



热门文章

用多元线性

间借贷利率

何加强会计

如何处理银行

国外汇储备

章

章

品市场竞争

业银行走混

国存款保险

国创业板市

华夏并购案

[2009年11月]互联网数据自动采集系统设计

【字体：大 中 小】

作者：[陈亚楠 谢声时 王凌飞] 来源：[本站] 浏览：

一、前言

随着我国加入世界贸易组织，我国金融业正在走向市场，业务经营逐步规范化、市场化、国量以表外业务为特征的金融创新业务不断涌现，使得金融机构的经营方式更为灵活。在这种金融运行环境越来越复杂，形势变化越来越快，金融机构经营中的潜在风险也不断增加且更和控制。从金融行业的管理来说，建立一个有效的决策支持系统（俗称DSS，Decision Support System）显得十分必要。

建立决策支持系统的基础是对数据的分析。在当前的形势下，而有效并全面的收集和处理互是数据分析的重要方面。在当前情况下，由于不同网站、不同指标数据的更新频度不同，采为不得不采用人工行为或使用简单工具辅助，同时采集工作被动地反复进行。采集到本地的数据没有统一的清洗和归整机制，无法统一高效地利用，无法为更深层次的数据分析提供条此，建设互联网数据自动采集系统，数据采集，数据处理，数据入库和查询，可以更好持服务。

二、数据自动采集系统方案

互联网数据采集系统主要包含以下内容：（一）数据采集，指允许用户根据需要，管理目标所有信息，实现定向、定时、自动化采集；（二）数据归整，从采集到本地的各种异构的源准确抽取指标数据，并保存到指标数据库中。（三）数据查询，为用户提供查询功能。

（一）系统逻辑方案

如下图，系统使用BS架构对浏览器设置无特殊要求；并采用模块化的体系结构，标准化接口放，全面支持二次开发与系统集成；提供友好的管理员界面，简化操作。

在本方案中，先将采集到的目标源文件在服务器硬盘上集中存放，再通过ETL工具设计模板，文件进行抽取、清洗、转换，并装载到指标数据库，形成符合要求的指标数据，为查询、导据基础。

（二）系统模块设计

1. 任务定制管理：用户可新建采集任务并自定义采集任务相关细节：访问资源协议、资源地址类型、资源有效性验证方式等属性配置。

2. 数据采集：根据定义好的某个任务，到互联网上采集数据，通过HTTP、FTP等网络协议，未定位采集相关数据资源（html、excel、word、PDF等），并将这些资源存储到本地系统中时目录或缓冲区中。本模块需要高效利用网络资源，需要有相关网络连接调度池管理及功能相关日志信息。最后执行数据有效性检查，若资源无效则抛出相关异常，由调用者负责处理

3. 数据预处理：将任务执行采集单元采集到的相关数据文件做首次整理归档，主要是为了方植和其他子系统的导入。在数据采集后需要将临时目录下的数据与文件资源进行归档，并根索要求，查找文档内容里的关键字，创建索引。同时根据文档内容生成符合全文检索要求的据。最后将整理归档好的数据移动到归档区，方便数据转移。

4. 任务调度管理器（任务池）。该模块处本系统核心，所有采集作业（任务）都是有此模块发。同时还支持以下功能：查看正在进行中的任务、终止某个正在进行中的任务、按时间表任务、通知数据预处理单元工作。本模块还包含了数据核对确认功能，未通过核对的数据不据预处理工作，这部分数据将继续保留在临时工作目录下。

5. 手工补录：补录采集数据提供的方式有在线填报指标数据和文件上传。在线填报指标数据面按指标填入需要补录的数据，在线提交。文件上传方式可上传一个EXCEL报表，系统将报析装载。作为一个上报数据的一个其他解决方案，用户还可以制定固定的模版，用相同的模不同数据时间的数据。

6. 日志管理。本模块提供系统的安全审计功能，相关的管理操作都会形成日志记录在数据库三、安全管理

（一）应用系统安全

实现方法是通过部署防火墙和审计系统的方式来实现。防火墙保证网络边界安全，实现了网的访问控制，避免由于涉密数据被非法窃听或者嗅探后所导致的泄密情况发生，在数据链路保证了数据传输的安全。

（二）数据安全

本系统数据部分具有敏感性，对于业务数据的查询，保存与转移要有高度的安全保障体系作为支撑。为了防止对系统数据的侦听、窃密、非法修改、冒名顶替、恶意破坏等，需要采取必要的技术措施，保证交换数据的机密性、完整性、可靠性、不可否认性。金融运行分析系统在数据安全保障体系中有如下方案：

（三）校验数据载体的正确性

系统在处理数据载体是首先会校验这些载体的正确性，包括文件名称，文件格式，数据时间等。在验证数据载体没有错误以后，系统会校验数据的精度等。只有所有验证都通过，数据才会进行入库操作。

（四）数据备份

数据库定时备份防止数据因为意外情况而丢失。由于金融数据的重要性，定期的数据库备份可以防止数据因意外情况丢失所带来的隐患。数据库备份分为脱机备份和联机备份两种。

四、小结

本系统初步实现了互联网数据收集、汇总整理，为进行更深层次的研究分析，支持专家决策提供了数据基础。在下一步的工作中，我们将对数据地址智能匹配、数据清洗和数据整合方面进行更深入的研究，对系统进行完善和优化。

（作者单位：中国人民银行海口中心支行）

120+ renowned advisors reveal what to buy and what to sell

Meet face-to-face with top investment experts

Acquire a global market perspective

Discover profitable investment insights...

The WORLD MONEY SHOW Utilizing the Global Investment Community

WORLD MONEY SHOW 13th MONTH

insights... investment boutique discover

正在读取...



笔名:



评论:

[评论将在5分钟内被审核, 请耐心等待]

【注】 发表评论必需遵守以下条例:

- 尊重网上道德, 遵守中华人民共和国的各项有关法律法规
- 承担一切因您的行为而直接或间接导致的民事或刑事责任
- 本站管理人员有权保留或删除其管辖留言中的任意内容
- 本站有权在网站内转载或引用您的评论
- 参与本评论即表明您已经阅读并接受上述条款