

嵌入式桌面信息服务及其系统实现

乐小虬 李宇 张晓林 张智雄 李春旺

(中国科学院国家科学图书馆, 北京, 100080)

[摘要]: 在用户操作现场感知用户需求, 采用情景驱动, 将科技文献服务即时嵌入到用户的工作流程之中。介绍了这种嵌入式桌面信息服务的基本思想及系统设计、实现方法。

[关键词] 嵌入式服务, 桌面工具, Web2.0 桌面, 信息检索

[分类号] TP393

Approaches to implement Services-embedded Desktop Information Tools

Le Xiaoqiu Li Yu Zhang Xiaolin Zhang Zhixiong Li Chunwang
(National Science Library, CAS, Beijing 100080, China)

[Abstract] This paper introduces a services-embedded desktop information system, which put science and technology literature services to desktop by tracing user's operation and his ongoing workflow scene. While a string is selected by mouse, the retrieval results pop up. The paper also provides design ideas and implement approaches to the system.

[Keywords] Embedded Service Desktop Tools Web2.0 Desktop Information Retrieval

1 研究背景

为用户提供准确、方便、快捷的信息服务是众多信息服务机构的共同目标。桌面系统是一种以用户的实际操作环境为中心、以个体的应用需求为主导、以提供个性化服务为己任的软件系统。随着信息技术的迅猛发展, 尤其是搜索引擎技术、Web2.0 技术的广泛应用, 桌面系统日愈成为许多重量级信息技术企业争夺的焦点。近 2 年来, 微软、GOOGLE、YAHOO、百度、中搜等先后推出了不同形式的桌面搜索工具^[1], 一些概念型“Web2.0 桌面”“i桌面(网络桌面)”也开始浮出水面^[2,3]。这些桌面搜索工具充分利用了搜索引擎自身的资源体系, 以网络资源搜索和本地资源搜索为主体, 辅以其它常用功能(如: 即时信息服务、病毒防控、小秘书等), 在个体环境中形成一个相对完善的软件服务体系。

尽管如此, 桌面系统离用户的实际需求仍然有较大距离。一个重要原因就是这些系统多以独立系统的方式存在, 用户在实际工作时如果想要获取某种信息或服务不得不开“现场”, 进入或登录到其它服务系统中进行获取, 导致当前工作流程的中断以及时间成本的增加。

嵌入式服务是文献情报机构的基本服务模式^[4], 它强调将信息资源和服务直接“到桌面、进现场”, 有机融入用户工作过程, 随时随地可被利用^[5]。就桌面系统而言, “到桌面”就是到具体的应用软件系统中; “进现场”就是进入软件的操作流程之中。为此, 本文对嵌入式桌面信息服务做如下定义:

以 Web 资源、本机资源、专题资源为基础, 以当前操作的软件系统为现场, 在不干扰、不中断用户当前工作状态和行为的前提下, 将用户想要的资讯即时呈现于现场。

为了实现这种服务, 嵌入式桌面信息服务工具(以下简称**桌面工具**)一方面需要感知用户操作现场; 一方面还需具备强大的信息获取、组织、检索功能。同时, 能依应用的不同进行功能的组合和裁减, 以满足不同用户个性化的需求。另外, 桌面工具本身应处于“侍服”

状态，远离用户的桌面工作视线，只有当用户需要时才浮出视线，做到“呼之即来，挥之即去”。

为此，本文尝试以嵌入式桌面信息服务的方式，将系统提供的服务即时融入到用户的工作流程之中，力求使用户在最短的时间内“所要即所得”。以下具体介绍其设计、实现方法。

2 系统框架及基本流程

桌面工具的总体框架如图 1 所示。该框架以用户的应用环境为中心，将信息的获取（感知需求、接入服务等）、加工、服务流程有机衔接起来，构成一个较为完整的信息处理体系。系统可划分为五大功能模块：

