

유통 상품의 데이터 품질 관리를 위한 데이터 표준화에 대한 연구

송장섭*, 류성렬**

An Empirical Study on Quality Improvement by Data Standardization for Distributed Goods

Jang-Seop Song*, Sung-Yul Rheo**

요약

데이터 품질 관리는 매우 중요하다. 본 연구는 효율적인 기업 데이터의 품질 관리를 위한 데이터 표준화 설계를 유통 상품 사례로 구축 방안을 제시하고 그 효과를 검증 하였다. 데이터 표준화 설계 방안으로 데이터 표준화 체계와 데이터 사전설계를 설계 하였다. 데이터 표준화 체계 설계를 위하여 데이터를 분류, 속성, 식별하였으며, 데이터 사전 설계를 위하여 데이터 사전 설계 프로세스와 단어·용어·도메인·코드사전을 구축하고, 데이터 표준화 설계 방안을 제시하였다. 제시한 데이터 표준화 방안의 효율성을 정량적, 정성적 방법으로 검증한 결과 데이터 표준화로 인한 데이터 품질은 24% 및 데이터 사전의 속성 설계인 일관성에 대한 데이터의 구조적 품질은 7% 향상되고, 유효함을 입증하였다.

▶ Keywords : 상품, 데이터 품질, 데이터 표준화, 데이터 사전, 마스터 데이터

Abstract

Data quality management is extremely important. In this study, we proposed data standardization for effective quality management of enterprise-owned data about distributed goods and validated its effectiveness by case study. For the standardization of data, we designed data category and data dictionary. Additionally, we categorized data and identified its attributes for data category design, and we developed design process for data dictionary and built the dictionary of word, term, domain and code for data dictionary design. And then we proposed output documents which have to be written for data standardization. Proposed data standardization approach is validated its efficiency by quantitative and qualitative measurement, and as a result

•제1저자 : 송장섭 •교신저자 : 류성렬

•투고일 : 2013. 5. 20, 심사일 : 2013. 6. 4, 게재확정일 : 2013. 6. 25.

* 송실대학교 대학원 컴퓨터학과(Dept. of Computer Science, Soongsil University)

* 송실대학교 대학원 컴퓨터학과(Dept. of Computer Science, Soongsil University)

the data quality of the data standardization improved 24% and the data quality of the consistency of the data dictionary improved 7%.

▶ Keywords : Goods, Data Quality, Data Standardization, Data Dictionary, Master Data

I. 서 론

마스터 데이터 연구기간 보고서에 따르면 “잘못된 마스터 데이터로 인한 문제를 경험한 적이 있습니까?”에 대한 질문에 대해서 741개 회사 응답 결과 83%가 “Yes”, 9%는 “No”, 나머지 8%는 “Don't Know”라고 응답했고, 어떤 문제였습니까? 질문에 데이터의 “신뢰성 및 부정확한 보고”에 대해서 81%가 응답했고, 그러면, “어떤 이점이 있습니까?”에 대한 질문은 데이터 품질 개선에 76%가 응답했다[1].

마스터 데이터 관리는 비즈니스와 IT 조직들이 상호 협력하는 기술 주도의 비즈니스 영역으로, 기업 조직 내 공식적이면서 공유된 마스터 데이터 자산에 대한 균일성, 정확성, 책임성, 시멘틱 일치성을 보장하도록 하는 방법 중 하나이며, 전사 차원에서 활용하면 데이터 통합과 동기화에 있어 IT 비용을 30% 절감 할 수 있다고 가트너 보고서는 말한다.

유통 상품은 식품과 비식품으로 나누며, 규격과 비격으로 구분된다. 신선·즉석·기획 상품에 대한 특성과 이것을 사용하는 다양한 사용자를 고려한 데이터 표준화가 요구된다.

본 연구는 데이터 품질 관리의 구축 단계에서 유통 상품의 데이터 표준화에 목적을 가지며, 목적을 달성하기 위하여, 데이터 표준화 설계 방안을 제시한다. 데이터 표준화 체계는 분류·속성·식별 체계로 작성되며, 표준화된 데이터는 데이터 사전인 단어·용어·도메인·코드사전으로 관리한다.

본 연구에서 제시한 유통 상품의 데이터 표준화는 실증적 연구의 유효함에 기여할 것으로 생각한다. 데이터 표준화는 의사소통의 혼란방지, 데이터의 정확성과 일관성을 유지하며, 품질의 데이터를 확보한다.

