

Internet 上的个性化信息服务

中国人民大学信息学院 丁浩
中南财经大学信息系 林云

在浩如烟海的混乱的信息世界中，人们逐渐迷失了方向。于是人们开始寻找一种有效的方式，以带领人们在信息海洋中确认正确的方向，以充分利用这些宝贵的资源。个性化信息服务是一个非常好的出路。在张扬个性，倡导创造性的现代信息社会，个性化服务更是倍受人们的关注和青睐。那么究竟什么是个性化信息服务呢？对于这个概念，我们通过逐层的分析，对其进行揭示。

个性化信息服务的概念

“某某化”是指使事物具有某种属性或某种趋势。因此，个性化的含义是使事物具有个性，或使其个性凸显。这里包含了两层含义，其一，个性是需要经过培养而逐步形成的，这个过程我们可以称之为使个体个性化的一个过程；其二，个体总是具有一定的个性的，让这种个性得到别人的了解、认可，并在一定的空间中得以体现、展示，是每个个体都拥有的潜在需求，这个过程我们也称之为个性化的过程。

下面我们来看一看什么是个性化信息。这个概念可以从两个角度分析。第一，个性化信息是指反映人类个性特性的一切信息，这些信息包括了这个个体的一些各种属性的描述；第二，个性化信息是指由人类个体特性所决定的其对信息的需求的一种信息组合，也就是由人类个性对信息需求的决定关系而产生的一系列对个体有用的信息。

个性化信息服务，既是一种个性化服务，又是一种信息服务，因此，我们可以根据前面阐述的两种个性化过程和两类个性化信息，找到个性化信息服务的定位。个性化信息服务首先应该能够满足用户的个体信息需求的一种服务，即根据用户提出的明确要求提供信息服务，或通过对用户个性、使用习惯的分析而主动地向用户提供其可能需要的信息的服务。其次，个性化信息服务也应该成为用户展现自我，宣传个性的一个窗口。具有一定个性的个体通常都有表现自我，表达自我，让他人了解自己的愿望。最后，个性化信息服务应该是一种培养个

性,引导需求的服务,这样可以帮助个体培养个性、发现个性,引导需求,促进社会的多样性和多元化发展。

个性化信息服务系统的一般模式

个性化信息服务系统本身也是一个信息系统,因此,个性化信息服务的基本模式 and 设计方法也应该是信息系统的模式和方法。从信息系统工程的角度来说,信息系统一般分为流水线式和工具箱两种,前者常见于业务信息管理系统和管理信息系统等简单的小型信息管理系统中。后者的系统结构常见于 DSS 系统以及一些 OA 系统中,这些系统的功能灵活多变,需要根据使用者的需求随时进行组合和变更。这种系统的动作、功能和系统扩展都是以用户需求为中心的,其主要特点是‘随用随成长’。是一个开放型的系统。

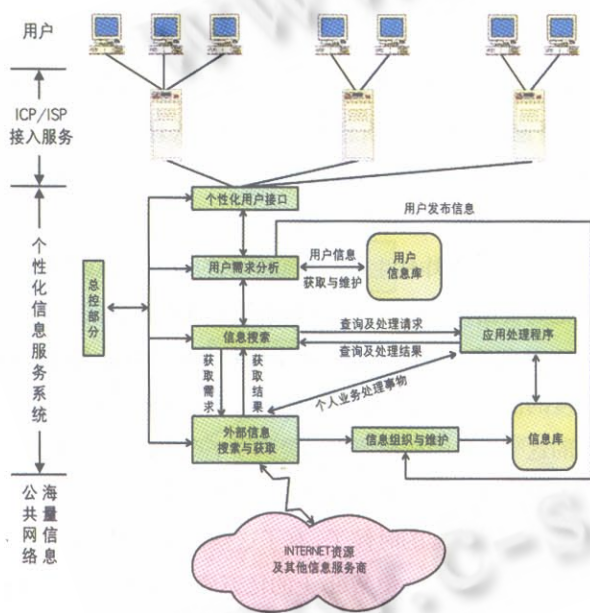


图 1 个性化信息服务系统框架

根据 INTERNET 上信息的特点,以及个性化信息服务的特殊性,我们认为工具箱式的信息系统模式更适合于这种服务。原因在于个性化信息服务系统是以用户需求为中心,而用户的需求无论在数量上还是内容上都是不断变化的。另外,个性化信息服务系统对数据的维护相对简单,获取到的信息本身都无需做加工处理,只需要对元数据进行处理,以便于系统的查询、用户需求的匹配。因此,对于个性化信息服务系统,我们需要的系统模式和工具箱式结构有一定的类似性,但又不完全一样,因为我们的系