2.1 需求感知

主要是获取用户需求。包括实时检测用户的应用场景、捕获用户的操作及焦点词串、分析用户操作日志、获取用户资源使用权限、记录用户使用偏好（相关设置）以及对焦点文本中命名实体进行智能识别等。

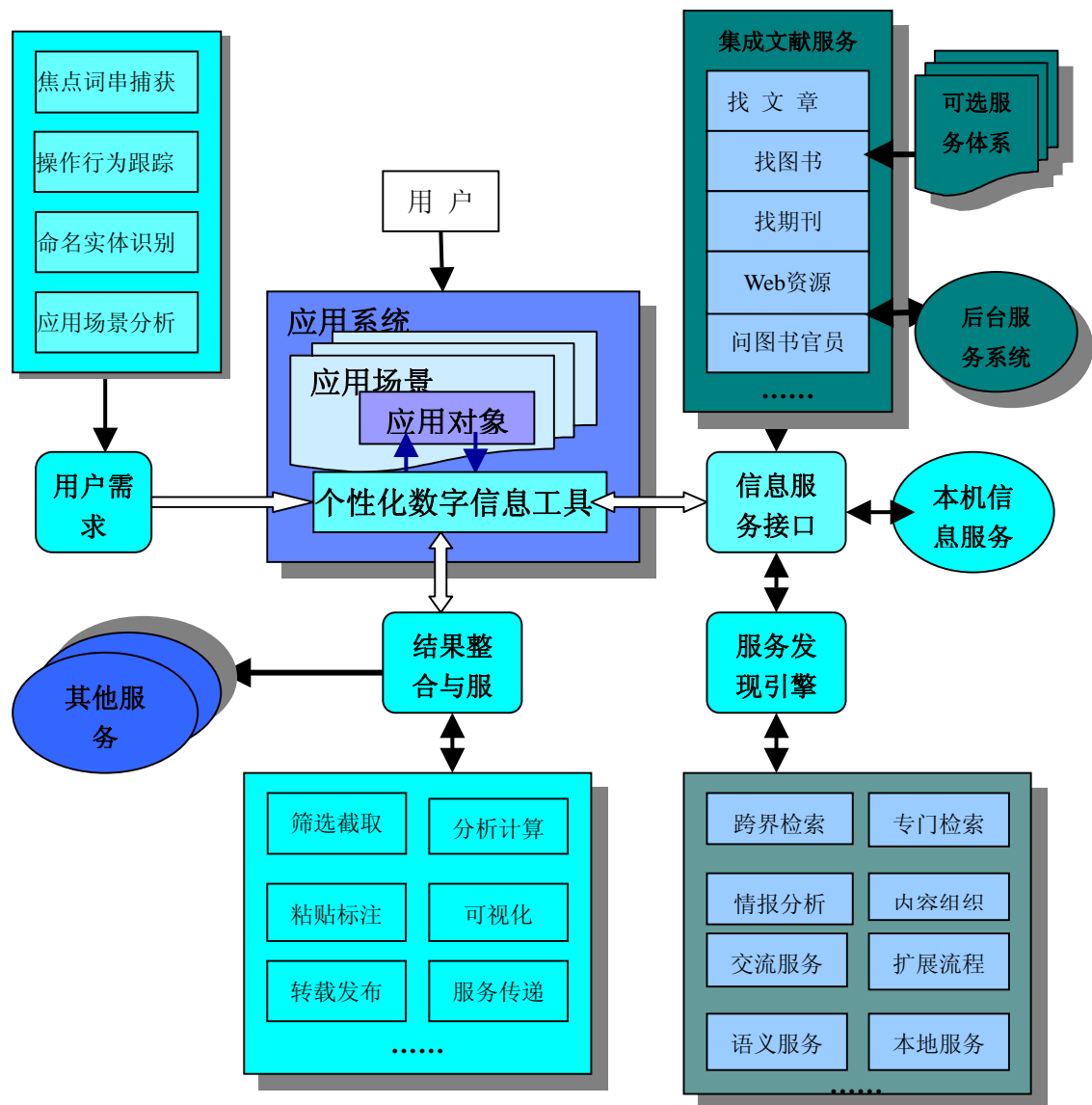


图 1 系统总体框架

2.2 集成文献检索

提供文献的查找和传递功能。利用国家科学图书馆相关系统中的数据资源（如 CSDL、OPAC、NSTL 等）以及购买的数据库，为用户提供授权范围内的中西文全文资料查找、期刊查找、图书（包括馆藏目录）查找以及学位论文、会议文献、科技报告、标准文献、专利文献的查找等功能。