본 연구의 구성은 다음과 같다. 다음 II 장에서 본 연구 내용과 유사한 기존 연구들을 분석하며, 이를 통해 본 연구와 기존 연구와의 차이점을 말한다. III 장에서는 데이터 표준화 설계에 대해서 제시하며, IV 장은 사례를 통한 검증한다. 마지막으로, V 장은 본 연구의 결론 및 한계점과 향후 연구에 대해서 제시한다.

II. 관련 연구

마스터 데이터 품질관리에서 유통 상품의 데이터 표준화 설계 제안에 앞서 기존 관련 연구에 대한 분석과 고찰이 필요하다. 본 연구에서는 이러한 관련 연구를 3 가지 분야로 나누어 분석한다.

첫 번째는 마스터 데이터 품질관리 프레임워크에 대한 분석이다. 데이터의 품질관리 대상은 데이터, 데이터 품질관리 항목, 데이터 관리 프로세스의 세 요소로 구성된다[2]. 또한 ISO 8000에서는 데이터를 제품(Product) 데이터, 측정(Measurement) 데이터, 거래(Transaction) 데이터, 마스터(Master) 데이터로 분류한다[3].

두 번째는 데이터 표준화 체계와 대상인 코드·데이터사전·도메인 표준화에 대한 분석이다. 데이터 표준화 체계는 문헌 분류 체계를 분석하며, 코드 표준화는 정보시스템에서 동일한 의미로 해석되고, 공통으로 적용될 수 있도록 중복 정의되어 있는 코드들을 표준화 하는 것이다. 데이터사전 표준화는 데이터 사전인 단어와 용어로 구성되며, 단어는 일정한 의미를 갖는 최소 단위를 의미하며, 용어는 단어의 조합으로 구성된다. 도메인 표준화는 속성(칼럼)이 갖는 공통적인 특성(예, 허용 값 범위, 데이터 타입, 길이 등)을 그룹화 하여 특성을 정의한 것이다[4].

1. 마스터 데이터 품질관리 프레임워크

마스터 데이터 품질관리 프레임워크는 [그림 1]과 같다. 이것은 데이터 활용, 품질모니터링, 품질개선의 3가지 기능과 관리자, 통제자, 실행자의 3그룹 역할로 3 x 3 조합의 9개 프로세스로 구성되어 있다[5].

프레임워크의 주요 3가지 기능은 다음과 같다. 데이터 활용은 분산된 데이터를 식별, 통합 관리하여 데이터를 공유 및 활용이 되도록 지원하는 기능을 말한다. 품질모니터링은 품질 관리 방향, 목표, 측정항목 그리고 기준을 설정하고, 지속적인 품질 측정기능을 말한다. 품질개선은 발견된 오류수정, 추적과 분석을 통해서 근본 원인 파악 및 재발하며, 데이터 접

3. 데이터 표준화 검증

품질 측정을 위한 메트릭은 전체 데이터 구조의 일관성 측면에서 오류가 발생한 비율을 이용하여 구하도록 정의한 메트릭을 적용 하였다(10). 각 컬럼별 오류 데이터(cij)는 표준데이터와 일관성이 맞지 않는 경우 오류로 판단하며, 가중치(wj)는 각 컬럼의 개수가 전체 컬럼 개수에서 차지하는 비율로 계산한다.

- Q : 데이터 품질 측정값
- N : 총 품질 측정 대상 속성개수
- T : 컬럼별 가중치를 부여한 총 오류속성 개수
- mj : 총 컬럼 수
- cij : 각 컬럼에 해당하는 오류 속성의 개수
- wj : 각 컬럼에 해당하는 가중치

$$Q = 1 - \frac{T}{N}$$

$$T = \sum_{j=1}^{m_j} c_{ij} \times w_j \quad (식1)$$

본 연구에서는 관련 연구를 기반 연구로 수용하며, 유통 상품의 특성에 대한 데이터 표준화를 제안한다는 것이 본 연구의 독창성이다.

III. 표준화 설계 방안

데이터 표준화 체계는 분류·속성·식별체계로 작성되며, 표준화된 데이터는 데이터 사전인 단어·용어·도메인·코드사전에 의해 관리된다.

