

인공지능 교육을 위한 데이터셋 아카이브 설계

이세훈*, 노예원^o, 노연수*

*인하공업전문대학 컴퓨터시스템과,

^o인하공업전문대학 컴퓨터시스템과

e-mail: seihoon@inhac.ac.kr*, yewo2nn16@gmail.com^o, clear0755@naver.com*

Design of Dataset Archive for AI Education

Se-Hoon Lee*, Ye-Won Noh^o, Yeon-Su Noh*

*Dept. of Computer Systems & Engineering, INHA Technical College,

^oDept. of Computer Systems & Engineering, INHA Technical College

● 요약 ●

본 논문에서는 효율적인 AI 교육을 위한 데이터셋 아카이브와 데이터 활용을 위한 프로그래밍 플랫폼과의 연동 모듈을 제안한다. 데이터셋 아카이브는 공공데이터를 전처리하여 생성한 데이터를 모아 설계하며, 프로그래밍 플랫폼 코드비(CodeB)와 연동하여 데이터를 활용할 수 있도록 한다. 코드비(CodeB)는 파일선 블록 프로그래밍 플랫폼으로 연동을 통해 데이터를 활용한 프로그래밍이 가능하다.

키워드: 인공지능 교육(Artificial Intelligence education), 데이터셋 아카이브(Dataset Archive), 교육용 데이터셋(Educational Datasets)

I. Introduction

초중고교를 대상으로 AI 교육이 확산함에 따라 교육용 데이터셋의 필요성이 대두되고 있다[1]. 다양한 공공데이터가 개방되어 있으나 전처리 과정이 필요하여 교육에 바로 대입하기에 까다롭다.

또한 AI 교육을 위해서는 데이터와 프로그래밍 플랫폼이 연동되어야 할 필요가 있다.

이에 본 논문에서는 교육에서 활용 가능한 데이터를 생성하여 데이터셋 아카이브 서버 구축을 제안하고, 프로그래밍 플랫폼 코드비(CodeB)와의 연동 모듈을 제안한다.

통합관리를 위해 규정되었으며 방대한 교육정보의 검색 속도와 능률에 매우 도움이 된다[2].

III. The Proposed Scheme

본 논문에서 설계한 시스템의 전체 구조는 Fig 1과 같이 코드비 서버와 사용자 PC 그리고 데이터셋 아카이브로 구성된다.

II. Preliminaries

데이터셋을 제공하는 기존 플랫폼으로 CORGIS, 엔트리(Entry) 등이 있다. CORGIS는 다방면의 데이터를 다양한 형태로 제공하나 프로그래밍 플랫폼과 연동이 되어있지 않다. 엔트리는 제공하는 데이터셋을 활용한 프로그래밍이 바로 가능하나 블록 프로그래밍만 가능하다.

데이터의 검색에 용이하게 하기 위해서는 이터에 관한 구조화한 데이터인 메타데이터(Metadata)가 필요하다. 표준화된 메타데이터에는 LOM, KEM 등이 있다. LOM은 학습객체에 대한 메타데이터 인스턴스의 구조를 정의하는 개념적 데이터 스키마를 상세화하여 설계에서 높은 수준의 상호운용성을 가지도록한다. KEM은 교육정보

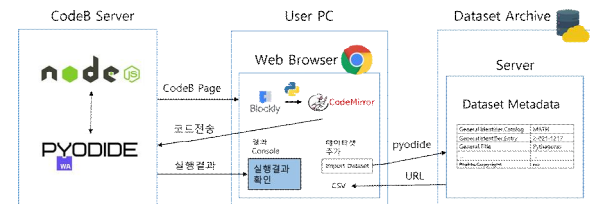


Fig. 1. System Architecture

1. 코드비(CodeB)

본 논문은 코드비 플랫폼의 아키텍처를 확장하는 구조를 갖는다. 코드비 플랫폼은 Fig 2와 같다.

