

2010 中国计算机大会 参展系统

一、ScholarSpace: 面向计算机领域的中文文献集成系统

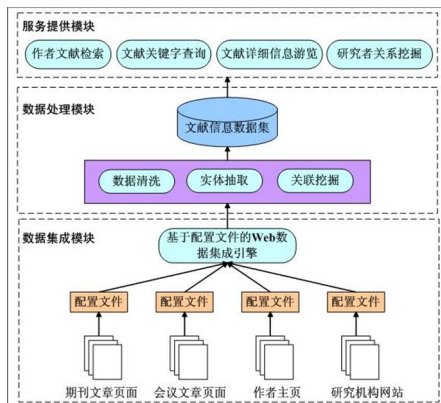
ScholarSpace: 面向计算机领域的中文文献集成系统

Web数据管理是实验室持续近十年的研究方向,在Web数据集成研究我们提出了一种面向领域的Web数据集成技术框架,并利用该技术框架开发了ScholarSpace等多个应用系统。

ScholarSpace是一个以作者为中心的自动集成、增量更新的计算机领域中文文献集成系统,在业内颇受好评。

系统特点

- ❖ **精选数据源**: 只收录国内计算机领域权威期刊和学术会议的数据
- ❖ **以作者为中心**: 系统内容组织以作者为中心,自动构建各种关联
- ❖ **集成丰富信息**: 集成作者图片、论文列表、发表数量曲线、合作作者列表、单位信息、承担项目信息等,构建作者个人学术页面
- ❖ **可视化展示**: 系统基于Flash技术动态展示一名作者与其合作作者的关系
- ❖ **同名作者区分**: 利用GHOST算法及基于作者单位信息的聚类算法,实现同名作者区分功能



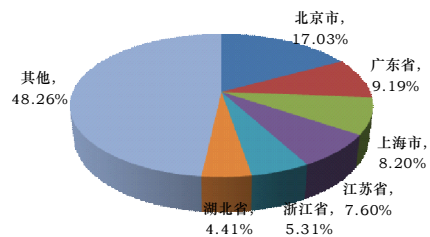
系统架构

- ❖ **数据集成模块**
利用面向领域的数据集成技术构建基于配置文件的Web数据集成引擎,集成多数据源
- ❖ **数据处理模块**
通过数据清洗、实现发现和关联挖掘建立以“实体-关联”组织的文献信息数据库
- ❖ **服务提供模块**
提供学术主页生成、作者/文献检索、实体关联发掘等应用服务

社会关注



中国计算机学会会刊/网站新闻

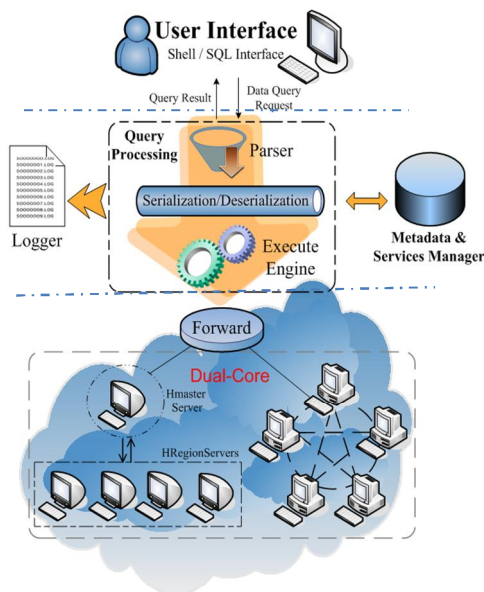


系统访问来源分布图

二、TaijiDB: 云数据管理系统

云数据管理系统TaijiDB

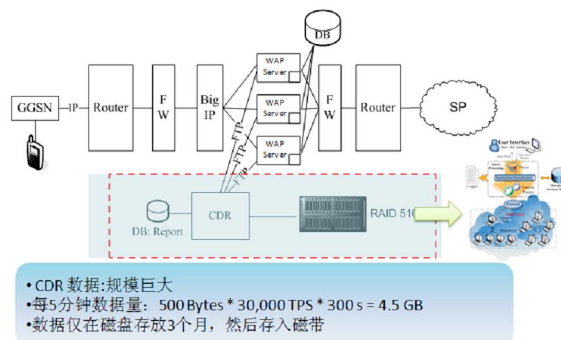
TaijiDB系统综合主从（Master-slave）和点对点（P2P）两种云数据存储架构，继承了两种架构在CAP、数据写操作、MapReduce、系统性能、应用场景等方面的优势与特点。TaijiDB支持SQL和Shell两种接口，可在低端机器上进行部署，可靠性高，扩展性强，并能自动进行冗余备份和错误恢复，管理海量的纷繁复杂的混乱数据，使之有条不紊，便于查询。