1. 데이터 표준화 체계 설계

1.1 분류체계 설계

분류체계는 [그림 2]과 같다. 분류 체계는 한국유통물류진흥원의 KAN 상품분류코드 기반으로 설계 한다.

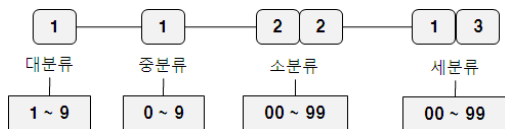


그림 2. 분류체계 설계
Fig. 2. Classification structure design

분류단계는 4단계로 대분류(1~9), 중분류(0~9), 소분류(00~99), 세분류(00~99) 단계로 분류한다.

1.2 속성체계 설계

[그림 2]중에서 세분류는 속성을 갖는다. 속성체계는 공통 속성 및 개별속성으로 구분하며, [표 4]과 같다.

표 4. 속성체계 설계
Table 4. Attribute Category Design

속성구분	정의	비고
공통속성	1. 모든 상품이 공통적으로 가지는 속성정보 2. 마스터의 공통된 속성을 공통관리를 위한속성	품명, 단위, 제조원
개별속성	1. 각 상품이 가지는 개별적 고유의 속성정보 2. 마스터 각각의 상품의 특질과 관련된 정보를 관리하기 위한 속성.	용량, 크기, 재질, 당도

공통 속성의 정의는 모든 상품이 공통적으로 가지는 속성이며, 필수적인 항목이다. 예로 상품의 품명, 단위, 제조원 등을 말한다. 개별 속성의 정의는 각 상품이 가지는 개별적 고유의 속성정보를 관리하는 속성이며, 상품에 따라 선택적인 속성이다. 예로 상품의 용량, 크기, 재질, 당도 등을 말한다.

1.3 식별체계 설계

식별체계는 [표 5]과 같다. 식별 체계는 최소한의 의미부여로 마스터간의 구분과 무의미 식별 체계의 장점을 포함 한다.

표 5. 식별체계 설계
Table 5. Identity Category Design

마스터	형태
원부자재	M+□□□□□□□□□□ Material9DigitsS/N
제품/상품	P+□□□□□□□□□□ Product9DigitsS/N
공급업체	M+□□□□□□□□□□ Vendor9DigitsS/N

[표 5]에서 제품/상품 식별 체계를 형태를 보면 "P"와 "9 Digits S/N" 구성된다. "P"는 마스터 구분하기 위한 최소한의 의미미 코드이며, "9 Digits S/N"은 무의미 코드로 식별 체계로 구성한다.

2. 데이터사전 설계

데이터 사전 개발 프로세스는 [그림 3]과 같다. 우선 데이터 표준화를 위해 원칙 수립하며, 데이터 표준 기본 원칙 정의한다. 기본 원칙을 수립함으로써 표준화에 대한 방향을 정의한다. 그리고 데이터 표준 지침 작성한다. 모든 사용자들이 참고해야 하는 데이터 표준화에 대한 구체적인 지침 문서를 작성한다. 각 데이터 표준 대상에 대한 세부 지침 사항은 데이터 표준 기본 원칙에 부합되도록 작성한다.

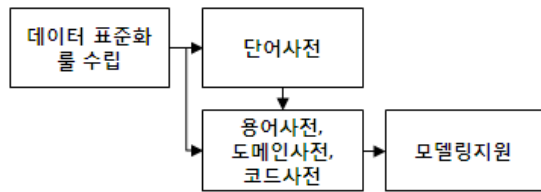


그림 3. 데이터 사전 개발 프로세스
Fig. 3. Data Dictionary Develop Process

[그림 4]은 데이터 표준화 대상이다. 용어사전 “상품구분 코드”은 단어사전에서 조합된 “상품, 구분, 코드” 단어의 조합 용어이며, 도메인 사전은 용어사전의 상품구분코드의 데이터 형태인 “Char(2)”을 나타낸다. 표준코드는 상품구분 코드인 “01:식품, 02:비식품”인 항목을 나타낸다.

