博物馆新闻采集分析

子系统需求规格说明

**目录**

1. 导言...................................................................................................................3
   1. 版本修订历史..........................................................................................3
   2. 概述............................................................................................................3
2. 功能模块..........................................................................................................3
3. 需求分析..........................................................................................................3

3.1 数据获取功能.........................................................................................3

3.2 数据加工功能.........................................................................................4

3.3 数据分析功能.........................................................................................4

3.4 数据定制服务功能................................................................................5

1. 性能要求..........................................................................................................5

4.1 可靠性.......................................................................................................5

4.2 安全性.......................................................................................................6

4.3 可维护性...................................................................................................6

4.4 高效............................................................................................................6

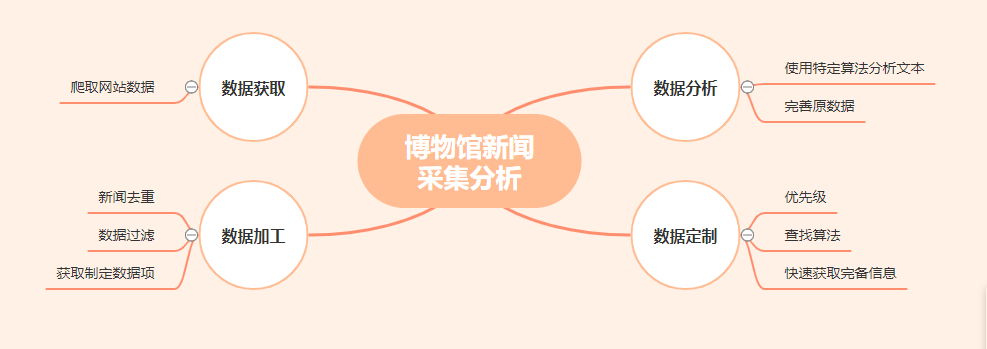
4.5 有效性........................................................................................................6

1. **导言**
   1. **版本修订历史**

|  |  |  |
| --- | --- | --- |
| **版本号** | **修订内容** | **修订者** |
| 1.0 | 系统需求规格说明 | 罗朝旭 |

* 1. **概述**

本系统主要功能是通过程序能从主要的新闻网站上爬取博物馆相关的新闻信息，进行加工处理。主要包括数据获取、数据加工、数据分析、数据定制服务四项功能。

1. **功能模块**
2. **需求分析**
   1. **数据获取功能**
      1. 数据

主要的正规新闻网站中有关博物馆的新闻。

* + 1. 获取要求

①每一篇新闻至少包括标题、正文、发布时间三项内容，可以适当爬取新闻评论内容。

②同一篇新闻不同的发布时间的情况下，使用发布时间最新的优先原则。

③数据内容清晰。

④对于新闻中的图片，获取并根据新闻标题和发布时间存储。

* + 1. 实现方法

**不知道写啥**

* 1. **数据加工功能**
     1. 加工要求

①根据文本相似度和设定的阈值对新闻进行去重。

②过滤掉无用的信息，例如新闻页面的广告等。

③至少抽取新闻标题、正文、发布时间、新闻涉及的博物馆这四项内容，可以试着抽取新闻的评论内容。

④加工后存储的数据项定义为博物馆、发布时间、新闻标题、新闻内容或者其他能够满足后续处理的形式。

* + 1. 实现方法

**不知道写啥**

* 1. **数据分析功能**
     1. 分析要求

①分析每一篇新闻，判断其属于正面新闻还是负面新闻，并存储判断结果。

②确保每一篇新闻的判断是有效的（即是正面和负面中的一个）。

③不能改变数据原本的存储结构。

* + 1. 实现方法

不知道写啥

* 1. **数据定制服务功能**
     1. 数据

①确保已经被完整、有结构地存储在数据库中。

②对于存储结构可以根据实现的需要，设置相应优先级，处理效率。

* + 1. 定制服务要求

①支持根据指定的博物馆进行新闻搜索。

②支持根据指定的时间段、新闻性质（正面或者负面）搜索。

③服务接口与获取信息并展示的子系统保持一致。

④使用高效的算法，确保信息获取的敏捷、完备。

* + 1. 实现方法

**不知道写啥**

1. **性能要求**
   1. **可靠性**

①数据均来源于正规网站。

②用户端获取的信息正确、完备。

* 1. **安全性**

①数据传输过程中受到保护。

②数据库中的数据不会被恶意篡改。

* 1. **可维护性**

①支持定期更新和维护数据。

* 1. **高效**

①用户端所获取的数据能够在短时间内被筛选和显示。

②数据具有良好的空间存储结构，不会造成大面积空间浪费。

* 1. **有效性**

①根据用户端的要求，传送的数据是有效的，不包含无意义的信息。