2.3 Web 资源及服务发现

自动或半自动搜索互联网中相关服务接口，提供 Web 资源搜索及可配置服务的接入功能。分析相关 web 站点提供的 API，确定分类并存储于元数据库中。

2.4 本机资源检索

对用户指定目录下的本机文档进行全文检索。

2.5 个性信息管理与组织

管理用户个性化信息的设置、请求、推送、存储、编辑等功能。完成返回结果后处理，包括结果的转换、截取、重组、标注、引用、反显、转发、激发应用等。将最终结果呈现于桌面。

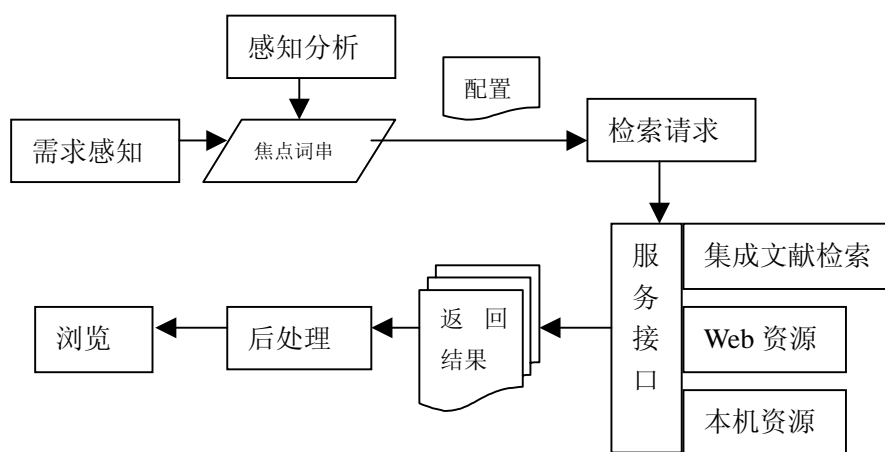


图 2 系统基本处理流程

按照信息的获取、加工、服务的处理顺序，系统的基本流程（图 2）可简单描述为：首先，需求感知模块实时捕获用户操作现场焦点词串，如果成功获取则在焦点位置提供服务窗口。如用户选取了某项服务，则进行感知分析，对其关心的词串进行中英文识别、命名实体识别（如人名、地名、术语、机构名称）等处理，同时记录日志供偏好评估。然后根据配置信息构造检索请求，并向相应的服务接口发送。获取返回结果后，先对其进行筛选、转换、截取、反显等后处理，然后重新构造显示形式，供用户浏览、保存和发布服务。

3 软件架构

系统总体上采用模块式多层 C/S 结构（见图 3）。从逻辑上看，Client 端由应用层、模块接口层、系统层构成；Server 端由接口服务层、应用服务层、数据层构成。Client 端与 Server 端间的通信采用 TCP/IP 协议。

在 Client 端，应用层由托盘控制工具、划词嵌入工具以及 IE Toolbar 工具组成。这三者之间的关系为：①划词工具不单独存在，它作为嵌入工具，主要是解决现场即时操作问题。②托盘程序和 IE Toolbar 均可单独存在，也可和划词工具捆绑，它们都有控制操作和相应的服务功能，是同一种系统功能的不同表现形式，相当于不同的操作平台。③在处理单元方面，划词工具侧重于词串，IE Toolbar 侧重于网页。模块间的接口采用类库、DLL、COM 等方式，操作系统为 Windows 平台。

