

●张树良 冷伏海

## Web 环境下个性化信息的获取和 个性化服务的实现

**摘 要** Web 环境下服务和用户都极具目标指向性与高度个性化,因此对任何一个服务方而言,对用户个性化信息的抽取均是不可避免的关键环节,关系到服务方对用户个性化信息的预知和占有直至其个性化服务的最终实现。基于外部链接、资源整合、情景关注系统和本体系统等 4 种不同层次是实现 Web 个性化服务的方式。个性化 Web 服务系统的实现不仅是 WIS 发展的要求,而且也是当前有效减轻用户负担、促进 WIS 应用拓展的理想途径。图 4。参考文献 5。

**关键词** 网络信息系统 个性化服务 信息获取 个性化信息

**分类号** G252

**ABSTRACT** In the Web environment, both services and users have clear orientation and high personalization. Therefore, the extraction of user's personalized information is a key factor for any information service providers. The four levels of external links, resource integration, situational awareness system and ontology system are the ways for the realization of Web personalized services. The realization of personalized Web service systems is not only a requirement of Web information systems, but also an ideal approach to alleviate user's burdens and promote the development of WIS applications. 4 figs. 5 refs.

**KEY WORDS** Web information system. Personalized service. Information acquisition. Personalized information.

**CLASS NUMBER** G252

Web 环境下服务与用户都极具目标指向性和高度个性化,一方面,用户对服务的需求与选择趋向于易用性与个人偏好的结合;另一方面,用户对服务的个性化需求使得服务更为细化和更具用户针对性。Web 环境下,服务方之间的竞争已经转向对用户个性化信息占有和预知程度的竞争,而对任何一个服务方而言,对用户个性化信息的抽取均是其不可避免的关键环节,关系到服务方对用户个性化信息的预知和占有直至其个性化服务的最终实现。

### 1 Web 环境下个性化信息的特征及其分类

Web 个性化信息具有以下特征:

(1)差异性。个性即意味着差异,因而差异性是个性化信息的最显著特性。用户个体或用户群之间由于其性别、职业、性格、行为习惯等等的不同必然导致其信息行为方式、兴趣等的不同,因而所产生的个性化信息必然存在差异。事实上,个性化信息的差异性是其最具商业竞争价值的特性。

(2)易变性。Web 资源中个性化信息的易变性主要由三方面因素所致:①用户的兴趣取向:用户的兴趣取向是导致用户个性化信息发生改变的主导因

素。用户的兴趣取向并不是一成不变的,它是随着用户的实际需求发生改变的。这种改变有两种情况:一是用户的兴趣主要方向不变而是围绕主要兴趣产生了外围或分支需求;二是用户的兴趣点完全改变,这往往同用户的阶段性需求有关。②服务模式:为适应 Web 发展和基于服务方自身商业竞争的需要,服务方不得不适时调整或改变其服务模式。服务方服务模式的调整 and 改变必然导致用户获取相应所需服务方式的改变进而导致用户个性化信息的改变。③Web 结构:Web 结构的改变将直接导致服务方服务模式和服务程序的变革,因而使得用户通过 Web 实现需求的行为方式发生改变最终使得其个性化信息发生改变。

(3)动态性。个性化信息的动态性体现在服务和用户随 Web 发展和时空变化双方状态的演变上。在不同 Web 层次和不同时空,由于双方状态的不同,因而用户的个性化信息也必然不同,这种不同在特定时间段上表现为连续性序列。

