

## 공공데이터 품질진단 지원을 위한 진단규칙 매핑모델 제안

정하나<sup>0</sup>, 김재웅<sup>\*\*</sup>, 이윤열<sup>\*\*\*</sup>, 채의근<sup>\*\*\*</sup>, 정영석<sup>\*</sup>

<sup>0</sup>공주대학교 대학원 컴퓨터공학과,

<sup>\*</sup>공주대학교 대학원 컴퓨터공학과,

<sup>\*\*</sup>공주대학교 소프트웨어학과,

<sup>\*\*\*</sup>공주대학교 컴퓨터공학과

e-mail: konghanaj@gmail.com<sup>0</sup>, jwkim@kongju.ac.kr<sup>\*\*</sup>,  
{alphaone, ygchae}@kongju.ac.kr<sup>\*\*\*</sup>, merope@kongju.ac.kr<sup>\*</sup>

## Proposal of diagnosis rule mapping model to support public data quality diagnosis

Ha-Na Jeong<sup>0</sup>, Jae-Woong Kim<sup>\*\*</sup>, Yun-Yeol Lee<sup>\*\*\*</sup>, Yi-Geun Chae<sup>\*\*\*</sup>, Young-Suk Chung<sup>\*</sup>

<sup>0</sup>Dept. of Computer Engineering, Kongju National University,

<sup>\*</sup>Dept. of Computer Engineering, Kongju National University,

<sup>\*\*</sup>Dept. of Software, Kongju National University,

<sup>\*\*\*</sup>Dept. of Computer Engineering, Kongju National University

### ● 요약 ●

정부는 공공데이터 개방을 통해 신산업, 일자리 창출 등 경제 활성화를 위한 도구로 활용하는 것을 목표로 한다. 정부는 고품질의 공공데이터 보유를 위하여 품질 개선 활동을 통해 공공데이터 품질 향상을 진행하고 있다. 그러나 공공데이터 품질관리 수준 진단을 진행하는 담당자의 데이터에 대한 전문성과 이해도에 따라 품질진단 결과에 격차가 발생하여 진단 결과의 신뢰성을 보장하기 어렵다. 본 논문은 공공데이터의 원활한 품질진단 지원을 위해 품질진단규칙 매핑 모델을 제안하여 공공데이터 품질진단의 안정성과 신뢰성을 높인다.

**키워드:** 공공데이터(public data), 데이터품질(data quality), 품질진단(quality diagnosis), 품질개선(quality improvement), 공공데이터 품질평가 지표(public data quality indicator)

## I. Introduction

대한민국 정부는 1980년대부터 다양한 공공데이터를 수집, 개방하고 있다. 이러한 공공데이터의 개방은 공공데이터의 재사용을 통하여 신산업, 일자리 창출 등을 도와 창조경제 활성화에 기여하는 것을 목표로 가진다[1][2]. 정부는 이러한 목표를 달성하기 위하여 2013년 공공데이터 법을 제정하였고 최근 데이터 댐 사업을 진행하며 공공데이터 개방에 박차를 가하고 있다. 그러나 공공데이터의 품질관리 미흡으로 인하여 공공데이터의 활용성이 저하되는 문제가 발생하였다[3]. 이를 해결하기 위하여 행정안전부가 공공기관 데이터 품질관리 가이드를 제공하고, 한국지능정보화진흥원이 데이터품질진단 도구를 배포하는 등 공공데이터의 품질개선을 위한 노력을 기울이고 있다[4]. 그러나 사용자가 데이터품질진단 규칙을 직접 지정해야하는 품질진단 도구의 한계와 도구를 사용하는 담당자의 데이터에 대한 이해도 격차로 인하여 진단결과의 신뢰성이 떨어진다.

본 논문은 공공데이터의 품질진단 도구 사용자의 데이터 이해도 격차가 데이터품질진단 결과에 끼치는 영향을 줄여 공공데이터 품질에 대한 신뢰성을 높일 수 있는 공공데이터 품질 진단 규칙 매핑 모델을 제안한다.