在 Server 端，数据层由国家科学图书馆相关系统自建数据库、购买数据库、元数据库、

知识库构成。在数据层之上是应用服务层，建有众多应用系统（如 CSDL、OPAC、NSTL、参考咨询、学位论文系统等），这些系统不但直接向外界提供服务，还提供软件检索接口（如

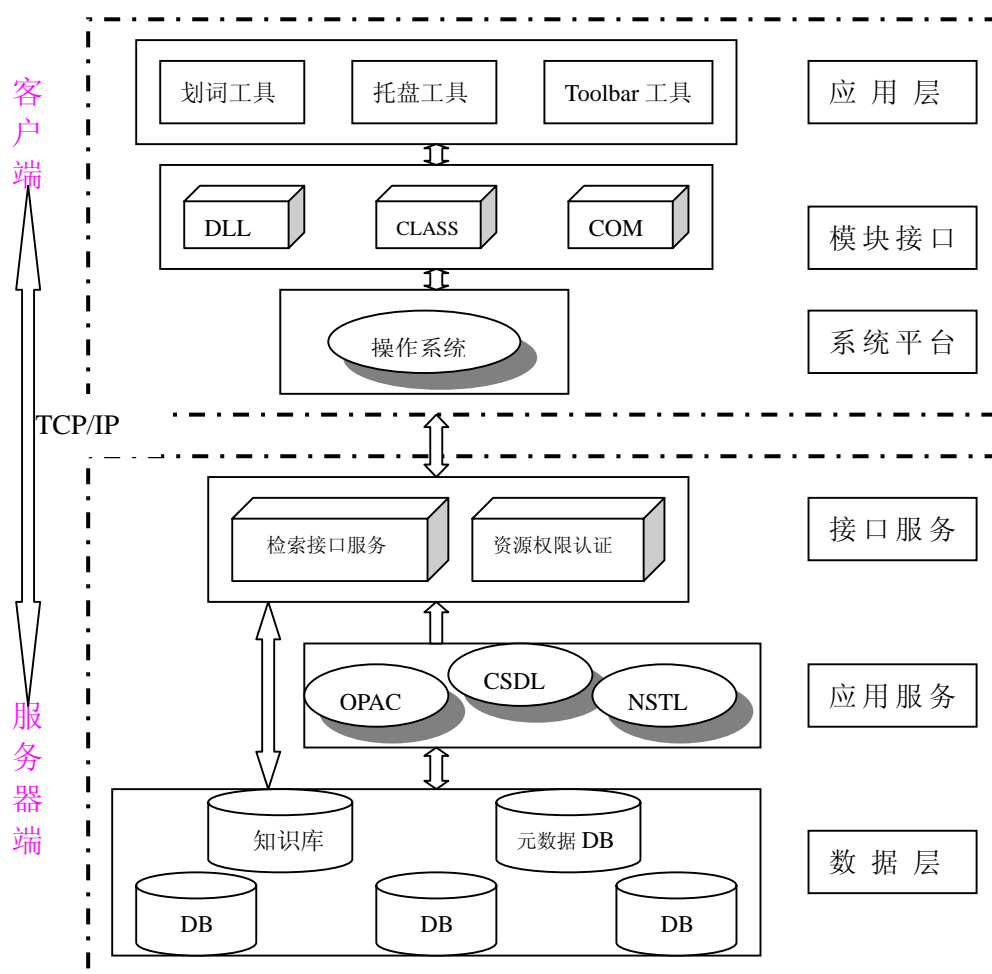


图3 桌面工具软件架构

URL 接口、SRU、SRW 等)。在应用服务层之上是接口服务层，该层是客户端与服务器端连接的枢纽，它管理应用服务层提供的检索接口、资源使用权限认证以及用户身份认证等。

由于系统采用了 C/S 结构，客户端软件需具备完整的自我更新机制，包括数据更新、软件版本更新等，其中涉及服务器端资源使用权限变更监测以及检索接口变更监测。限于篇幅，本文不再赘述。

