目录

**第一章 概述1**

1.1 选题背景及意义2

国内外研究现状

1.2 论文的主要内容5

**第二章相关技术与开发工具4**

2.1 OWL 本体语言简介5

2.2 Protégé本体建模工具5

2.3 Jena技术5

2.4 jess介绍5

2.5 ontology与数据质量5

2.6 熵权法5

2.7 Java语言与Java EE5

2.8 B/S模式5

2.9 MySQL数据库5

2.10 IntelliJ IDEA5

**第三章 数据质量本体构建1**

3.1 本体的概念、关系以及约束公理2

3.2 本体的构建准则2

3.3 数据质量本体构建流程2

3.4 数据质量本体类以及属性2

3.5 数据质量规则对在本体中的描述2

**第四章 基于jess的数据质量评价方法4**

4.1 jess规则5

4.2 熵权法设计5

4.3 熵权法在本体中的设计实现

4.4 数据质量的jess规则实现5

**第五章 系统实现1**

5.1 数据质量本体2

5.2 基于熵权法的数据质量评价实现2

5.3 评估实现2

**第六章 系统测试4**

6.1测试基本原则5

6.2软件测试步骤5

6.3 测试用例5

**结论1**

**参考文献4**

**致谢4**

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 数据属性 | 定义域 | 值域 |
| 1 | UpperLimit | AssessedColumn | Float |
| 2 | LowerLimit | AssessedColumn | Float |

1．完整性  
完整性（Integrity）用来描述数值信息的完整程度[21]。完整性包括完整性约束规则、非空约束规则、连续性约束规则三个维度的规则。  
完整性就是我们日常所说的“数据要全”。在数据迁移和数据采集的过程中要把需要的数据采集完整了，比如产量只采集一个数字而不记录单位，那么就无法获取产量的完整信息，只有信息完整才能对企业管理和生产决策有意义。  
2．一致性  
一致性（Consistency）用来描述同一记录在多个数据集上存储的同一变量的一致程度 。一致性包括三层含义：一种是数据集之间的一致性关系，是指数据集中的某些属性与其他数据集的属性集合的数据存在等值关系或逻辑关系；一种是数据集内部的一致性关系，也就是数据集的属性之间应该满足等值或逻辑关系；最后一种是数据集之间的存在关系，具体指表之间的父子引用关系。一致性包括等值一致性依赖约束规则、存在一致性依赖约束规则、逻辑一致性依赖约束规则三个维度的规则。一致性的三种含义对应三个维度的约束规则。  
3．准确性  
准确性（Accuracy）是描述待测数据与真实数据的符合程度，数据的准确性体现在数据内容和形式上的准确，即数据的类型、格式、精度及值域范围的准确[22]。例如在企业的实际生产中，准确性往往是指在对数据的操作过程中采集的数据与真实值之间的差异，值的误差越小准确性越高。  
影响准确性的原因很多，在数据的录入、修改、审核、传输每个环节都有可能出现。我们需要从产生误差的原因来分析影响准确性的因素，使误差尽可能的控制在理想的范围内，提高数据的准确性。  
准确性包含值域约束规则、代码约束规则、逻辑依赖约束规则、词法约束规则、非空约束规则、等值函数依赖约束规则六个维度。