统结构中,并不是完全以一个模块为核心的。因此,我们把下面给出的结构称为一种准工具箱式模式。下面我们根据个性化信息服务系统的特殊性,在工具箱式结构的基础上,给出一个较为一般的系统结构(见图 1)及各部分的功能。

其中各部分的任务、功能及可使用技术包括以下一些内容。

1. 个性化用户接口

根据用户个人信息及其需要的数据信息有效地组织成页面,并传送到用户的浏览器;准确有效地获取用户的输入信息,并记录用户动作,以供用户个性分析。

2. 用户需求分析

接收用户请求;通过对用户使用习惯和需求调查数据对用户个性和需求进行分析,调整系统服务角度和内容;维护用户信息数据。

3. 用户信息库

主要包括有关用户个性特征的信息及元信息。个性特征数据结构的设计是这一部分的关键问题,决定了个性化信息服务系统的服务好坏。

4. 信息搜索

是系统中信息流控制的中心,其主要功能是接收上一部分传递过来的用户信息需求后,区分出需求的是已存在信息还是未存在信息,前者到本地信息库查询,后者将需求传送到外部信息搜索与获取部分,以便到 INTERNET 上获取相应信息。

5. 外部信息搜索与获取

获取本系统没有的用户需求的信息;与外界其他信息服务商的联合品牌服务,以为用户提供多样化的个人业务信息处理服务。

6. 信息的组织与维护

这里的信息指的是存储在信息库里的能够为用户查询的信息。这部分信息的组织与维护工作关系到系统运行效率和对用户需求的满足程度。其中主要包括信息的格式化处理、存储和维护。

7. 信息库

这部分主要存储系统所维护的已有的有可能成为用户需求的数据。此处信息的查找和分类比较重要,元数据的维护是工作重点,包括数据的存储信息、分类信息、被访问特征等。

8. 应用处理程序

根据用户需求或所提取的信息做相应的处理工作。

包括信息的再生性处理和与用户个人业务相关的个人事务处理。

9. 总控部分

用于各模块之间的协调工作。上述个性化信息服务系统的一般模式中,各个子系统之间几乎都有交互活动发生,要有有效的控制这些交互动作,严格控制各模块对数据库部分的操作,就需要一个总控模块来实现这一控制功能。

当然,我们上面所给出的结构应该说是一个个性化信息系统应有的几个基本部分及其一种组合方式,不同的系统目标不同,也会有不同的解决方案,以更适合本系统的特点。另外不同系统各部分及部分间的组合关系的复杂程度也不同,需要具体问题具体分析。而且很多问题都是非确定性的,是需要在实际工作中解决的,不一定能在理论上提取出太多共性来。但总的来说,综合性服务需要更强大的信息管理和职能程度,在设计的初期,系统研制人员可以将综合性服务分解成几个方面的领域服务,将几个部分衔接在一起的工作单独分离出来对待,这样可以相对降低系统研制的复杂性。

个性化信息服务系统存在的问题

1. 信息搜索

在INTERNET上进行搜索信息目前主要的方法有两种。一种比较简单的方案是按照现有的搜索引擎中常用的传统方法,即根据原始资料提供者向搜索引擎等大的信息服务商提交的索引信息来获取该信息的链接。这种方法比较适用于大型的信息服务商,如YAHOO、INFOSEEK等等,而且能提供的查询方式有限,对需求的满足精度不高,对智能化查询的满足程度相对较低。另外一种方法是,使用智能代理技术,在INTERNET上搜索所需信息。目前的主要浏览器和信息检索工具还没有智能搜索功能。例如在Yahoo!、Altavista等这些类似的站点上,当用户仅输入“LIRA”这个词,想要了解“LIRA”这种具有学习功能的信息搜索程序的有关资料时,他却得到了意大利货币“LIRA”方面的信息。当前在智能信息搜索方面的研究方向主要是将机器学习与经典的信息检索技术相结合。例如:LIRA、NewsWeeder、Webwatcher、WebDoggie、InfoFinder、Inquery、Letiza、firefly、Syskill&Webert、DICA、Remembrance Agent、ILA以及ShopBot等系统。

2. 信息分类

分类问题涉及两个方面。一是系统内部对信息的分

类。我们说信息库中存储的大量信息必然需要一种分类,以便于信息的管理和查询。这里应该采取一些目前INTERNET上比较流行的分类方式,例如YAHOO、EXCITE、INFOSEEK等采取的办法,他们所采取的分类方式比较类似于图书管理中的分类方式,涉及面比较广,通用性比较强。二是用户的个性化分类。每个用户对信息所属类型的理解不同,那么他们需要的信息分类方式也不尽相同,因此,信息服务系统提供给用户的应该是一种可以由用户自己决定的分类。