4 系统实现

桌面工具的客户端和服务端分别由 VC++ 和 Java 语言开发而成。

客户端采用对话框模式，利用 IE 控件进行结果显示。划词窗口与检索结果窗口分离，当鼠标离开后，划词窗口在设定的时间内自动关闭，以减少对用户当前操作环境的干扰。划词串的获取主要是通过截获 Window 系统中的消息以及一些应用程序接口获得。获取词串后系统需要构造相应的检索式向服务器发出检索请求。检索式通常由资源服务商提供，它是一系列程序检索接口，如 SRU、SRW、URL 等等。如果某一资源库没有提供这种接口，则需要由人工分析获得，重点是分析网页中提交检索的 URL 中的参数种类及语义信息。然后将划词串替换检索式中对应参数的值，向服务器发送 HTTP 请求后即可得到检索结果。例如在 GOOGLE 搜索中，

输入检索词“computer”后可获得如下提交 URL：
<http://www.google.cn/search?complete=1&hl=zh-CN&ie=GB2312&q=computer&meta=&aq=nurl>，分析可知其中的参数q(q=computer)接受检索词，因此只要替换其中的“computer”即可得到检索式。在实际系统中，由于检索式会经常发生变动，需要有一套自动跟踪更新机制，用于确保系统检索的有效性和正确性。

通常，系统返回的检索结果为 HTML 或 XML 两种形式，为了使结果呈现方式简洁明了，需对接受的 HTML 网页进行分析，去除其中的干扰内容（如广告、图片等），取出结果集，对 XML 进行解析，重新组织后在 IE 控件中显示出来。

服务器端用 Servlet 实现，它获取客户端的 IP 地址并据此从数据库中读取用户的权限及可访问的资源。同时利用数据库的触发机制监视资源的变动、软件版本的变更、软件是否有新补丁等信息，使客户端能实时获取到最新资源。

目前，系统已推出了体验版、BETA 版和 1.0 正式版本。在功能上，除客户端的本机资源搜索和 IE Toolbar 工具尚未实现外，其余设计功能均基本实现。图 4 是系统运行时的划词窗口，图 5 为结果显示窗口。从用户的调查反馈来看（来自中科院半导体所、力学所、物理所），用户对其易用性、实用性等方面给予了较高的评价。



图 4 划词工具处理窗口



图 5 检索结果界面

5 结论与讨论

嵌入式桌面信息服务拉近了信息使用者与提供者之间的距离，它将诸多孤立的系统联系起来即时嵌入用户的工作流程之中，疏通了用户与信息服务系统间的“最后 50 米”。实践证明，该系统是一种方便、快捷的科技文献信息查找工具，得到了众多一线科研人员的认可。

由于系统刚建成，系统在结构和功能上还不能完全满足各种用户的使用习惯。如何融入

“Web2.0 桌面”的理念，真正做到以用户自我为中心，自我管理的同时又能吐陈纳新，不断吸收互联网上的知识和“智力”，成为一个开放式桌面工具，是系统下一步改进目标。

致谢：本系统在设计、实现及测试过程中得到了中科院遥感所研究员杨崇俊老师的精心指导和大力帮助，在此表示衷心感谢！此外，国家科学图书馆王小梅、管仲、郭文丽、李书宁等也参与了部分工作，谢谢他们！

参考文献：

- [1]桌面盛宴：主流桌面搜索工具横向评测. <http://tech.sina.com.cn/s/2005-03-23/0952558793.shtml>
- [2]刘素英. 玩转WEB 2.0桌面. 网络与信息, 2006(20/10):62-63
- [3]闰辉, 刘洪洁. i桌面, 与微软Live争锋. 程序员, 2006(9):74-75
- [4]常唯. eScience与文献情报服务的变革. 图书情报工作, 2005(49/3):27-30
- [5]张晓林. 科研环境对信息服务的挑战. 中国信息导报, 2003(9):18-22