

Proposal of Public Data Quality Management Level Evaluation Domain Rule Mapping Model

Ha-Na Jeong*, Jae-Woong Kim**, Young-Suk Chung*

*Student, Dept. of Computer Engineering, Kongju National University, Cheonan, Korea

**Professor, Dept. of Software Engineering, Kongju National University, Cheonan, Korea

*Lecturer, Dept. of Computer Engineering, Kongju National University, Cheonan, Korea

[Abstract]

The Korean government has made it a major national task to contribute to the revitalization of the creative economy, such as creating new industries and jobs, by encouraging the private opening and utilization of public data. The Korean government is promoting public data quality improvement through activities such as conducting public data quality management level evaluation for high-quality public data retention. However, there is a difference in diagnosis results depending on the understanding and data expertise of users of the public data quality diagnosis tool. Therefore, it is difficult to ensure the accuracy of the diagnosis results. This paper proposes a public data quality management level evaluation domain rule mapping model applicable to validation diagnosis among the data quality diagnosis standards. This increases the stability and accuracy of public data quality diagnosis.

▶ **Key words:** Public data, Data quality, Quality improvement, Standards for public data quality diagnosis, Quality diagnosis, Open data

[요 약]

정부는 공공데이터의 민간 개방, 활용을 장려함으로써 신산업, 일자리 창출 등 창조경제 활성화에 기여하는 것을 주요 국정과제로 삼고 있다. 그리고 고품질 공공데이터 보유를 위해 공공데이터 품질관리 수준평가 진행 등의 활동을 통해 공공데이터 품질 향상을 도모하고 있다. 그러나 품질진단 도구 사용자의 데이터 전문성, 이해도에 따라 공공데이터 품질관리 수준평가 결과에 격차가 발생하기 때문에 진단 결과의 정확성을 보장하기 어렵다. 본 논문은 데이터 이해도가 낮은 사용자의 진단 결과에 대한 정확성을 보장하기 위해 데이터 품질진단 기준 중 유효성 진단에 적용 가능한 공공데이터 품질관리 수준평가 도메인규칙 매핑 모델을 제안하였다. 또한 모델에 실제 데이터를 적용한 결과 공공데이터 품질진단의 안정성과 정확성을 높이는 것을 확인하였다.

▶ **주제어:** 공공데이터, 데이터품질, 품질개선, 공공데이터 품질진단 기준, 품질진단, 개방데이터

• First Author: Ha-Na Jeong, Corresponding Author: Jae-Woong Kim

*Ha-Na Jeong (konghanaj@gmail.com), Dept. of Computer Engineering, Kongju National University

**Jae-Woong Kim (jykim@kongju.ac.kr), Dept. of Software Engineering, Kongju National University

*Young-Suk Chung (merope@kongju.ac.kr), Dept. of Computer Engineering, Kongju National University

• Received: 2022. 09. 22, Revised: 2022. 11. 18, Accepted: 2022. 12. 12.

• This paper is an extension of the paper ("Proposal of diagnosis rule mapping model to support public data quality diagnosis") presented at the 66th Summer Conference of the Korea Computer Information Society in 2022.

I. Introduction

정부는 공공데이터의 민간 개방, 활용을 장려하기 위해 각 기관에서 공공데이터를 수집하여 무료로 제공하고 있다. 이러한 공공데이터의 개방을 위해 공공데이터 포털을 구축하였고 신산업, 일자리 창출 등을 도와 창조경제 활성화에 기여하는 것을 주요한 국정과제로 추진하고 있다 [1-3]. 정부는 데이터를 축적, 분석하여 인공지능/네트워크/데이터의 생태계를 강화하기 위한 데이터 댐 사업을 진행하며 공공데이터 개방에 박차를 가하고 있다. 그러나 공공데이터의 품질관리 미흡으로 인한 데이터의 저품질 문제로 공공데이터의 활용이 저하되는 문제가 발생하였다 [4-5].