(4)情景敏感性。个性化信息是因时因地而异的,用户的行为方式、兴趣取向等都可能由于时间的改变、地点的转移而发生改变进而导致信息需求和信息

行为的改变。这种现象是短暂的、临时的、频繁发生的,但并不影响用户的主要兴趣取向。正是 Web 资源个性化信息的情景敏感性导致了对其抽取的复杂性。

另外,根据 Web 环境下个性化信息的记录与处理方式,可将 Web 资源中个性化信息分为静态信息和动态信息两大类。

(1)静态个性化信息。静态个性化信息是指所有以历史记录形式反映用户行为和历史状态的信息,如用户的登记资料、网络浏览记录等等。

(2)动态个性化信息。动态个性化信息是指实时的关于用户即时状态的信息,它通过同网络相联的跟踪用户实时状态的物理传感器等实时监测设备获得。

## 2 Web 环境下用户个性化信息的获取

近年来,随着 WWW 的迅猛发展,Web 应用集和数据密集的趋势愈发明显,由此导致用户个性化信息量巨大,如何有效获取用户个性化信息为 Web 环境下实现个性化服务创造条件成为备受关注的研究热点。

### 2.1 用户静态个性化信息的获取

(1)通过网络登记系统。一般情况下,在使用者要求作为某一特定服务的正式用户初次登录某一系统或申请某些特定服务时,网络登记系统都会要求用户填写用户资料,通常以包含用户一般特征,如姓名、性别、年龄、职业等等的数据表单的形式被存贮。因此有关用户一般特征的信息可以通过网络登记记录获取。作为用户个性化信息重要组成部分的用户一般特征信息对于用户分类、目标用户的选择、用户需求预测等都至关重要。

(2)通过网络浏览日志。用户的网络浏览行为会被 Web 系统以网络浏览日志的形式自动记录。网络浏览日志犹如行程记录将用户每一次网络浏览中所浏览过的网页、达到的网站、所进入过的网络结点或系统以列表的形式详细记录。通过网络浏览日志可以获取用户网络浏览行为习惯、需求偏好、兴趣取向以及用户兴趣取向变迁和需求转换等有关用户的特定特征的信息。它对于服务方掌握用户需求与行为特征、及时掌握用户兴趣取向变迁和需求转换,为用户提供准确及时的服务有重要意义。

(3)通过用户个性化网络信息行为管理系统。网络浏览日志虽记录了用户每一次用户的网络浏览行为,但这种记录是孤立的、机械的不能系统反映特定用户的需求与行为特征,特别是在当用户终端权限共享状态下,由于系统无法就某一特定用户予以区分和识别,因而网络日志记录内容是所有共享该终端用户的

网络浏览行为,在这种情况下网络日志所记录的用户个性化信息仅能作为服务方就个性化服务内容选择的参考,对于特定目标用户个性化信息的掌握和就如何为其提供个性化服务进行决策并无意义。而专门的用户个性化网络信息行为管理系统则可以解决此问题。

专门的用户个性化网络信息行为管理系统,如 PowerBookmarks<sup>[1]</sup>等,可以实现对特定用户的个性化网络信息行为进行全程跟踪和记录,并自动建立用户个性化信息行为档案。由于它对用户信息行为的记录是连续的和一对一的,所以可以准确反映特定用户的网络浏览行为习惯、需求偏好、兴趣取向以及用户兴趣取向变迁和需求转换等信息。

### 2.2 用户动态个性化信息的获取

(1)通过一般的实时监控设备。一般的实时监控设备包括各种物理传感器、状态监测仪、监控器以及能实时传递用户状态信息的各种物理设备,如 GPS、GPRS、GSM 等。它们同网络相联,一般处于网络信息系统的前端,通常作为用户个性化信息原始数据收集环节。

(2)通过高级的实时信息处理系统。高级实时信息处理系统用于多用户或用户群个性化信息收集和多任务处理,它内嵌由多个物理传感器和监控设备组成的传感器组、数据预处理系统以及多目标跟踪器,可以同时对不同目标用户进行实时跟踪获取实时状态信息<sup>[2]</sup>。