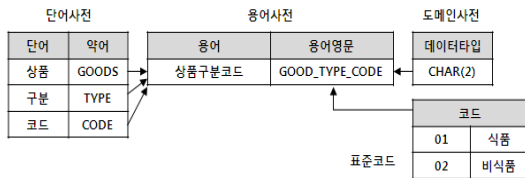


그림 4. 데이터 표준화 대상 및 관계
Fig. 4. Data Standardization Object and Relation

2.1 단어사전 설계

단어사전 정의는 데이터 사전을 통해 사용되는 모든 단어를 대상으로 한다. 추출된 단어는 단어 종류와 유형을 분류하고 용도를 고려하여 단어를 정의한다. 관리 기준은 표준성, 일반성, 대표성을 갖는다. 단어사전 작성은 단어명, 영문명, 약어, 설명 항목으로 구성되며, [표 6]과 같다.

단어 한글명은 명사로 작성하며, 영문명은 대문자 전체이름, 영문 약어는 대문자로 작성한다.

표 6. 단어사전 설계
Table 6. Word Dictionary Design

단어명	영문명	약어	설명
상품	GOODS	GD	사고파는 물품
제품	PRODUCTS	PROD	원료로 만든 것

2.2 용어사전 설계

용어사전은 단어사전을 바탕으로 용어를 구성하고, 용어사전 관리 기준은 표준성, 일반성, 업무 지향성 등으로 관리한다. 작성 형식은 용어 한글명, 영문명, 설명 형식으로 작성되며, [표 7]과 같다.

표 7. 용어사전 설계
Table 7. Term Dictionary Design

용어명	영문	설명
상품코드	GOODS CODE	사고파는 물품 코드
제품코드	PRODUCTS CODE	원료로 만든 제품 코드

용어명은 단어 사전에서 정의된 단어의 조합으로 구성된다.

2.3 도메인사전 설계

도메인사전 정의는 업무적인 용도, 사용빈도, 데이터의 물리적 특성을 고려하여 도메인을 분류하고 데이터 형태를 부여한다. 관리 기준의 표준성은 공통적으로 사용되는 속성을 대상으로 정의한다(예, 계좌번호). 유일성은 동일한 내용의 도메인이 서로 다르게 선언되지 않도록 하며, 업무 지향성은 일반화하여 정의하기 보다는 업무 특성을 반영 할 수 있도록 한다. 예로 계좌번호 ‘-’ 이용하여 표현한다. 도메인 사전의 작성은 도메인명, 정의 데이터 타입, 길이, 값, 측정단위 형식으로 작성하며, [표 8]과 같다.

표 8. 도메인사전 설계
Table 8. Domain Dictionary Design

도메인명	데이터			측정단위
	타입	길이	값	
순번	NUMBER	12	-	-
코드	VARCHAR	30	-	-

2.4 코드사전

코드사전은 용어로부터 코드를 선별하여 코드 값을 정의한다. 정의된 코드별로 오너십을 정의하여 향후 관리 할 수 있게 한다. 코드사전 관리 기준은 재사용성은 정의된 코드를 재사용하는 것이 데이터에 대한 이해력을 높이고, 코드 관리

를 용이하게 한다. 일관성은 코드는 업무 범위 내에서 가능한 유일하게 정의되어야 한다. 코드 데이터의 불일치는 문제를 발생한다. 정보 분석성은 가능한 범위의 데이터는 모드 코드 화하여 관리한다. 즉 사용자가 텍스트로 직접 입력하는 값을 최소화하고 정의된 범위 안에서 선택하도록 함으로서 정보 분석 할 때 효과적이다. 코드사전 작성은 코드명, 코드 값, 설명 형식으로 작성하며, [표 9]과 같다. 코드사전은 유의미 코드 1자리, 무의미코드 5자리로 구성한다.

표 9. 코드사전 설계
Table 9. Code Dictionary Design

코드명	코드값	설명
상품마스터	M00001	-
제품마스터	M00002	-

IV. 데이터 표준화를 통한 효율성 검증

1. 사례 연구 환경

사례 연구 환경은 [그림 5]과 같이 네트워크 개발 환경은 내·외부 망, 웹서버, 미들웨어, DB서버는 이중화 및 검색서버 구성되며, S/W 개발 환경은 윈도우 NT서버에 웹투비, 제우스, MSSQL, BizTalk 연계 서버로 구성한다.