데이터 품질은 '특정한 상황에서 데이터를 사용할 때 암시적, 명시적 요구를 충족시킬 수 있는 데이터의 능력'을 뜻한다[6]. G8 오픈 데이터 헌장(G8 Open Data Charter)은 개방 데이터의 품질 향상을 위하여 데이터의 지속적인 유지, 관리가 필요하다고 강조한다[7-8]. 정부는 이러한 공공데이터의 품질 향상을 위해 노력을 기울이고 있다. 행정안전부는 각 기관에 데이터 품질관리 가이드를 제공하여 공공기관의 데이터 품질관리 가이드라인을 제시하고 있으며, 한국지능정보화진흥원은 2016년부터 매년 품질관리 수준평가를 진행하여 공공데이터의 품질개선을 위한 노력을 기울이고 있다[9-11]. 그러나 사용자가 데이터품질 진단규칙을 직접 지정해야 하는 품질진단 도구의 특성상, 품질진단을 수행하는 담당자 개인별 품질진단 개념과 데이터 대한 이해도 격차로 인하여 진단결과의 신뢰성이 떨어지고 진단결과의 정확성을 보장할 수 없는 문제가 발생한다[12]. 동일한 DB를 진단하더라도 데이터에 대한 이해도가 낮을수록 데이터별 적절한 진단규칙을 설정하는 것에 어려움이 있고, 그로 인해 고득점을 받기 불리하기 때문이다.

본 논문은 데이터 품질진단 결과의 신뢰성을 높이기 위해 공공데이터 공통표준용어 데이터를 기반으로 공공데이터 품질진단 도구 사용자의 진단규칙설정을 하여, 공공데이터 품질진단 결과의 정확성을 높이는 공공데이터 품질관리 수준평가 도메인규칙 매핑 모델을 제안한다. 본 논문을 다음과 같이 구성하였다. 2장에서 관련 연구에 대해 논의한다. 3장에서 본 논문에서 제안한 공공데이터 품질관리 수준 평가 도메인규칙 매핑 모델을 논의한다. 4장에서 데이터를 적용하여 제안된 모델을 검증한다. 5장을 결론 및 향후 연구과제에 대해 논의한다.

II. Preliminaries

1. Related works

1.1 Public data quality diagnosis criteria

품질진단 기준이란 데이터가 사용되는 목적을 충족시키기 위해 품질점검을 통하여 관리되어야 하는 진단 기준이다. 2021년 공공데이터 품질관리 수준평가의 품질진단 기준은 크게 정합성 진단, 완결성 진단, 유효성 진단으로 나뉜다. 정합성 진단은 데이터의 특성별로 진단기준이 달라지고, 완결성 진단은 오류율 산정 시 제외되는 기준이므로 본 논문에서는 다루지 않는다. Table 1. 은 2021년 공공데이터 품질관리 수준평가에서 지정한 품질진단 기준 중 유효성 진단 항목을 정리한 것이다[13].

Table 1. Public data quality indicator

Validation diagnosis types	Explanation
Date domain	Measurement of data errors where date data values are out of valid range or format is out of standard
Number domain	Error measurements when number generation rules are violated
Whether domain	Measurement of data errors outside the range of valid values
Code domain	Measurement of data errors other than code values defined by the source DB as standard
Amount domain	Measurement of data errors that contain characters other than amount data
Quantity domain	Measurement of data errors with characters other than quantity data
Percentage domain	Measurement of data errors that contain non-rate characters

유효성 진단은 도메인규칙 진단이라고 지칭하기도 하며 데이터베이스 테이블의 속성값으로 저장된 값이 속성 도메인이 부합하는 유효한 값인가를 진단한다. 예를 들어 날짜 도메인으로 지정된 '저장연도' 컬럼 내에 'YYYY' 형식 외에 'YYYYMM' 등의 다른 형식 데이터가 존재한다면 데이터 무결성을 위배한 데이터이므로 오류데이터로 판별한다. 수량 도메인으로 지정된 컬럼 내에 영문, 한글 등 숫자가 아닌 데이터가 존재한다면 오류데이터로 판별한다. 여부 도메인으로 지정된 컬럼 내에 Y/N, 0/1 등 지정된 두 종류 외의 다른 데이터가 존재한다면 오류데이터로 판별한다. 이러한 데이터 무결성을 위배한 데이터들은 해당 데이터를 사용하는 서비스의 에러로 이어지기 때문에 고품질의 데이터를 유지하기 위해서는 유효성 진단이 이루어져야 한다.

