踪通过分析使用记录来了解用户的访问模式和倾向,以改进站点的组织结构;而个性化的使用记录追踪则倾向于分析个别用户的偏好,其目的是根据不同用户的访问模式,为用户提供个性化的定制服务^[10]。

用户访问挖掘针对用户对 Web 的访问记录,利用路径分析、分类聚类、关联规则和序列模式的发现等方法进行分析,寻找其中蕴涵的用户知识和模式。Web 访问信息挖掘的数据对象主要是日志数据。当用户访问 Web 服务器时,服务器方将会产生三种类型的日志文件:Server logs, Error logs 和 Cookie logs,记录用户的访问情况^[11]。

Cookie 是由互联网服务器产生的记号并由客户端持有,用于识别用户和用户的会话。通过 Cookie 分析,获取用户个性化需求、兴趣和爱好,目前还是一种较为有效的方法。但由于 Cookie 存储在客户计算机上,使用中会由于各种原因导致 Cookie 内容失真,因而通过对 Cookie 的挖掘获取用户的个性化需求,也有一定的局限性。

3.2 利用 Bookmark(书签)获取用户的个性化需求

当用户在因特网上用浏览器浏览主页时,经常会遇到自己需要的和喜欢的站点和主页,只要点一下浏览器的 Bookmark 菜单,就能将当前的站点存入 Bookmark 中。下次访问时,只要从 Bookmark 中取出

即可使用。用户可以对 Bookmark 进行多级管理,不同的目录反映用户不同的兴趣,这给对用户需求的分析提供了依据。

利用用户在浏览器上存储的 Bookmark 获得用户信息需求是一种有效的方法。在 Bookmark 中存储的信息往往是用户最关心的,需要记录下来以便以后再读。Bookmark 结构化的信息存储更能够确切地表达用户需求。我们可以利用用户对文章的评价来表达用户的信息需求。对存入 Bookmark 的某一推荐目录的文章,可以认为是用户喜欢的文章,作为学习中的正例;在其他目录的文章,或是经过推荐没有被选中的文章将作为学习的反例。根据向量空间法将正反例表示成向量,就可以利用机器学习的方法对新文章进行推荐^[12]。

利用 Bookmark 进行信息过滤有很多优势。用户可以导入自己已有的 Bookmark 文件或直接在浏览网络时通过客户端软件加入的新书签,并且可以对其进行各种编辑操作。用户可以明确地对文章的内容与自己的需求是否一致作出评价,通过反馈信息将使信息推荐更为准确。用户可以灵活地设置推荐的范围,系统对用户信息的推荐信息直接并入 Bookmark 结构中,完全符合用户的习惯。另外,系统使浏览器的 Bookmark 功能实现网络化,即利用网络服务器存储用户的 Bookmark 信息,这样,可以使用户在任何地方、各种平台上拥有一致的 Bookmark 信息。

3.3 利用智能 Agent 获取用户个性化需求

监视用户的信息查询过程的方法能自动获得用户的信息需求。其方法是在用户的终端上运行一个监视的信息代理(agent)。信息代理将用户在 web 浏览时的相关信息不断传送给远端的服务器,服务器将信息进行整理、组织,并从中分析出用户的信息偏好。

4 个性化信息服务模型的构建

要实现个性化信息服务,系统必须具备两方面的能力:一是构建个性化信息 Agent 模型,即将个性化信息从全局信息空间中分离出来;二是构建用户 Agent 模型,即跟踪用户行为,学习、记忆用户兴趣,通过描述用户的兴趣特征来建立个性化用户模型。此外,还需要有功能强大的网络信息搜索能力和友好的用户界面,这既是构建信息 Agent 和用户 Agent 的基础,也是实现个性化信息服务的保障。

4.1 个性化信息模型的构建方法

信息 Agent 具有多种特性,从不同的特性角度出

发,可以用不同的方法来构建个性化信息模型。这里主要介绍三种常用的方法^[13]。

(1) 信息 Agent 的自主学习方法。个性化服务的对象是一个个具体的用户,因而建立个性化信息模型时,信息 Agent 必须对用户感兴趣的信息类型有所了解,并以此作为信息模型构建的基准,通过不断的学习积累形成一个有规模的模型。信息 Agent 可以直接探求用户的信息兴趣和爱好,也可以从用户信息检索和利用行为中提取这个基准,然后由 Agent 执行一定的算法,将这些信息转化为知识存入用户信息模型中。信息 Agent 有两种不同的学习方法:一是关键词学习。这种学习方法是首先对搜索来的信息进行处理,将信息中的关键词按出现频率、出现位置提取出来,根据关键词的频率来确定唯一表征信息体的关键词,将具有相同表征关键词的信息聚类,并存入相应的个性化信息模型中。采用关键词学习方法提高了建立多个个性化信息模型的效率。二是字典学习。这是一种对相似概念的学习。在这种学习方法中,信息 Agent 为个性化信息模型提供的每个基准关键词创建一个字典,字典根据关键词的权值来确定其权限:搜索权和学习权。一个信息单元是否被学习首先得看其关键词的权值是否达到搜索权所规定的权值,否则首轮即被模型所淘汰。即使是过了搜索关,也只有当关键字的权值达到学习权时,该关键字才能被纳入字典用来表征个性化信息模型。信息 Agent 通过不断的学习,并在学习中修改字典,最终可得到一个关键字的相对稳定的字典。