## 3 Web 环境下个性化服务的实现

面对资源、用户多样化和数据密集的 Web 环境,如何在准确确定用户个性化需求的基础上为其提供便捷的个性化服务成为 Web 服务方所要考虑的首要问题和面临的挑战。

### 3.1 通过外部链接

这是 Web 服务方向用户提供个性化服务最便捷的方式,即 Web 服务方并不实质性拥有满足用户个性化需求的资源而是仅仅向用户提供指向各种服务的 out-links,具体服务的获得需要用户通过这些链接自行查询和申请(在该 Web 系统之外用户个人则没有进入相应服务界面的权限),对于用户而言该服务方系统只是一个多种个性化服务的入口。

### 3.2 通过 Web 信息抽取与整合

该方法是在对目标用户(群)的个性化需求进行深入了解的基础上,首先从 Web 资源中将目标用户(群)所关心的信息抽取出来,然后将其整合到统一平台上提供给用户,这样就省去了用户通过系统所提

供的链接服务自行在系统外部寻求所需服务的努力,更便捷和人性化。具体实现过程是:系统向用户提供一个检索入口,根据用户的检索请求,系统通过信息抽取工具从 Web 资源中抽取相应的信息,对抽取到的信息进行特定处理和分类之后,将其存储在系统数据库中,这一过程可以在用户发出请求之前完成,也可以通过用户实时进行。系统提供给用户的界面是各种服务项目的详细列表,用户只需根据所需点击相应项目名称即可获取相应的信息和服务。

Web 信息抽取使用 Web 资源打包引擎(wrapper engine),它通过创建 Web 抽取资源数据库实现 Web 信息的抽取。在数据库构建过程中,打包引擎将每一个相关的抽取资源或信息赋予一个相应的说明文件。该文件包含了同抽取 Web 资源或信息实际属性值规则和其它相关规则相关联的实际关联框架。一旦任何相关资源被打包,数据将通过遵从说明文件所包含的规则 SQL 表达式的使用(用户终端)而被抽取。简言之,对于给定的合法的检索请求,打包引擎随即按照既定抽取规则从目标 Web 资源中抽取相应的数据。其抽取结果以 XML 或 HTML 表单形式表达<sup>[3]</sup>。通过 Web 信息抽取与整合方式实现对用户的个性化服务的构架如图 1 所示。

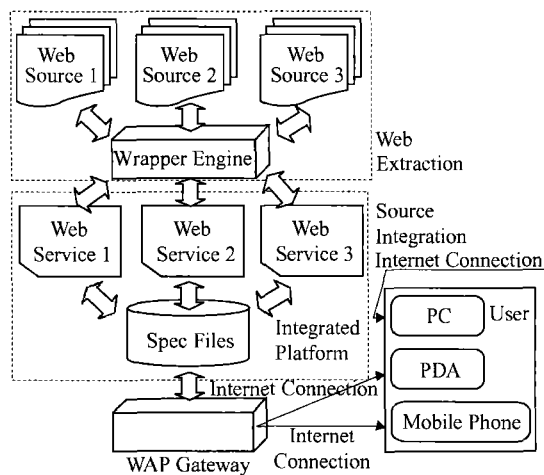


图 1 Web 资源整合个性化服务系统结构

(据 Nor Adnan Yahaya et al., 2005)

### 3.3 构建情景关注系统

如前所述,个性化信息的一个显著特征就是情景敏感性,情景敏感性是个性化信息动态、易变的集中反映,对于用户日益增长的即时服务需求而言,忽视个性化信息的情景敏感性将导致服务方不能实时掌握反映用户即时状态的情景信息,因而无法满足用户的即

时需求。因为用户所处的具体环境和状态不同,其需求也将会有所不同,比如,如果当前某手机厂商的某一目标用户正在一家食品店,而如果该厂商忽视其“在食品店”的情景信息,即使此时立刻为其提供有关手机的最新信息,该用户由于没有对手机的即时需求,所以服务方的信息提供是失败的。因此,对服务方而言,用户的个性化情景信息不仅关系到其对有即时服务需求的目标用户的确定,而且直接关系到其所提供服务的质量与成败。只有构建实时关注用户状态的情景关注系统才能满足用户的个性化即时需求。