1.2 Public data common standard term

행정안전부는 ‘공공데이터의 제공 및 이용 활성화에 관한 법률’ 제 23조 공공데이터의 표준화 ‘공공기관의 데이터베이스 표준화 지침’에 따라, 공공데이터를 누구나 같은 의미로 이해하고 같은 방식으로 사용할 수 있도록 공공기관이 데이터 구축 시 공통으로 적용하여야 하는 공통표준용어를 정의하였다[14]. 이는 2020년 제정되었으며, 주기적으로 검토되어 공통표준의 추가 제정 또는 폐기가 이루어지고 있다. ‘공공기관의 데이터베이스 표준화 지침’ 제8조(공공데이터베이스 표준 수립 및 적용)에 의거하여 공공기관이라면 신규 공공데이터베이스 구축 시 바로 적용해야 하며 기존의 운영 중인 공공데이터베이스의 경우에도 신규 전면 구축 또는 재구축 업무 추진 시에 적용해야 한다. 즉, 공공 데이터베이스라면 행정안전부에서 제공하는 공공데이터 공통표준용어, 공통표준도메인, 공통표준단어를 최대한 따르도록 규정되어 있다. Table 2. 는 행정안전부에서 정의한 4차 제정 공공데이터 공통표준용어로 공통표준용어명, 설명, 영문약어명, 공통표준도메인명, 허용값 등을 관리한다.

Table 2. 4th enacted public data common standard term

Item	Example
Common standard term name	Review status
Explanation	Whether any facts or contents have been analyzed
English abbreviation name	RVW_YN
Common Standard domain name	WhetherC1
Allowed value	Y:Yes N:No
Storage Format	1 character
Form of expression	Y or N

Table 3.은 행정안전부에서 정의한 4차 제정 공공데이터 공통표준도메인으로 공통표준도메인명, 설명, 그룹명, 분류명 등을 관리한다[15].

Table 3. 4th enacted public data common standard domain

Item	Example
Common standard domain name	WhetherC1
Explanation	Code to distinguish yes and no
Group name	Code
Classification name	Whether
Data Type	Char
Data length	1
Storage Format	1 character

III. The Proposed Scheme

1. Proposed model diagram

Fig. 1. 은 본 논문에서 제안하는 공공데이터 품질관리 수준평가 도메인규칙 매핑모델의 시스템 구성도이다.

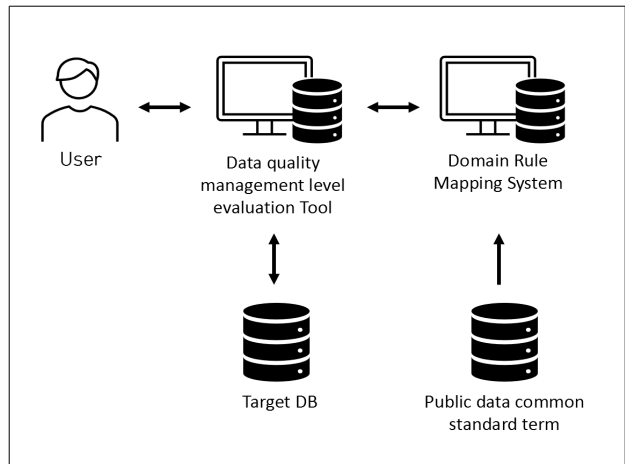


Fig. 1. Diagram of Public data quality management level evaluation domain rule mapping model

사용자가 공공데이터 품질관리 수준평가 도구를 사용하여 공공데이터 품질관리 수준평가의 진단으로 선정된 진단대상 DB를 진단할 때 도메인규칙 매핑 시스템이 활용되는 구조이다. 이러한 구조를 통해 사용자에게 도메인규칙의 진단규칙을 제안함으로써 사용자의 진단대상 DB에 대한 낮은 이해도로 인한 진단규칙 설정 오류 발생 확률을 줄인다.