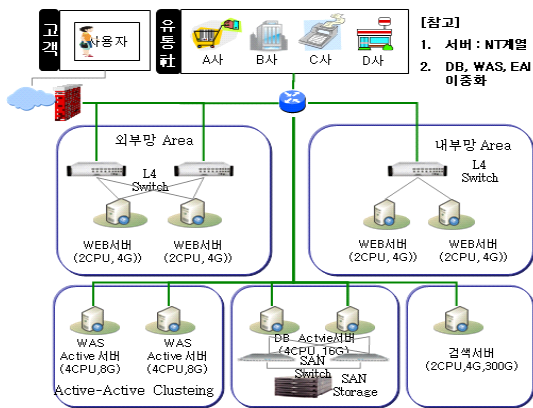


그림 5. 사례 연구 환경
Fig. 5. Empirical Study Environment

데이터 표준화의 연구 대상은 '규격 식품위주인 가공식품 및 규격 신선식품'이며, 정비 대상은 '이력데이터 2년 118,510건'이며, 데이터 대상은 '4개사의 규격 식품'과 같다.

데이터 표준화 대상 상품 중 '정비제외' 상품은 비규격 신선식품, 비식품은 운영중지 및 품명이 오류 상품, 한시적인 판매인 기획 상품. 그리고 코드 및 속성(품명, 용량 등)의 불일치로 중복처리 할 수 없는 코드를 말하며, 정비제외 한 데이터 표준화 대상은 90,978건 이다.

2. 데이터 표준화 체계의 설계

데이터 표준화 체계는 분류·속성·식별체계를 말하며, 분류 체계는 [표 10]와 같이 분류 개수는 총 1,813개 이며, 대분류 6개, 중분류 33개, 소분류 245개, 세분류 1,529 개 이다.

표 10. 분류체계 사례
Table 10. An Empirical Study of Classification

분류	대분류	중분류	소분류	세분류
개수	6	33	245	1,529
분류명	식품	가공식품	장류 식용유 축산가공품 면류	고추장 올리브유 햄 붕지라면 츄잉검 맥주
	일용품	과자, 냉과류	과자류	추잉검
	문화용품	음료및주류	주류	맥주
	내구소비재	신선식품	농산물	잡곡
	의류잡화스포츠	화장품	두발용제품	샴푸
	기타상품계	정보기기	디지털형	컴퓨터
		가전	냉각기	냉장고
		의류	여성의류	원피스
		기타상품군	기타상품류	기타상품

[표 10]의 분류체계에서 세분류가 갖는 속성체계는 [표 11]과 같이 공통속성과 개별속성으로 설계한다. 공통 속성명은 7개로 품목코드, 표준품명, 품목명, 분류코드, 등록자, 등록일, 삭제Flag 이며, 개별 속성명은 10개 이며, [표 11]과 같다.

표 11. 속성체계 사례
Table 11. An Empirical Study of Attribute Category

속성명	유형	비고
품목코드	공통속성	시스템 생성
표준품명	공통속성	상세분류의 개념
품목명	공통속성	브랜드명, 상표명, 영문명 포함
GTIN코드	개별속성	단품의 GTIN코드
규격	개별속성	분류별 개별 속성의 조합
단위	개별속성	
제조사	개별속성	
모델번호	개별속성	
분류코드	공통속성	등록시 우선택/6자리.
등록자	공통속성	시스템 생성
등록일	공통속성	시스템 생성
수정자	개별속성	시스템 생성

수정일	개별속성	시스템 생성
삭제Flag	공통속성	Y/N(Default : N)
삭제사유	개별속성	마스터 삭제사유
대체코드	개별속성	대체품목의 품목코드
이미지명	개별속성	.gif/.jpg

식별체계는 유의미와 무의미 코드로 구성되며, 유의미는 최소한의 마스터 정보를 표시한다. 예로, 상품/제품은 P(Product), 원부자재 M(Material), 공급업체 V(Vendor)의 유의미 코드이며, 9개 Digits S/N 무의미 코드로 구성된다. 공통속성은 필수이며, 개별속성은 상품에 따라 선택사항이다. 식별체계의 사례는 [표 10]에서 대분류(식품), 중분류(가공식품), 소분류(축산 가공식품), 세분류(햄)에 대한 식별체계이며, '품목코드(식별코드)'가 'P000000001'의 의미는 'P'는 상품/제품이며, '000000001'은 무의미 코드를 말한다. [표 12]과 같다.