情景关注系统是建立在实时信息处理系统和复杂数据库之上的 WIS 系统,由于其实时更新和即时服务的特性,因而被列为 3G - WIS 发展的目标之一。其语义模型要素有:

story:即用户同服务方发生关系的整个过程。它说明了发生了什么事件、事件发生的原因、地点、发起者、如何发生以及事件发生的时间。

scenario:即上述整个过程的各个环节,整个事件即 story 由诸 scenario 组成,而所有 scenario 的整合便形成 story space。

scene:即 customer-WIS 交流、对话(dialogue)的场所。每一个 scene 由其 scene-ID 来区分,并且每一个 scene 代表一个同特定情景相对应的说明。其中 dialogue 由 dialogue step 说明和定义。

media object:即中介实体,每一个 scene 由若干中介实体所支撑。中介实体是中介类型(media type)的实例。一个中介类型的核心由基于一定数据库架构的 view 所定义,它包括一个 view 架构和一个定义表达式。

context:情景(context)通常在数据库的实体组中被定义。为使 context 的表达更为完整和合理,通常要使用 context 词典。一个情景关注 WIS 一般有以下四种情景(context):用户的个性化情景(包括同具体服务相关的用户的全部细节)、服务方 WIS 情景(包括服务方的目的、网站主题、网站任务与特性以及访问者所在场合与意图)、开发方 WIS 情景(包括所有可能的开发环境、信息系统、事件空间(the story space)、场景(scenes)、交流过程(dialogue steps)、主体(roles)、权限(rights)以及具体任务(task))和 WIS 情景(包括被当前用户使用的 WIS 的历史和现状、由相应的情景所确定的 WIS 当前环境、由实际参与者所确定的用户群、WIS 的目标、细节与规则)<sup>[4]</sup>。系统的语义模型结构如图 2 所示。

图 3 则给出了一个情景关注系统的应用模型。

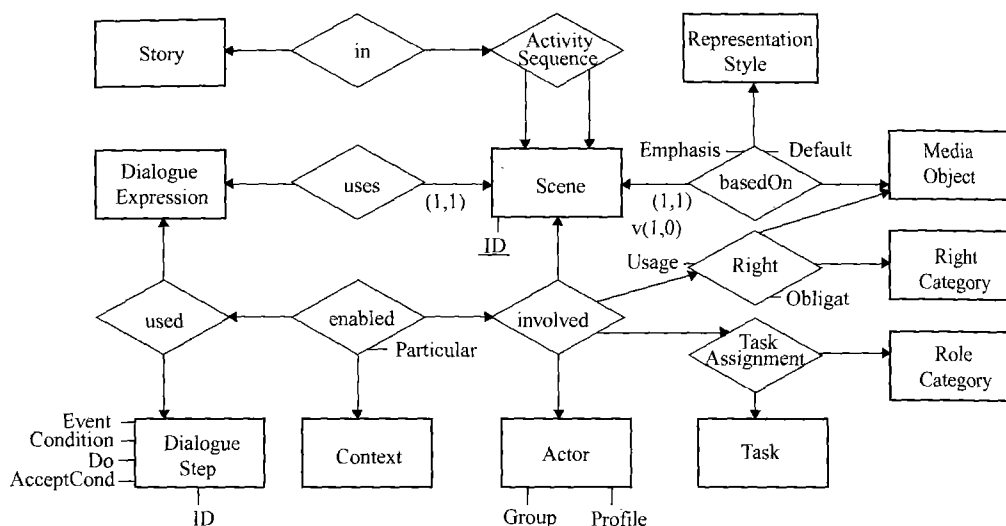


图2 情景关注系统语义模型结构  
(据 Aleksander Binemann-Zdanowicz et al., 2004)