도메인규칙 매핑 시스템은 공공데이터 공통표준사전 정보를 입력받아 도메인규칙 진단 선택에 활용한다. 또한 품질진단 도구가 수집한 진단대상 DB의 샘플 데이터를 공유받아 컬럼의 상세 진단규칙의 설정에 활용한다. 이렇게 도메인규칙 매핑 시스템이 설정한 진단규칙을 공공데이터 품질관리 수준평가 도구에 적용하여 진단대상 DB를 진단한다.

2. Proposed model process

Fig. 2. 는 본 논문에서 제안하는 공공데이터 품질관리 수준평가 도메인규칙 매핑모델 프로세스이다.

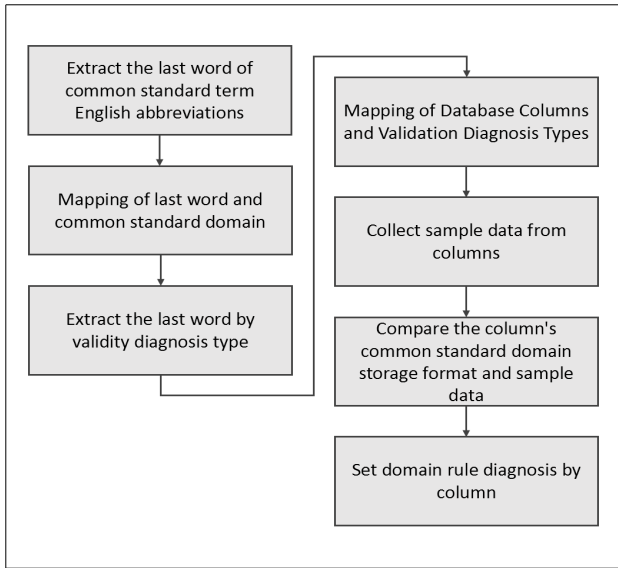


Fig. 2. Process of public data quality management level evaluation domain rule mapping model

프로세스의 단계별 상세 내용은 다음과 같다.

첫 번째, 공공데이터 공통표준용어의 영문약어명 마지막 단어를 추출한다. 이때 영문약어명의 마지막 단어를 추출하는 이유는 표준용어 특성상 마지막 단어가 표준도메인을 지정할 수 있는 의미를 담기 때문이다. 예를 들어, Table 2. 의 검토여부 공통표준용어의 경우, 영문약어명 RVW_YN의 마지막 단어인 YN이 추출된다. 이 YN이란 단어는 여부 도메인을 지정할 수 있는 ‘여부’라는 의미를 지니고 있다.

두 번째, 추출한 영문약어명 마지막 단어와 공공데이터 공통표준도메인을 매핑한다. 공공데이터 공통표준용어의 관리 항목에는 공통표준도메인명이 존재하므로 이를 활용하여 매핑한다.

세 번째, 데이터 품질진단 기준 중 하나인 유효성 진단의 진단유형 별로 공통표준용어의 마지막 단어를 추출한다. 4차 제정 공공데이터 공통표준용어에서 추출한 진단유형 별 마지막 단어 데이터는 Table 4. 와 같다.

Table 4. List of last words by type of validation diagnosis

Validation diagnosis types	Last word
Date domain	YMD, DT(DATE), TM, YR(YEAR), HR, YM, MM, BRDT, FYR
Number domain	ACTNO, NO, BNO, BMNO, BSNO, TELNO, OZIP, ZIP, FXNO, RRNO, CRNO, MNO, SNO, BRNO, PNO, FRNO, DLN, SN, VHRNO, VIN
Whether domain	YN
Code domain	CD(CODE)
Amount domain	AMT, OALP, OVPMNT, FFNLG, INCTX, PJTCO, FEE, LBRCO, CPTL, RSDTX, MCPBND, GRAMT, TPCO
Quantity domain	CNT, BDAR, LOT, CMAR, SVLF, SIAR, LDAR, FCAR, NOCS, AGE, LAT, NOPE, XUAR, NOFL, SZ, GFA
Percentage domain	BDCVRT, FART

이때 Table 4. 의 괄호 표기는 빈번하게 사용되는 축약형 단어의 원단어로, 추가적으로 허용하였다.