표 12. 식별체계 사례
Table 12. An Empirical Study of Identity Category

품목코드	표준품명	품목명	규격	단위
P000000001	장조림 통조림	장조림	105g	EA
P000000002	계육 가공품	닭가슴살	150g	EA
P000000003	장조림 통조림	메추리알	140g	EA
P000000004	장조림 통조림	장조림	115g	EA

3. 데이터 사전의 설계

데이터 표준화 결과 데이터 사전인 단어사전, 도메인사전, 용어사전과 모델 데이터에 사용되는 건수는 [표 13]과 같다.

표 13. 표준데이터 및 모델데이터 사례
Table 13. An Empirical Study of Standard and Model Data

데이터 구분	데이터 사전	건수
표준데이터	단어사전	2,588
	도메인사전	72
	용어사전	2,135
모델데이터	테이블	39
	컬럼	484

단어사전 정의는 데이터 사전에 사용되는 모든 단어가 정의되어야 한다. 단어사전의 사례는 [표 14]과 같다.

용어사전은 단어사전인 [표 14]에서 정의된 단어로 구성한다. 단어사전에서 '상품(GOODS)'과 '코드(CODE)'를 구성해서 '상품코드(GOODS CODE)'라는 용어명과 영문명을 구성하며, 사례는 [표 15]과 같다.

표 14. 단어사전 사례
Table 14. An Empirical Study of Word Dictionary

단어코드	단어명	영문	약어	설명
W000001	상품	GOODS	GD	상품코드
W000002	제품	PRODUCTS	PROD	제품코드
W000003	코드	CODE	CODE	코드
W000004	구분	TYPE	TYPE	구분

표 15. 용어사전 사례
Table 15. An Empirical Study of Term Dictionary

용어코드	용어명	영문	설명
T000001	상품코드	GOODS CODE	상품 코드
T000002	제품코드	PRODUCTS CODE	제품코드
T000003	상품구분	GOODS TYPE	구분 코드

표준 데이터 사전을 적용한 데이터 모델이다. 논리적 모델은 데이터 사전인 단어, 용어사전으로 구성하며, 물리적 모델의 영문 컬럼명은 단어사전에 정의한 약어로 정의하며, 데이터 타입은 도메인 사전으로 정의한다. 데이터 모델의 사례는 [표 17]과 같다.

표 16. 데이터 모델 사례
Table 16. An Empirical Study of Data Model

컬럼명(한글)	컬럼명(영문)	타입
용어사전 코드	WORD_DICT_CD	varchar(30)
한글 명	KOR_NM	varchar(256)
영문 명	ENG_NM	varchar(256)
약어명	ABBR_NM	varchar(20)

4. 사례를 통한 효율성 검증

4.1 정량적 효과

본 논문에서 제시된 데이터 표준화가 수행된 설계 결과를 평가하기 위해, 규모 및 일정이 유사한 데이터 표준화를 수행하지 않은 L 유통 4사 상품 마스터 관리 시스템 프로젝트를(식1)에 적용하여 일관성에 대한 측정 결과는 오류 속성 개수인 T는 각각 0, 0.073(94개)이고, 데이터 품질 측정값인 Q는 1(100%), 0.927(93%) 나타난다. 결과는 [표 17]과 같이 비교하였다. 데이터 표준화를 통한 사례가 그렇지 않은 경우보다 데이터베이스 속성 설계의 일관성에 대한 데이터의 구조적 품질이 7% 향상됨을 나타낸다.

데이터 표준화 결과 [표 18]과 같다. L사의 품목 관리 마스터 거래력 2년 이내 데이터이며, 데이터 표준화 정비 대상에서 제외된 항목은 신선식품, 비식품, 기획 상품, 중복확

인 상품이다. [표 18] 과 같이 데이터 표준화 결과에서 As-is 보다 데이터 중복에 대한 품질이 24% 향상이 됨을 보여준다.

표 17. 프로젝트 비교 사례
Table 17. An Empirical Study of Project Compare

구분	사례 프로젝트	L 유통 4사 프로젝트
테이블 수	39	46
N	484	1287
T	0	0.073
Q	1	0.927

표 18. 표준화 대상상품 사례
Table 18. An Empirical Study of Standard Goods

구분	매핑 대상 (이력2年 内)	정비제외				정비 완료
		신선 식품	비 식품	기획 상품	중복 확인	
A사	59,213	13,794	436	805	353	43,825
B사	33,417	3,190	54	311	328	29,534
C사	20,997	7,762	236	62	199	12,738
D사	4,883	0	12	21	61	4,789
합계	118,510	24,746	936	999	851	90,978

4.2 정성적 효과

본 연구의 데이터 표준화를 적용하여 수행한 결과 품질 향상 측면 외에 다음과 같은 정성적 효과를 갖는다.