3. 站点的信息组织标准化问题

标准无疑是计算机技术得以发展的一个决定性条件:网络的爆炸性扩展,计算机的普遍存在几乎都是标准的结果。然而,我们目前存在的标准还远远不够,换言之,在许多方面我们还需要更多的标准。Web就是一个最好的例子。TCP、HTTP等机制控制着点到点的数据穿梭,HTML、XML提供了单个页面的组织结构标准(至少在理论上是可以做到的),而这两者之间却存在着一个大的漏洞——站点组织的标准。实际应用中一直没有一种公认的、普遍接受的站点信息组织的标准,其直接后果就是现在的热火朝天的无控制的一片混乱——WEB上最困难的工作就是上面寻找你想要的东西。

Web的爆炸性生长给我们带来的是成百万的站点和上亿的页面,却没有一个逻辑系统来从中查找任何东西。Web从贫瘠乏味的千篇一律中拯救我们的代价却是一种极度的混乱。事实上我们完全可以避免这种局面。上面给出的这个粗略的标准为我们提供了一个灵活的折衷方案,使得创造性和同一性可以稳定一致的“和平共处”。现在是考虑类似标准的时候了,有了这些标准,我们才能填补Web组织的漏洞,使Web上的亿万资源从无序中走出来为用户服务。

4. 信息汇集者与信息提供者的角色分配

随着web出版者的不断增加,web上的信息会变得更加纷繁杂乱,人们会因此而逐渐减少网上冲浪,而只去查看一些品牌站点。这样那些有足够力量的品牌站点就会转向保持站点内容的新鲜以刺激重复访问。如此下去,我们的目光就全被束缚在那些站点强加于我们的内容和商业广告上了。因此说,现在INTERNET上缺少的不是信息,而是对信息的组织,是一种能够帮助用户在信息海洋中找到自己目的地的信息服务。而信息是这一服务的基础和出发点,因此,获取、收集以及维护这种信息的服务也是必不可少的。由此,我们可以得到一个INTERNET上信息

服务的两层模式，如图3。

这个结构中的第一层目前已经十分多见，这一层在数量上已经趋于饱和（中国市场除外），下一步需要做的主要是对信息的有效组织，也就是我们上一个问题谈到组织标准问题。结构中的第二层是目前信息服务的一个新的突破口，具备极大的市场需求，前景十分明朗。实力强劲的信息服务商可以纵贯这两个服务层次，对自己的信息组织和存储做进一步的导航式加工，其在技术上是占有一定优势的。相对较小的信息服务商，就可以从第二个层次做起，因为这里需要的更多的是新概念、新模型和新技术，而不是大量的资源。

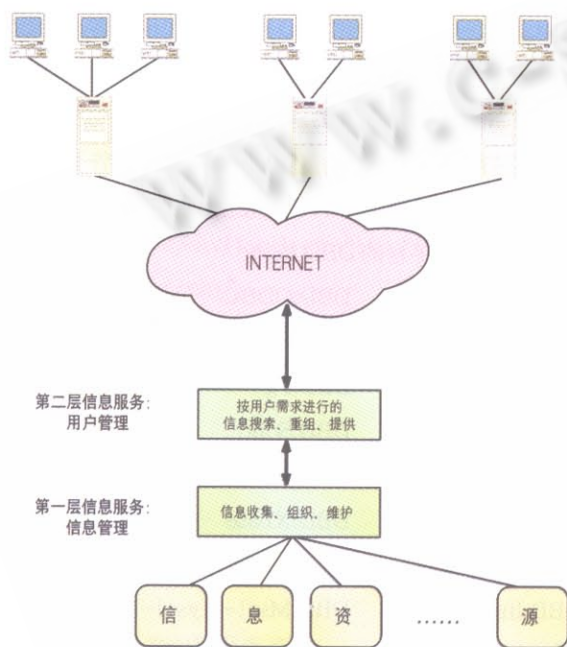


图3 信息服务的两层结构

对下一代个性化信息服务的展望

个性化WEB信息服务是个性化信息服务中最常见也是目前比较流行的形式之一。早期版本的个性化WEB信息服务主要由新闻剪裁、股票报价和目录推荐等内容组成，这些内容都是通过浏览器的cookie文件和一些信息服务商提供的智能后端工作实现的，这些信息中介通过一个综合的固定的视角提供过滤信息，并以此为读者创造价值。第二代的服务，是为用户建造创立和管理自己的信息或兴趣群组。

而下一代的个性化WEB信息服务将向更加深入的方向发展——将WEB扩展为真正的个人信息，并提供应用服务以帮助用户理解和管理这些信息。在竞争中获胜的可

能会是那些能够以对用户真正有意义的使用价值来抵消其对用户隐私侵犯的个性化WEB信息服务商。下一代个性化服务不再仅仅停留于大众化的娱乐性或专业性资讯类信息的提供，而是把用户的许多个人业务处理也加入了服务的范畴。■