네 번째, 진단대상 DB의 컬럼명 마지막 단어와 공통표준용어의 마지막 단어를 비교해 진단대상 DB 컬럼의 유효성진단 유형을 매핑한다.

다섯 번째, 공공데이터 품질관리 수준평가 도구를 활용하여 진단대상 DB 컬럼의 샘플데이터를 수집한다.

여섯 번째, 앞서 매핑해둔 진단대상 DB 컬럼 공통표준도메인의 저장형식 데이터와 진단대상 DB 컬럼의 샘플데이터를 비교한다.

일곱 번째, 여섯 번째 프로세스에서 비교한 데이터를 활용하여 진단대상 DB 컬럼별 도메인규칙 진단을 설정한다. 이때 기본적으로 진단대상 DB 컬럼의 샘플데이터를 이용하여 설정하되, 샘플데이터 내의 데이터 유형별 분포도가 동일하다면 공통표준도메인의 저장형식을 따른다. 예를 들어 개정일자 컬럼의 샘플데이터 유형이 ‘YYYYMMDD’와 ‘YYYY-MM-DD’의 분포도가 동일하게 30.1%이고, 개정일자 컬럼의 공통표준도메인인 ‘연월일C8’의 저장형식이 ‘YYYYMMDD’라면 ‘YYYYMMDD’ 형식으로 날짜도메인 진단을 설정한다. 이는 ‘공공데이터의 제공 및 이용 활성화에 관한 법률’ 제 23조의 내용에 따라 공공 데이터베이스가 공통표준도메인을 최대한으로 따라야 하기에 공통표준도메인을 우선하였다.

3. Database structure

Fig. 3. 은 도메인규칙 매핑 시스템의 데이터베이스 구조이다.

COMM_STND_TERM			
No.	Field	Description	Data Type
1	STND_TERM_ID (PK)	Standard term ID	Varchar(15)
2	STND_TERM_NM	Standard term name	Varchar(200)
3	STND_TERM_ENG	Standard term english abbreviation	Varchar(500)
4	LAST_WD	Last word	Varchar(200)
5	STND_DMN_ID (FK)	Standard domain ID	Varchar(15)

COMM_STND_DMN			
No.	Field	Description	Data Type
1	STND_DMN_ID (PK)	Standard domain ID	Varchar(15)
2	STND_DMN_NM	Standard domain name	Varchar(200)
3	GRP_NM	Group name	Varchar(200)
4	CLASS_NM	Classification name	Varchar(200)
5	STG_FORM	Storage format	Varchar(500)

TARG_COL_LST			
No.	Field	Description	Data Type
1	COL_ID (PK)	Column ID	Varchar(15)
2	DB_NM	Database name	Varchar(200)
3	SCH_NM	Schema name	Varchar(200)
4	TBL_NM	Table name	Varchar(500)
5	COL_NM	Column name	Varchar(500)
6	COL_LAST_WD	Column last word	Varchar(200)

TARG_COL_DATA			
No.	Field	Description	Data Type
1	OBJ_ID (PK)	Object ID	Varchar(15)
2	COL_ID (FK)	Column ID	Varchar(15)
3	COL_DATA	Column data	Varchar(4000)
4	DISTR_RATE	Distribution rate	Number(3,1)

DMN_RULE_COL_MAP			
No.	Field	Description	Data Type
1	OBJ_ID (PK)	Object ID	Varchar(15)
2	COL_ID (FK)	Column ID	Varchar(15)
3	VAL_TYP_NM	Validation diagnosis type name	Varchar(300)
4	DIAG_TYP	Diagnostic type	Varchar(300)

Fig. 3. Domain rule mapping system database tables

도메인규칙 매핑 시스템은 공통표준용어(COMM_STND_TERM), 공통표준도메인(COMM_STND_DMN), 진단대상 컬럼목록(TARG_COL_LST), 진단대상컬럼데이터(TARG_COL_DATA), 진단대상컬럼-도메인규칙매핑(DMN_RULE_COL_MAP) 5개의 테이블로 구성된다.