系统架构

系统功能

- ❖ 建立适配层桥接底层的主从结构和点对点结构。上层通过统一API调用进行操作，底层的差异对上层透明。
- ❖ 支持部分SQL查询语句，包括建表、插入数据、选择数据、删除数据和数据表等。
- ❖ 支持部分文件操作



应用场景

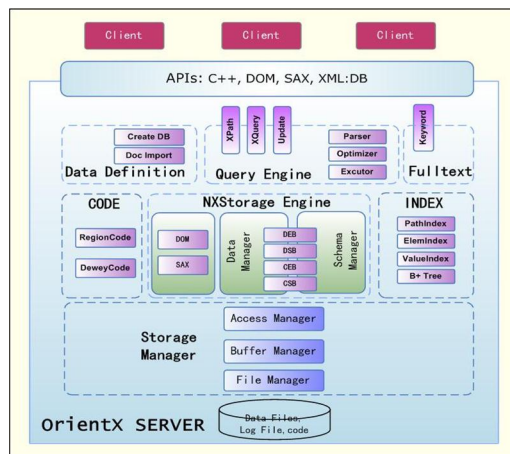
三、OrientX: XML 数据库系统

OrientX: 纯XML数据库系统

OrientX 是由中国人民大学网络与移动数据管理实验室开发的纯XML数据库系统。从2002年以来,我们不断完善OrientX系统,到目前为止已经陆续发布了6个版本。系统依据XML的数据模型和特性高效地存储数据,同时支持XQuery和XPath查询标准以及XQuery/Update更新标准。OrientX 采用C/S模式,客户端提供了图形化用户界面方便用户操作和查询数据,服务器端提供一套访问数据库的API接口。

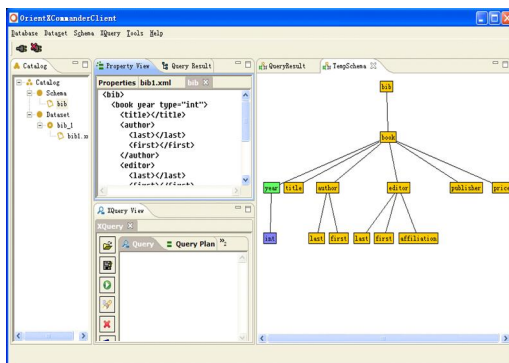
开发版本

- ❖ OrientX 1.0 (2002-2003):包括存储模块、模式管理和数据管理模块三个模块;
- ❖ OrientX 1.5 (2003-2004):增加了XPath查询处理模块、编码模块和索引模块;
- ❖ OrientX 2.0 (2004-2005):增加了基于导航的查询引擎;
- ❖ OrientX 2.5 (2005-2006):添加了基于代数的查询引擎;
- ❖ OrientX3.0 (2006-2007):支持XQuery/Update标准的插入、删除、替换以及重命名四个操作;
- ❖ OrientX3.5 (2008-2009):支持XQuery/Update的transform操作并加入整枝匹配算法。



OrientX3.5 系统结构图

系统功能



用户界面

- ❖ 数据库的建立和维护: 管理模式信息, 创建和删除数据库;
- ❖ 数据操作: 包括对数据库数据的检索以及增删改操作;
- ❖ 数据组织、存储和管理功能: 提高存储空间利用率以及查找、增删改等操作的时间效率;
- ❖ 应用程序接口: 提供了一套C/C++ API接口;
- ❖ 图形化用户界面: 方便用户进行数据库管理、查询和更新操作以及查看查询执行计划。

四、OrientPrivacy: 移动环境下的隐私保护系统

OrientPrivacy: 隐私保护系统

随着无线通讯技术和移动定位技术的发展,移动用户的隐私也越来越受到人们的关注。基于位置位置服务中有两种隐私类型:用户位置隐私和查询内容隐私。OrientPrivacy系统集成多种隐私保护方法,可以根据用户隐私需求保护其位置隐私和查询内容隐私。

系统特点

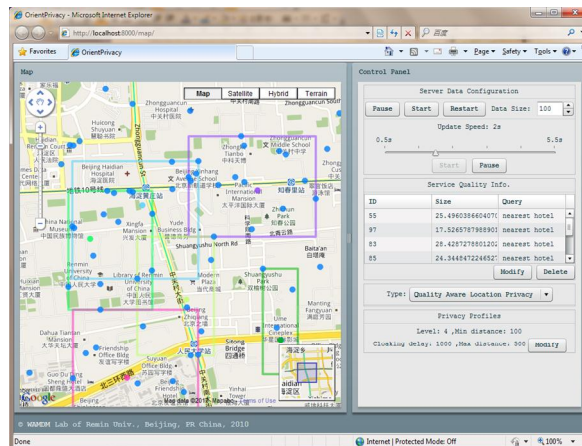
- ❖ 采用多种算法,可以保护用户的位置隐私和查询内容隐私
- ❖ 用户可以配置隐私需求参数,服务器根据其参数和不同查询类型自动选取合适的隐私保护算法
- ❖ 服务器界面可以实时显示隐私保护处理状态,对比匿名前后结果
- ❖ 服务器端可以在线访问
- ❖ 使用真实的地图数据