## II. Preliminaries

### 1. Related works

#### 1.1 공공데이터 품질평가 지표

품질평가 지표란 데이터 사용 목적을 충족시키기 위해 지속적으로 품질점검을 통해 관리되어야 하는 진단 기준이다. 다음 Table 1. 은 공공데이터 품질관리 수준진단 평가에서 지정한 공공데이터 품질평가 지표 중 유효성진단 항목을 정리한 것이다[5].

Table 1. Public data quality indicator

Sortation	quality indicator	Explanation
Validity diagnosis	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

유효성진단 품질평가 지표는 데이터베이스 테이블의 컬럼에 저장된 값이 컬럼의 의미상 유효한 값인지를 진단한다. 예를 들어 수량 도메인으로 지정된 컬럼인데 데이터가 음수라면 유효하지 않은 값이라고 판단한다.

### III. The Proposed Scheme

다음 Fig. 1. 은 본 논문에서 제안하는 공공데이터 품질 진단 규칙 매핑모델의 프로세스이다.

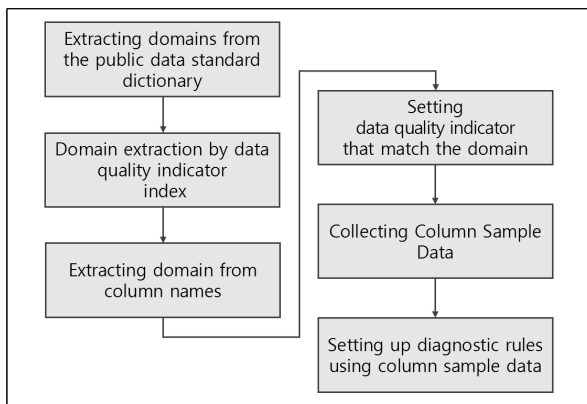


Fig. 1. Process of Diagnostic Rule Mapping Model

첫 번째, 진단규칙 매핑모델에서는 공공데이터에 대한 진단규칙 매핑 데이터 추출을 위하여 정부에서 제공하는 공공데이터 표준사전 데이터를 이용한다.

두 번째, 공공데이터 표준사전 데이터에서 도메인을 추출하여 공공데이터 품질평가 지표별로 분류한다. 세 번째, 진단대상 컬럼에서 도메인을 추출해낸다. 네 번째, 앞서 분류한 데이터를 이용하여 이전 단계에서 진단대상 컬럼에서 추출해낸 도메인의 품질평가 지표를 매핑한다. 다섯 번째, 진단대상 컬럼의 샘플데이터를 추출하여 데이터 형식에 대해 수집한다. 마지막으로 진단대상 컬럼의 품질평가 지표 매핑 정보와 샘플 데이터를 이용하여 진단규칙을 지정한다.

### IV. Conclusions

본 논문은 정부에서 제공하는 공공데이터 품질평가 지표와 공공데이터 표준사전을 활용하여 공공데이터 품질진단 지원을 위한 진단규칙 매핑모델을 제안하였다. 제안된 모델을 활용할 경우 공공데이터 품질진단 결과의 안전성과 신뢰성을 높일 것으로 기대된다. 향후 제안된 모델을 구현하여 공공데이터 품질평가의 안정성과 신뢰성 향상에 대한 연구를 진행할 예정이다.

### REFERENCES

- [1] Go-Eun Park, Chang-Jae Kim, "Quality Characteristics of Public Open Data", The Korea Society of Digital Policy & Management, Vol. 13, No. 10, pp. 135-146, 2015.
- [3] Dongjun Kim, Hyeonji Kim, Chaeun Song, Jiwoo Yang, Haklae Kim, "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.
- [3] Ministry of Government Administration and Home Affairs, "Completion report for diagnosis and improvement of public data quality", 2014.
- [4] Hyeonji Kim, Haklae Kim, "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, March 2022.
- [5] National Information Society Agency, "Evaluation Manual for Quality Management Level of Public Data", 2021