이 중 공통표준용어 테이블, 공통표준도메인 테이블, 진단대상컬럼목록 테이블, 진단대상컬럼데이터 테이블은 진단

단 도구와의 연계를 통해 가져온 각각의 데이터가 저장되는 테이블이다.

공통표준용어 테이블과 공통표준도메인 테이블은 공통표준용어와 공통표준도메인이 저장되는 테이블로, 표준도메인ID(STND_DMN_ID) 값을 기준으로 공통표준도메인 테이블을 참조하여 공통표준용어의 도메인정보를 가져온다.

진단대상컬럼목록 테이블은 진단대상DB의 컬럼에 대한 데이터가 저장되는 테이블이다.

진단대상컬럼데이터 테이블은 진단대상DB의 각 컬럼의 샘플데이터가 저장되는 테이블로, 컬럼ID(COL_ID) 값을 기준으로 진단대상컬럼목록 테이블을 참조하여 해당 컬럼에 대한 정보를 가져온다.

진단대상컬럼-도메인규칙매핑 테이블은 각 컬럼의 도메인규칙을 저장하는 테이블로, 컬럼ID(COL_ID) 값을 기준으로 진단대상컬럼목록 테이블을 참조하여 해당 컬럼에 대한 정보를 가져온다.

또한, 공통표준용어 테이블과 진단대상컬럼목록 테이블은 참조관계는 아니지만, 공통표준용어 테이블의 'LAST_WD' 컬럼과 진단대상컬럼목록 테이블의 'COL_LAST_WD' 컬럼을 비교하여 데이터를 가져온다.

IV. Evaluation

4차 제정 공공데이터 공통표준용어를 기반으로 본 논문에서 제안한 공공데이터 품질관리 수준평가 도메인규칙 매핑모델을 샘플DB에 적용하여 결과를 도출하였다. 샘플 DB 데이터는 공공기관에서 주로 사용되는 용어를 공공데이터포털, 공공기관 사이트 등을 참고하여 생성하였다. 모델 적용 시 컬럼명의 마지막 문자열이 숫자로 이루어진 컬럼의 경우는 예외처리를 하여 숫자를 삭제하고 모델을 적용하였다. 샘플DB의 컬럼 개수는 3849건이며, 이 중 유효성진단 대상 컬럼의 개수는 2602건이다. 즉, 샘플DB에서 유효성진단 대상 컬럼은 53.5%이다. 이것은 개인정보 등 보안 문제 가능성이 있는 컬럼과 이력성 컬럼을 유효성진단 대상에서 제외한 결과이다. 샘플DB의 경우 '작성일자(WRT_DD)', '대상여부(TGT_YN)' 등은 유효성진단 대상 컬럼으로 분류되었고 '주민등록번호일련번호(JNO_SNO)' 등은 유효성진단 대상 제외 컬럼으로 분류되었다.

Table 5. 는 샘플DB를 모델에 적용한 결과이다.

Table 5. Model application result of sample DB

Diagnostic target Y/N	Cnt	Mapping Y/N	Cnt	suit-ability Y/N	Cnt
Y	2602	Mapped	1751	Suitable	1509
				Unsuitable	242
		Not mapped	851	-	-
N	1247	-	-	-	-
Total	3849	-	-	-	-

유효성진단 대상 컬럼 2062건 중 공공데이터 품질관리 수준평가 도메인규칙 매핑모델 실행 결과 유효성진단 유형이 매핑된 컬럼은 1751건으로, 84.9%이다. 매핑되지 않은 851개 컬럼은 매핑 기준이 되는 진단유형 별 마지막 단어와 일치하지 않는 컬럼명을 가져 유효성진단 매핑에 실패하였다. 즉, 진단유형 별 마지막 단어 데이터가 충분하지 않은 것이 유효성진단 매핑 실패의 원인이다. 샘플DB의 경우 작성일자(WRT_DD)의 마지막 단어인 'DD'는 Table 4.의 마지막 단어에 포함되어있지 않기 때문에 유효성진단 매핑에 실패하였다. Table 6.은 유효성진단 유형이 매핑된 컬럼 1751건의 유효성진단 유형별 건수이다.