手机主界面



手机端隐私保护参数设置



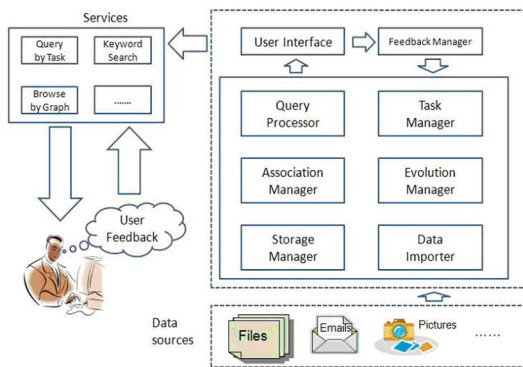
服务器端主界面

五、OrientSpace: 个人数据空间系统

OrientSpace: 个人数据空间系统

在信息社会, 计算机和通信技术的发展使数据量急速膨胀, 每个人都面临着数据海量以及高度异质带来的挑战, 却缺乏有效的管理工具。例如, 用户很难在电脑中找到一个特定的文件。因此, 如何管理海量、分布、异构数据成为一个值得关注的问题。

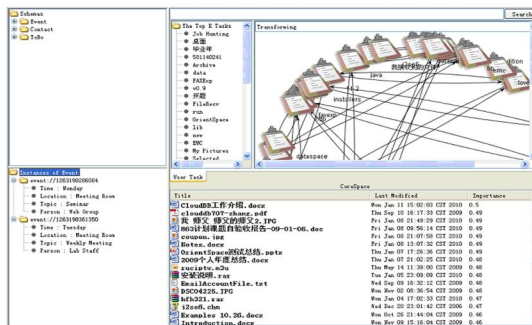
个人数据空间系统OrientSpace即基于数据空间理论, 以个人数据空间管理为应用背景, 关注数据管理中的任务管理、演化管理、数据关联管理、用户反馈管理等, 致力于为用户提供高质量的数据服务。



OrientSpace系统框架图

关键技术

- ❖ 通过用户行为监控引擎识别用户访问行为
- ❖ 自动形成用户任务
- ❖ 构建并自动更新核心数据空间
- ❖ 构建并自动更新任务空间
- ❖ 基于三元组的全局式存储策略
- ❖ 多级索引、选择性索引与混合索引相结合的索引策略



OrientSpace主界面

系统特色

- ❖ 灵活的数据模式, 允许用户灵活自由地创建和修改数据模式
- ❖ 支持复杂多样的数据关联, 并支持基于关联的查询
- ❖ 支持基于用户行为的数据空间演化, 充分考虑了用户的个性化信息
- ❖ 提供自动建立和更新数据空间的功能
- ❖ 支持基于任务的数据操作, 便于用户浏览任务及其关联的文件

下载地址: <http://idke.ruc.edu.cn/projects/pds.htm>

六、Flash DB: 闪存数据库系统

Flash DB: 闪存数据库系统

随着闪存容量的快速增长, 闪存上的数据管理问题已成为新的挑战性问题。从数据库技术的角度研究闪存数据管理; 从闪存的器件特性入手, 构建新型的闪存数据库系统(简称Flash DB)。

闪存特性

- ❖ 非易失性、高速、高抗震性、低功耗、小巧轻便;
- ❖ 无机械延迟, 即没有寻道时间和旋转时间;
- ❖ 原位更新代价过大, 一般采用换位更新方式;
- ❖ 读/写/擦除速度存在较大差异, 读操作速度明显高于写操作, 而擦除操作则更慢;
- ❖ 块具有擦除次数限制, 超过限制后, 该块将成为坏块, 不再可靠。



固态硬盘



闪存数据库系统体系结构

闪存数据库系统

- ❖ 可剪裁闪存数据库体系结构;
- ❖ 闪存数据库存储管理;
- ❖ 闪存数据库索引;
- ❖ 闪存数据库缓冲区管理;
- ❖ 闪存数据库查询处理;
- ❖ 闪存数据库事务处理;
- ❖ 闪存数据库隐私保护;

基于闪存的索引

- ❖ 基于FTL的索引: FTL为闪存转换层, 把闪存转换为块设备来使用, 基于FTL的索引主要分为Hash、B-tree和R-tree三种;
- ❖ 基本的索引方法: 主要分为Hash、B+-Tree和Pbfilter三种;