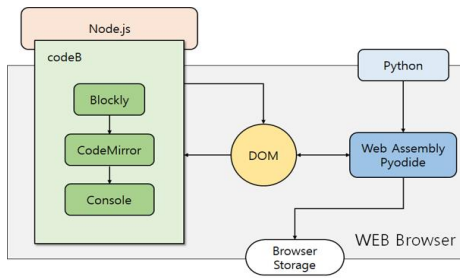


Fig. 2. CodeB Platform Architecture

node서버에 코드비 플랫폼이 동작하며 블록클리를 통해 DOM에 블록 프로그래밍 환경을 구축한다. 사용자가 블록을 생성하게 되면 실시간으로 코드미러에 파이썬 코드가 출력되며 실행 시 코드미러의 파이썬 코드가 Web Assembly Pyodide에 전달되어 처리 후 CodeB Console창에 출력값을 보여준다.

2. 데이터셋 아카이브

데이터셋은 Fig 3과 같이 공공데이터를 전처리하는 과정을 거쳐 생성한다[3].

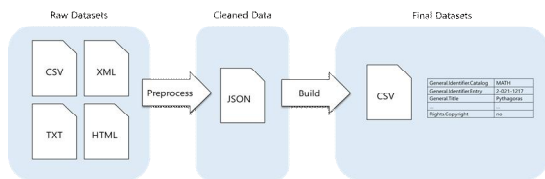


Fig. 3. Dataset Archive Design

메타데이터는 KEM3.0 표준을 활용하여 설계하며, Table 1과 같이 4개의 큰 범주로 분류된다[2].

Table 1. Metadata Design

분류	요소	설명
1	General	일반적인 정보
2	Meta-Metadata	메타데이터의 정보
3	Technical	학습객체에 관한 요소
4	Rights	지적소유권과 조건

3. 프로그래밍 플랫폼 간 연동 모듈

코드비와 데이터셋 아카이브간의 연동 모듈은 Fig 4과 같은 구조를 가진다.

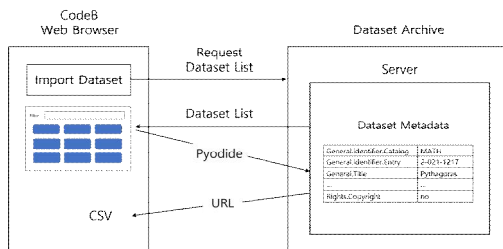


Fig. 4. Interface Module Architecture

코드비에서의 'Import Dataset'을 통해 데이터셋 리스트를 받아와 원하는 데이터셋을 선택한다. 데이터셋 아카이브는 해당 데이터셋의 URL을 코드비로 전달하며, 사용자는 Pyodide의 open_url() 함수와 Pandas의 read_csv() 함수를 통해 CSV 파일을 사용한다.

IV. Conclusions

AI 교육을 위한 데이터셋 아카이브 서버를 설계하였다. 활용 시 전처리 과정을 거치지 않아 AI 교육에서 쉽게 적용이 가능할 것으로 예상된다.

데이터셋 아카이브 서버와 프로그래밍 플랫폼을 연동함으로써 데이터를 활용하기에 편리하다. 코드비는 블록 프로그래밍 환경으로 초심자도 쉽게 데이터 활용이 가능하다.

REFERENCES

- [1] Seulki Kim, Taeyoung Kim, "Study on Educational Standard Dataset Requirements for Artificial Intelligence Education", Journal of The Korean Association Of Computer Education, Vol. 25, No. 2(A), pp. 43-46, Aug. 2021.
- [2] JinGon Son, Yongsang Jo, KwangSik Jeong, BeomSeok Go, "Metadata(KEM v3.0) Research for Higher Education Information and Copyright Management", Korea Education and Research Information Service, KR 2005-27, 2005.
- [3] Austin Cory Bart, Ryan Whitcomb, Dennis Kafura, Clifford A. Shaffer, Eli Tilevich, "Computing with CORGIS: Diverse, Real-world Datasets for Introductory Computing", In Proceedings of the 48th ACM Technical Symposium on Computer Science Education, SIGCSE '17, pp. 57-62, Mar. 2017.