中国肿瘤数据共享技术体系建设

曲宸绪 姜勇 李纪宾 陈可欣 贺宇彤 姚树祥 乔友林

中国肿瘤数据共享系统主体数据库前身是科技部2001年课题"中国肿瘤防治数据库"。2004年科技部启动了"医药卫生科学数据共享"项目，肿瘤数据共享系统成为其中一个子项目。该课题定位于建立全国性的肿瘤防治数据共享平台，及时、准确地提供我国肿瘤发病死亡、治疗、预后和防治情况信息，促进肿瘤科研数据的二次利用。

目前肿瘤数据共享系统的数据源目前包括70、90年代两次全国死因调查数据，国内11个地区肿瘤监测数据及河北磁县和天津市的详细肿瘤监测数据，我国食管癌、肺癌和宫颈癌高发现场研究数据，医科院肿瘤医院、天津市肿瘤医院和河北磁县肿瘤医院病案数据以及其它一些常用编码数据。

肿瘤数据共享服务主要采用基于JAVA的Web Service技术开发，服务平台应用Apache Tomcat，并且利用Apache jUDDI架设私有UDDI，实现共享服务的注册与查询服务。该系统通过设置Web代理及WAP代理实现SOAP、HTML与WML三种格式封装数据的相互转换。用户可以使用浏览器、手机、个人数码助理(PDA)及各种支持Web Service的软件(MS Office，SAS等)或应用获得共享数据，共享方式得到扩展。

除针对特定数据集的共享服务接口，设置专业接口提供所有数据集的原始数据。应用XSLT技术，系统将库结构(XML Schema)转换成各数据集的WSDL(服务描述文件)，同样利用WSDL文件动态生成查询页面(Html)。因此数据集共享的开发流程简化为数据导入、数据结构撰写、访问授权、说明文档编写四个步骤。

目前肿瘤数据共享平台主要应用Web Services及相关XML技术，注重系统的先进性和可用性。

关键词：肿瘤，数据共享

作者单位：100021 北京 中国医学科学院肿瘤研究所流行病研究室(曲宸绪 姜勇 李纪宾 乔友林)；天津医科大学附属肿瘤医院流行病室(陈可欣)；河北省肿瘤防治办公室(贺宇彤)；个旧锡业公司劳动防护研究所(姚树祥)

Methodology and Technology for China Cancer Data-sharing System

Chenxu Qu, Young Jiang, Jibin Li, Kexin Chen, Yutong He, Shuxiang Yao, Youlin Qiao

Cancer Data-sharing System is the subsequent project of China Cancer Database, and it is now one of subprojects within China Medicine and Health Data-sharing System sponsored by Chinese Ministry of Science and Technology. The objectives of this project is to establish a sharing platform by which users can obtain data about incidence, mortality, prognosis, therapy, prevention of cancer, and make full use of the research data.

The data sources of this project are now mainly composed of the data from two China national cause-of-death surveys conducted in 1970's and 1990's; Data of cancer register in 11 cities of China and detailed data from Ci County and Tianjin; The research data from large-scale epidemiological studies in cancers of cervix, esophagus and lung, the medical records from Cancer Hospital of Chinese Academy of Medical Science, Tianjin Cancer Hospital and Ci county Cancer Hospital as well as some classification data, etc.

The platform of service is established on Apache Tomcat. Core services are deployed with Apache Axis (JAVA version), a Web Service implementation, and a private UDDI(Universal Description, Discovery and Integration) service is constructed using Apache jUDDI for publishing and discovering of cancer data-sharing service. A web proxy and a WAP (Wireless Application Protocol) proxy were deployed to expand ways of data retrieve. Shared data can be encapsulated with SOAP (Simple Object Access Protocol), HTML (Hypertext Markup Language) or WML (Wireless Markup Language) and be transmitted by HTTP (Hyper Text Transfer Protocol). Therefore shared data can be obtained by software supporting Web Services (MS Office, SAS etc.), Web browser, Mobile phone or PDA (Personal Digital Assistant).

Two kinds of interfaces now serve the data-sharing. The first is specific interface developed for common dataset. The other one, expert data-sharing interface, gives service to professional staff with all detailed data. Dynamic WSDL (Web Service Description Language) will be generated for certain data structure using XSLT (eXtensible Stylesheet

Language Transformation), and can also be transformed to query interface(Html). So the procedure of data publishing can be simplified to 4 steps: importing data into database, writing data structure, writing policy of access control and related document.

To summarize, the data-sharing platform deployed with web services and XML technologies tries to facilitate the application and prospective development.
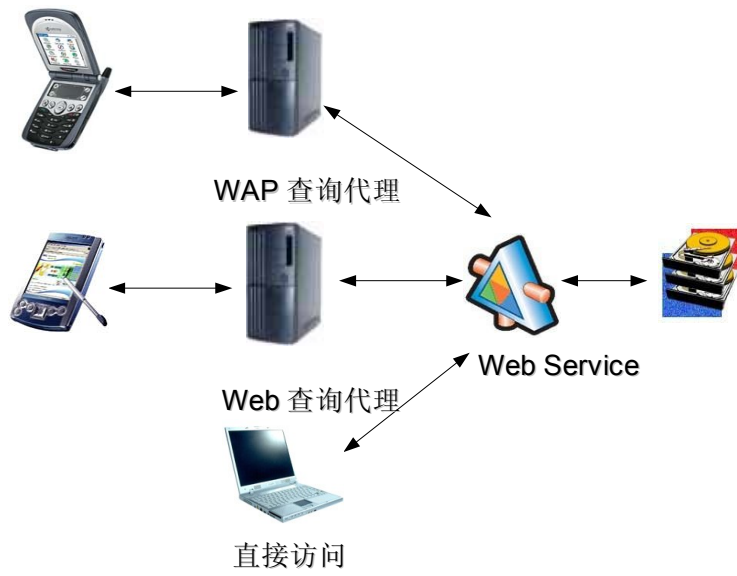
Keywords: cancer data-sharing

图1: 肿瘤数据共享系统体系结构

作者简介 曲宸绪(1975~) 男 博士研究生 流行病与卫生统计 13521848913