Table 6. Model application detailed result of sample DB

Validation diagnosis types	Count
Date domain	598
Number domain	326
Whether domain	378
Code domain	242
Amount domain	67
Quantity domain	140
Percentage domain	0

그러나 코드 도메인 진단의 경우 각 코드 데이터가 있어 야만 진단규칙 설정이 가능하다. 예를 들어 '국가코드'의 경우 'KR', 'US', 'JP' 등의 '국가코드'를 이루는 데이터가 있어야 한다. 수집한 샘플데이터 정보만으로는 진단규칙을 설정할 수 없으므로 품질진단을 수행할 수 없다. 이를 해결하기 위해서는 위의 예시와 같은 코드 데이터를 추가적으로 수집해야 한다. 그러나 본 논문의 모델은 공공데이터 품질관리 수준평가 도구에서 송신받은 데이터 외에 추가적인 데이터를 수집하지 않으므로 코드 데이터를 수집하지 않는다. 즉, 본 논문의 모델에 적합하지 않은 데이터이므로 코

드 도메인이 매핑된 컬럼 242건은 제외하여 결과를 도출하였다. 그러므로 유효성진단 유형이 매핑된 컬럼 1751건에서 코드도메인 매핑 컬럼 242건을 제외한 1509건이 본 논문의 모델에 성공적으로 적용된 데이터 건수이다.

Table 7.은 본 논문의 모델 적용 결과 유효성진단 유형이 매핑된 건 중 코드도메인 외의 유효성진단 유형이 매핑된 컬럼 1509건의 유효성진단 유형별 건수이다.

Table 7. Validity diagnosis type match/mismatch list

Validation diagnosis types	Match Count	Mismatch Count
Date domain	589	9
Number domain	216	110
Whether domain	378	0
Code domain	-	-
Amount domain	63	4
Quantity domain	140	0
Percentage domain	0	0
Total by case	1386	123
Total	1509	

컬럼의 유효성진단 유형과 일치하게 매핑된 건과 불일치하게 매핑된 건을 나누어 표기하였다. 1509건의 컬럼 중 매핑이 일치된 건은 총 1386건, 불일치된 건은 총 123건이다. 즉 본 논문에 적용된 데이터 중 유효성진단 대상 컬럼 2602개 중 모델에 적합하지 않은 데이터인 코드 도메인 유효성진단 유형이 매핑된 데이터 242건을 제외한 2360건을 본 논문의 모델에 적합한 데이터로 판단하였다. 이에 따라 2360건 중 1509건, 즉 63.9%를 성공률로 도출하였다.

V. Conclusions

본 논문은 공공데이터 품질진단의 안정성과 정확성을 높이기 위하여 데이터 품질진단 기준 중 유효성 진단에 적용 가능한 공공데이터 품질관리 수준평가 도메인규칙 매핑 모델을 제안했다. 해당 모델은 공통표준용어를 기반으로 도메인규칙을 매핑하여 매핑 기준 데이터에 대한 신뢰성을 제공하는 것이 특징이다. 제안된 모델에 샘플DB를 적용하였고 그 결과 63.9%로 모델 성공률을 도출하였다. 공공데이터 품질관리 수준평가를 수행하는 담당자가 데이

터베이스에 대한 전문지식이 높지 않은 일반 공무원인 경우가 대다수인 것을 고려하면 63.9%의 성공률을 보이는 공공데이터 품질관리 수준평가 도메인규칙 매핑모델은 공공데이터 품질진단 결과의 정확성 향상에 도움이 된다. 그러나 유효성진단 유형이 매핑되지 않은 것은 공공데이터 공통표준용어 데이터 만으로는 유효성진단 유형 구분을 위한 데이터로 부족하기 때문이다. 추후 연구과제로 유효성진단 유형을 구분하기 위한 데이터에 대하여 연구가 진행되어야 할 것이다.

