

ICS

CCS 点击此处添加 CCS 号

团 体 标 准

T/XXX XXXX—XXXX

互联网价格大数据抓取及分析系统技术规范

Technical Specification for Internet Price Big Data Capture and Analysis System

XXXX – XX – XX 发布

XXXX – XX – XX 实施

全国城市工业品贸易中心联合会 发 布

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 功能要求 1

5 性能要求 2

6 安全要求 3

7 系统测试 3

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由 提出。

本文件由全国城市工业品贸易中心联合会归口。

本文件主要起草单位：

本文件主要起草人：

互联网价格大数据抓取及分析系统技术规范

1 范围

本文件规定了互联网价格大数据抓取及分析系统技术规范的术语和定义、功能要求、性能要求、安全要求、系统测试。

本文件适用于互联网价格大数据的抓取及分析系统。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 21052 信息安全技术 信息系统物理安全技术要求

GB/T 25000.51-2016 系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第51部分：就绪可用软件产品（RUSP）的质量要求和测试细则

3 术语和定义

下列术语和定义适用于本文件。

3.1

互联网价格大数据抓取及分析 Internet price big data capture and analysis

自动化地抓取互联网上的价格数据，并进行分析和展示。

4 功能要求

4.1 数据抓取功能

4.1.1 支持多种网站类型和格式

系统应能够处理各种不同类型的网站和数据格式，包括但不限于HTML、XML、JSON等。

4.1.2 高效性

自动化采集可以节省大量时间和人力成本，系统应能够快速、准确地抓取目标数据。

4.1.3 实时性

系统应能够实时地获取最新数据，以确保价格信息的及时性和准确性。

4.1.4 可靠性

自动化采集应减少人工干预，从而降低错误率，提高数据的可靠性。

4.1.5 精确性

系统应能够准确地抓取目标数据，避免误抓或漏抓。

4.2 数据处理功能

4.2.1 数据清洗

对于抓取的原始数据，系统应能够进行有效的清洗和整理，去除重复、错误或无效的数据。

4.2.2 数据整合

系统应能够将来自不同来源的数据进行整合，形成统一的数据格式和结构，便于后续的分析和展示。

4.2.3 数据存储

系统应提供可扩展的数据存储方案，以容纳海量的价格数据，并确保数据的安全性和稳定性。

4.3 数据分析功能

4.3.1 数据分析算法

系统应内置多种数据分析算法，如趋势分析、关联分析、聚类分析等，以满足不同分析需求。

4.3.2 报表生成

系统应能够根据需求生成多样化的、有针对性的数据报表，如价格走势图、竞品对比表等。

4.3.3 数据可视化

系统应提供丰富的数据可视化工具，如柱状图、折线图、饼图等，以使用户更直观地了解价格数据。

4.4 用户交互功能

4.4.1 用户界面友好

系统应具有简洁、直观的用户界面，便于用户操作和使用。

4.4.2 定制化服务

系统应支持用户根据自己的需求定制抓取和分析任务，如设置抓取频率、选择分析算法等。

4.5 报告导出

系统应支持将分析结果导出为常见的文件格式（如Excel、PDF等），方便用户共享和使用。

4.6 管理和监控功能

4.6.1 系统应提供系统配置、用户管理、权限控制等管理功能。

4.6.2 系统应能对性能、数据质量、抓取效率等关键指标的实时监控和报警。

4.6.3 系统应提供提供日志管理和审计功能。

5 性能要求

5.1 功能性

系统的功能性应符合 GB/T 25000.51-2016中 5.3.1 的规定。

5.2 可靠性

系统的可靠性应符合 GB/T 25000.51-2016 中 5.3.5的规定。

5.3 易用性

系统的易用性应符合 GB/T 25000.51-2016 中 5.3.4 的规定。

5.4 效率

系统的效率应符合 GB/T 25000.51-2016 中 5.3.10 的规定。

5.5 维护性

系统的维护性应符合 GB/T 25000.51-2016 中5.3.7 的规定。

5.6 可移植性

系统的可移植性应符合 GB/T 25000.51-2016 中 5.3.8 的规定。

5.7 使用质量

系统的使用质量应符合系统操作手册的规定。

6 安全要求

6.1 身份安全要求

- 6.1.1 应具有机密性，可采用对称和非对称加密技术，确保敏感信息不被窃取。
- 6.1.2 应具有完整性，可采用数字摘要确保身份信息不被篡改。
- 6.1.3 应具有真实性，身份鉴别可采用数字证书确保身份真实性。
- 6.1.4 应具有不可抵赖性，可采用数字签名确保数据发送方不可否认自己的信息。

6.2 网络传输安全

- 6.2.1 应在参与联盟链的节点之间建立安全传输通道，保证数据传输的完整性和不可篡改性。
- 6.2.2 应对数据和信息采取相应的防护措施，保证其能抵抗篡改、重放等主动或被动攻击。
- 6.2.3 应保证节点间通信过程中敏感信息字段或整个报文的保密性，宜采用密码学技术。
- 6.2.4 应确保信息在存储、传输过程中不被非授权用户读取和篡改，可采用有权限的网络访问控制，在参与分布式账本节点之间构建虚拟专用网络（VPN），降低网络攻击造成的危害。

6.3 物理安全

应符合GB/T 21052的要求。场地安全部署物理数据中心及附属设施的，用于业务运行、数据存储和处理的物理设备需符合国家监管要求。

6.4 节点部署安全

- 6.4.1 应保证承担共识的节点冗余部署，保证系统可用性；避免将所有承担共识的节点部署在同一机房内，能在单一机房节点不可用时保证系统整体的可用性。
- 6.4.2 确保部署节点的硬件设备存储容量可扩展，避免因数据容量达到上限而无法同步数据。

6.5 硬件设备

- 6.5.1 应对设备运行状态、资源使用情况等进行监控，能在发生异常时发出告警。
- 6.5.2 设备和存储介质在重用、报废或更换时，能对其承载的数据进行清除且不可恢复。

7 系统测试

7.1 系统性能测试

7.1.1 时间特性

通讯相应时间:使用测试用例,对100条信息进行相应时间测试,总计用时/信息条数=平均相应时间。

7.1.2 安全稳定性

- 7.1.2.1 设置不同类型的用户权限，验证是否满足系统性能的要求。
- 7.1.2.2 使用非法认证信息旁路安全访问限制，验证是否满足系统性能的要求。

7.2 系统功能测试

7.2.1 软件功能测试包括单元测试和集成测试，测试方法和测试结果应满足 GB/T 25000.51 的要求，及本文件的功能要求，具体情况如下：

- 单元测试：由程序员对单元模块进行测试；
- 集成测试：测试各模块间的接口和联系，验证功能的完整性。

7.2.2 测试方法：采用单元测试、集成测试方法，在确定测试用例后，进行反复调测。测试结果应该满足本文件的功能要求。

7.3 系统安全性

- 7.3.1 采用黑盒测试方法，对涉及安全的软件功能进行测试，应满足系统性能要求。
 - 7.3.2 查看系统用户密码存储字段，三次输错用户密码进行测试，应满足系统性能要求；采用主机或系统漏洞扫描器自动检测远程或本机安全性弱点，应满足系统性能要求。
 - 7.3.3 采用侦听技术对数据进行截取分析，应满足系统性能要求。
 - 7.3.4 采用模拟攻击试验，冒充、重演、消息篡改、服务拒绝、内部攻击、外部攻击方法进行测试，应满足系统性能要求。
-