논리 데이터 모델링과 표준화된 테이블 설계가 자동 변경 된다. 논리적 모델은 데이터 사전인 단어, 용어사전으로 구성하고, 물리적 모델은 영문 컬럼인 단어사전에 정의한 약어, 도메인 사전으로 정의한다. 이것은 물리 속성명 변경에 대한 설계에 편의성을 제공한다. 표준화 컬럼을 통해 테이블 변경에 대한 영향도 파악을 할 수 있어 유지보수에 편의성을 제공하며, 표준 용어사전을 사용하므로 프로그램 용어 표준 활용하여 개발 표준화를 유도한다.

V. 결론

본 연구는 유통 상품의 마스터 데이터 품질 관리를 위한 데이터 표준화, 데이터 사전에 대한 설계를 제안하였다.

데이터 표준화 체계는 분류·속성·식별 체계로 작성되며, 표준화된 데이터는 데이터 사전인 단어·용어·도메인·코드사전으로 관리된다. 데이터 표준화는 의사소통의 혼란을 방지하고, 데이터의 정확성과 일관성을 유지하여 고품질의 데이터를

확보하게 된다.

본 연구에서 한계점으로는 데이터 표준화에서 유통 상품의 비규격 식품군인 비규격 신선, 비규격 즉석 식품에 대해서 표준화 대상에서 제외 된 것이며, 향후 연구 과제로 비규격 식품에 대한 데이터 표준화를 할 수 있도록 연구가 필요하다.

참고문헌

- [1] Philip Russom, "Master Data Management : Consensus-Driven Data Definitions for Cross-Application Consistency" pp.8-9, Oct. 2006. <http://tdwi.org/>
- [2] Sunho Kim, Changsoo Lee, Jehyun Lee, "Process-based e-Catalog Data Quality Management", Society for e-Business Studies, Vol. 14, No. 3, pp.39-57, Aug. 2009.
- [3] ISO 8000-1 Information Quality (draft), 2008.
- [4] National Information Society Agency, "Public Information Quality Management Manual", pp. 61,62,63,65,66, Jan. 2012., <http://www.nia.or.kr/>
- [5] Kim Sunho, Lee Changsoo "A Master Data Quality Management Framework", Entru Journal of Information Technology, Vol. 9, No.2, pp.109-121, Jul. 2010.
- [6] Sang-Hyuk Cho "Design and Implementation of a Metadata System for Financial Information Data Modeling", Journal of The Korea Society of Computer and Information, Vol 17, No.1, pp. 81-85, Jan. 2012.
- [7] KyungBae Yoon "Standardization Model Development and its Effect Analysis for Effective Available Stock Management Process of Automobile Parts Manufacturing Industry using the ERP System", Journal of The Korea Society of Computer and Information, Vol 16, No.2, pp. 279-288, Feb. 2011.
- [8] Yong Man Ko, Tae Sul Seo, SunYong Cho, "A Study on the Developing Standard Classification of the National Knowledge and Information Resources", Journal of the Korean society for library, Vol. 40, No 3, pp. 153, Sep. 2006
- [9] National Information Society Agency, "A Study

on New Digital Contents Identification Code Syntax - UCI", pp.1.17-18, Dec. 2003.,
<http://www.nia.or.kr/>

- [10] Jayoung Yang, Byoungju Choi, "A Data Quality Measuring Tool", Korean Institute of Information Scientists and Engineers, Vol 9, No.3, pp. 278-288, Jun. 2003.

저 자 소 개



송 장 섭

2000: 숭실대학교
 컴퓨터학부 학사.
 2007: 숭실대학교
 소프트웨어공학과 석사.
 2009: 숭실대학교
 컴퓨터학과 박사수료
 현 재: Dream Factory Lab
 Plant manager
 관심분야: S/W공학, 데이터 표준화
 Email : jssong@ssu.ac.kr



류 성 렬

1977: 숭실대학교 전자계산학과 학사.
 1980: 연세대학교 전자계산학과 석사.
 1996: 아주대학교 컴퓨터공학과 박사
 현 재: 숭실대학교 컴퓨터학과 교수
 관심분야: S/W공학, S/W 유지보수
 Email : syrheew@ssu.ac.kr