REFERENCES

- [1] GePark, CjKim, "Quality Characteristics of Public Open Data", The Korea Society of Digital Policy & Management, Vol. 13, No. 10, pp. 135-146, 2015. DOI: doi.org/10.14400/JDC.2015.13.10.135
- [2] DjKim, HjKim, CuSong, JwYang, HlKim, "Methods for Utilising Local Government's Public Data Released to The Public Data Portal", Digital Contents Society, Vol. 22, No. 3, pp. 445-452, March 2021. DOI: dx.doi.org/10.9728/dcs.2021.22.3.445
- [3] HcKim, GyGim. "A Study on Public Data Quality Factors Affecting the Confidence of the Public Data Open Policy" Korea Society of IT Services, Vol 14, No 1, pp. 53-68, 2015. DOI: doi.org/10.9716/KITS.2015.14.1.053
- [4] Ministry of Government Administration and Home Affairs, "Completion report for diagnosis and improvement of public data quality", 2014.
- [5] ShPark, khLee, ayLee. "An Empirical Study on the Effects of Source Data Quality on the Usefulness and Utilization of Big Data Analytics Results" Korea Data Strategy Society, Vol 24, No 4, pp. 197-214, 2017. DOI: doi.org/10.21219/jitam.2017.24.4.197
- [6] Rafique, Irfan, et al. "Information quality evaluation framework: Extending ISO 25012 data quality model." International Journal of Computer and Information Engineering Vol. 6, No. 5, pp. 568-573, 2012. DOI: doi.org/10.5281/zenodo.1072956
- [7] G8 Open Data Charter. G8 Open Data Charter and Technical Annex, www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex.
- [8] D'Aquin, Mathieu, Dietze, et al. "Building the open elements of an open data competition" D-Lib Magazine Vol. 20, No. 5/6, 2014. DOI: 10.1045/may2014-daquin
- [9] National Information Society Agency, "Open Government Data Quality Management Manual v2.0", National Information Society Agency, 2018.
- [10] Ministry of the Interior and Safety. Public Data Provision and Management Practical Manual, https://www.mois.go.kr/fri/bbs/type001/commonSelectBoardArticle.do?bbsId=BBSMSTR_0000000012&ntfId=87981
- [11] Public Data Portal, https://me2.kr/euxsj.
- [12] HjKim, HlKim, "Method for improving public data quality applying the organization code of standard code for administration", Digital Contents Society, Vol. 23, No. 3, pp. 481-488, 2022. DOI: 10.1016/j.giq.2016.02.001
- [13] National Information Society Agency, "2021 Public Data Quality Management Level Assessment Manual", 2021
- [14] Act on Promotion of Provision and Use of Public Data, https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률.
- [15] Information on the 4th enactment of common standard terms for public data, https://www.data.go.kr/bbs/ntc/selectNotice.do?originId=NOTICE_0000000002403.

Authors



Ha-Na Jeong received the B.S., M.S. degrees in Computer Science Engineering from Kongju National University, Cheonan, in 2019, 2021 respectively. She is currently pursuing a Ph.D. degree in Computer Science Engineering from Kongju National University, Cheonan. She is interested in database, data quality, quality diagnosis.



Jae-Woong Kim received the bachelor's degree and the M.S. degree in the Department of Computer Engineering from the Jungang University in 1983 and 1988, respectively. He received the Ph.D. degree in

the Department of Computer Engineering from Daejeon University in 2000. He has been a professor in the Department of Computer Engineering at Kongju National University since 1992. His current research interests include software engineering.



Young-Suk Chung received the M. S. degree in Multimedia Engineering from Kongju national university, in 2009. Ph. D degree in Computer Engineering from Kongju national university in 2013.

He is currently an adjunct professor in Daejeon Health Sciences College. He is interested in Big data, Cloud computing, Simulation, A.I and Predictive modeling.