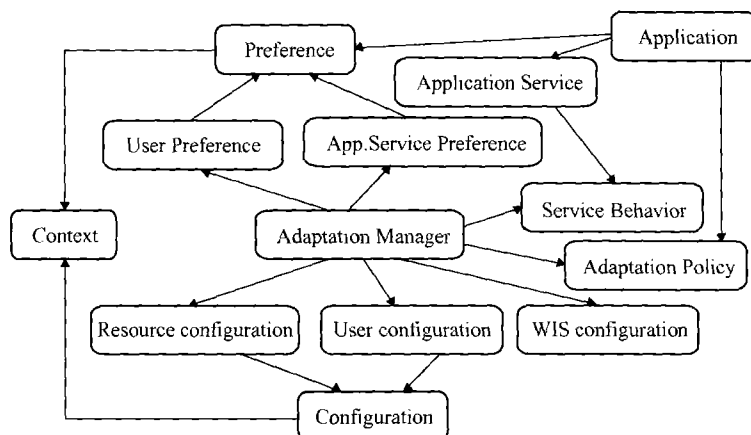


图3 情景关注系统应用模型  
(据 Soo-Joong Ghim et al., 2004)

### 3.4 基于本体(ontology)的个性化服务系统

面对 Web 环境下应用集中和数据密集给 Web service system 开发带来的挑战,基于本体的个性化服务系统的优势在于<sup>[5]</sup>:

① 本体能够如实反映实体的现实状态及其关系,因而使用本体可以更加清晰准确地描述和表达用户的个性化需求;

② 本体同时以人可理解的方式和机器可处理的方式对个性化服务所涉及的领域知识进行描述,同时形式本体能够自动检测数据的一致性从而提高系统的可靠性;

③ 在个性化服务所涉及的特定领域内,实体的定义、属性及其关系一经确定即能够被 WIS 的其它构成要素所共享和重复利用。

基于本体的个性化服务系统涉及两大本体:领域

本体和行为本体,前者定义和描述了一个个性化服务所涉及的特定领域内的诸实体要素及其之间的相互关系,后者则定义和描述了服务方(系统)同用户之间的行为要素及其之间的相互关系。

在 Web 环境下构建基于本体的个性化服务系统,除了使用 WIS 开发工具构建本体外,还有一个关键的基础环节是 Web 资源元数据的抽取,本体的构建即是以抽取获得的目标 Web 资源元数据为基础的。对 Web 元数据抽取包含两个环节:首先利用 Web 资源搜索工具 spider 收集以各种不同文件类型(html、xml、PDF、JPEG 等等)所表达的 Web 资源;然后元数据抽取工具从所获取资源中对元数据进行抽取并将抽取结果以 RDF 格式存储在元数据库中。整个系统结构如图 4 所示:

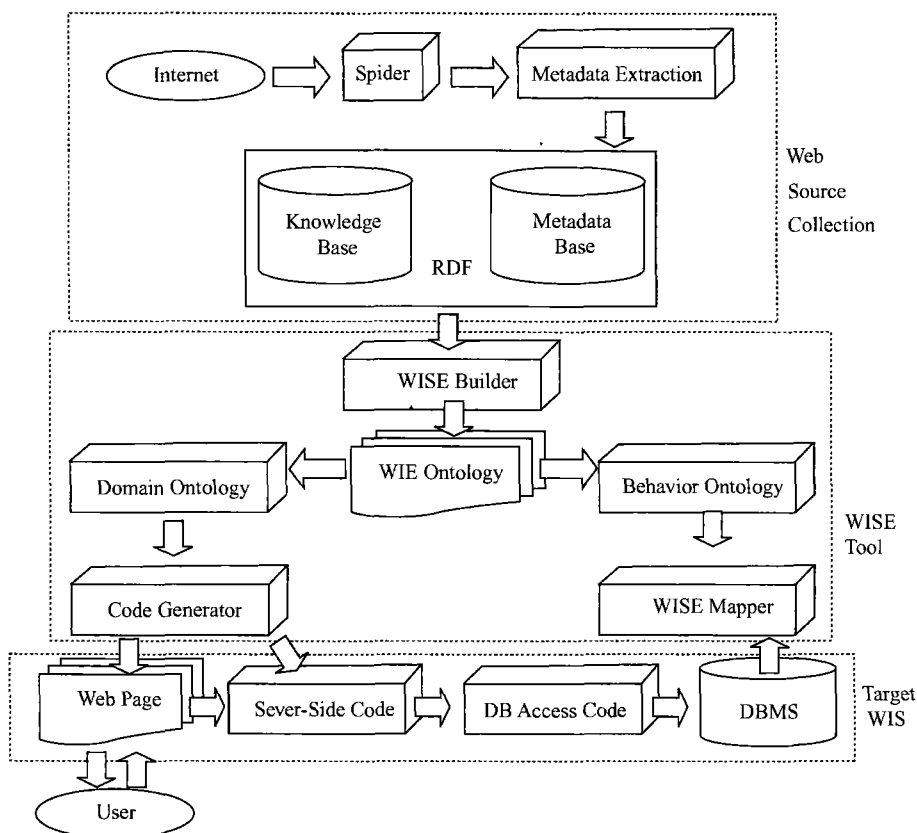


图4 基于本体的个性化服务系统结构

(据 Ting Wang et al., 2004; Lv-an Tang et al., 2006)

Web 环境下应用集中与数据密集的现实导致用户对适应自身的个性化服务需求愈来愈迫切,如何在跟踪用户个性化需求和全面准确收集用户个性化信息的基础上构建以用户为中心的个性化 Web 服务系统成为 WIS 开发和应用所面临的新挑战。个性化 Web 服务系统的实现不仅是 WIS 发展的要求,而且也是当前有效减轻用户负担促进 WIS 应用拓展的理想途经。本文旨在在对用户个性化信息及其收集方法进行系统研究的基础上,结合目前 WIS 开发与应用现状建设性地提出个性化 Web 服务系统实现的几种方式,为当前和今后个性化 Web 服务系统的实际开发和应用提供有益的思路。

## 参考文献

- 1 Wen-Syan Li, Quoc Vu, Divakant Agrawal et al. PowerBookmarks: A System for Personalizable Web Information Organization, Sharing, and Management. [2006-04-10]. <http://www8.org/w8-papers/3b-web-doc/power/power.pdf> (Accessed Apr. 10, 2006)
- 2 宋安军, 胡保生, 彭勤科. 实时信息处理系统体系结构的

研究. 计算机科学, 2004(9A)

- 3 Nor Adnan Yahaya, Goh Poh Gin, Chan Wai Choon. Developing Innovative Web Information Systems Through the Use of Web Data Extraction Technology. 19th International Conference on Advanced Information Networking and Applications (AINA'05)(2): 752-757
- 4 Aleksander Binemann-Zdanowicz, Roland Kascchek, Klaus-Dieter Schewel et al. Context-Aware Web Information Systems. Proceedings of the First Asian-Pacific Conference on Conceptual Modelling. 2004(31): 37-48
- 5 Lv-an Tang, Hongyan Li, Baojun Qiu et al. WISE: A Prototype for Ontology Driven Development of Web Information Systems. 8th Asia-Pacific Web Conference (APWeb2006), Frontiers of WWW Research and Development, LNCS3841. 2006(1): 1163-1167

张树良 中国科学院文献情报中心, 中国科学院研究生院情报学专业 2005 级博士研究生。通讯地址: 北京市海淀区中关村北四环西路 33 号。邮编 100080。

冷伏海 教授, 中国科学院文献情报中心博士生导师、情报研究部副主任。通讯地址同上。

(来稿时间: 2006-09-27)