(2) 信息过滤的方法。信息过滤系统是根据用户的信息需求对动态信息流进行过滤,把满足用户需求的信息传送给用户,可以提高获取信息的效率。信息过滤也是构建个性化信息模型的有效方法,它根据一定的标准将那些不相关的信息滤掉,仅保存与用户需求相关的信息。信息 Agent 根据自己的关键词域,提取几个权重比较大的关键词作为过滤的标准(E-profile),每个 E-profile 描述着一个方面的信息需求。E-profile 是否标准,可在实践中被检验。那些对信息需求描述不准确的 E-profile 个体在竞争中将被淘汰。反之则形成越来越大的规模,构成一个内容充实的个性化信息模型。常用的信息过滤方法有两种:基于关键词向量的过滤方法和基于文献集的信息过滤方法。基于关键词向量的信息过滤方法是一种相似过滤方法。其思想是比较表征信息单元的关键词与信息 Agent 给出标准关键词之间的相似

性,只有当相似性达到一定的度,该信息单元才被保留下来。确立关键词的相似度是这种过滤方法的核心。在此我们将信息单元的关键词与标准关键词都看作向量,其相似度就是这两个向量夹角的余弦值。

(3)基于多 Multi-Agent System 的合作方法。网络个性化服务不是面向单一用户的个性化服务,服务于每个用户的信息 Agent 之间并不是孤立的关系,它们通过相互合作构成一个 MAS。MAS 的合作求解能力远远超过单个 Agent,这是 MAS 产生的最直接的原因。通过 MAS 的合作,使每个用户拥有一个或多个 Agent,这些 Agent 推荐的信息在内容上都有相关性,克服了单个信息 Agent 功能不强的缺陷。多 Agent 之间的合作是通过多 Agent 协商协议来完成的。协商协议是 MAS 实现协同、协作、冲突消解和矛盾处理的关键环节。在 MAS 中,协商协议被描述成四元组: {目标 Agent,源 Agent,消息名,消息内容},这种四元组形式通过嵌套使用来扩大合作的范围。信息 Agent 之间一旦达成合作协议,多个信息 Agent 的个性化信息模型就构成一个逻辑体,每个信息 Agent 就将自己和伙伴 Agent 的信息文档推荐给相关用户,以充实用户信息需求的内容。

4.2 用户模型的构建方法

建立用户模型是实现个性化服务的另一个重要方面。用户模型是实现个性化服务的起点,用户模型不仅是个性化信息模型的构建依据,又是实现个性化服务的目标,个性化信息最终将被推向具有相应兴趣的用户。用户模型实际上就是对某一用户行为、兴趣倾向的描述,以确立他所需信息服务的类型。

在对用户的信息服务中,用户所处的领域不同,使用的信息服务系统不同,获取用户信息的方法不同,建模的方法也不同。通常情况下,通过与用户交互,获取用户信息来建立模型较为简单,较易实现。利用智能的方法,通过对用户需求行为(如用户浏览器中的 cookies 文件,用户的 bookmark 文件等)的分析,挖掘用户的需求、兴趣和爱好来建立模型较复杂,技术上也难以实现。这方面的研究还有待深入。对于隐式建模方法,用户模型的建设过程主要有四步^[14]:

(1)内容向量。内容向量是一个表示信息内容的向量,可以被用于向量空间模型成对的匹配。内容向量来源于初始用户的活动记录(以往信息利用行为),或用户的反馈等。一旦建立,就仅仅改变其

权值就可以了。

(2)一列当前活动的上下文(或用户以往信息行为的文档)。初始内容向量每个元素对应一些上下文。以后随用户的不同活动,每个用户的上下文就发生改变。

(3)一列当前活动的特征。

(4)一个用户关键词的集合。每一个关键词都有一定的权值,用于表征它对该用户的实际重要性。在用户反馈之后,关键词被动态插入、修改或删除。

用户模型建立后,对其更新也要遵循一定的规则。通常是根据用户的反馈信息相应地更新用户模型的不同部分。

个性化服务的关键是用户需求信息的获取。真正的用户动态需求信息的获取,必须利用智能代理技术。在这方面的研究,虽然取得了一定的成果,但距离实际应用或用户的要求还有较大的差距。

参考文献

- 1 薛蕊. 基于 web 数据库平台的图书馆个性化服务: MyLibrary. 图书情报工作, 2002(8)
- 2,13 张玉峰等. 基于 Agent 的个性化信息服务模型研究. 情报学报, 2001, 20(5)
- 3 史田华. 因特网个性化信息服务. 情报资料工作, 2002(1)
- 4 王悦. 新信息环境下图书馆的个性化信息服务刍议. 图书情报工作, 2002(8)
- 5 卢共平. 数字图书馆的个性化信息服务. 图书情报工作, 2002(8)
- 6 马献明等. 个性化网上信息代理技术的研究概述. 广西师范大学学报, 2000, 18(3)
- 7 黄晓斌. 网络文献的知识发现研究. 武汉大学博士学位论文, 2002
- 8,10 张晓林等. 基于 Web 的个性化信息服务机制. 现代图书情报技术, 2001(1)
- 9,11 田捷. 数字图书馆技术与应用. 北京: 科学出版社, 2002
- 12 卢增祥等. 利用 Bookmark 服务进行网络信息过滤. 软件学报, 2000, 11(4)
- 14 高文. 数字图书馆——原理与技术实现. 北京: 清华大学出版社, 2000

赵水森 洛阳师范学院图书馆馆长, 副研究馆员, 参加过多项省部级和国家级课题的研究. 通讯地址: 河南洛阳. 邮编 471022.

(来稿时间: 2002-12-18)