

수집 시스템간의 이기종 데이터 연계 방법 연구

박민우^o, 심형섭^{*}

^{*}한국과학기술정보연구원 개방형데이터융합연구단,

^o한국과학기술정보연구원 개방형데이터융합연구단

e-mail: {pminwoo, hsshim}@kisti.re.kr^{o*}

A study on the method of linking heterogeneous data between collection systems

Min-woo Park^o, Hyeong-seop Shim^{*}

^{*}Convergence Research Center for Data Driven Solutions, KISTI,

^oConvergence Research Center for Data Driven Solutions, KISTI

● 요약 ●

사회현안을 해결하기 위한 데이터 분석을 위해서는 많은 양의 데이터 수집과 데이터 분석에 활용할 수 있도록 데이터 전처리가 필요하다. 많은 양의 데이터를 수집 및 처리를 위해 데이터 수집, 데이터 저장, 활용 시스템이 기능적으로 분리하여 시스템을 구성하고, 이에 따른 시스템간의 데이터 상호 연계가 필요하게 된다. 또한 외부 네트워크에 구성되어 있는 시스템간의 데이터 연계나, OpenAPI와 같이 다양한 데이터 서비스에서도 적용이 가능할 수 있도록 확장성과 유연성을 고려할 필요가 있다. 본 논문에서는 부산 지역현안 해결을 위한 시스템 구성에 있어, 확장성을 고려한 데이터 수집 시스템간의 효율적인 데이터 연계 방법을 제안하고자 한다.

키워드: 플랫폼(Platform), 데이터수집(Data collection), 데이터연계(Data link)

I. Introduction

ICT의 급속한 발전으로 데이터에 기반을 둔 인공지능 기술 활용 등 데이터가 주도하는(Data-Driven) 사회로 빠르게 발전하고 있다. 이러한 인공지능을 이용하기 위해 공공데이터가 활용이 확대되고 있다. '공공데이터의 제공 및 이용활성화에 관한 법률'에 의해 공공데이터는 공공기관이 만들어내는 모든 자료나 정보로 국민 누구나 활용할 수 있도록 파일과 OpenAPI를 통해 데이터를 제공하고 있다.[1][2]

많은 분야에서 사회현안을 해결하기 위해 공공데이터와 자체 데이터를 활용하여 문제 해결할 수 있는 접근이 시도되고 있다. 이에 KISTI는 ETRI와 융합연구를 통해 AI기술을 활용한 공공데이터 기반 지역현안(침수,미세먼지,지진, 헬스케어)해결을 위한 솔루션을 개발하게 되었다. 전체 시스템은 수집 및 정제, 오픈지식플랫폼, 현안해결데이터솔루션으로 구성하였다

본 논문에서는 지역현안 해결을 위한 시스템 구성에 있어, 확장성을 고려한 데이터 수집 시스템간의 효율적인 데이터 연계 방법을 제안하고자 한다.

II. Preliminaries

1. Related works

1.1 수집데이터

오픈지식플랫폼은 AI기술을 활용할 수 있도록 정제된 데이터를 제공하는 기능을 담당하는 플랫폼이다. 플랫폼의 연계대상은 공공데이터 및 지역현안데이터 약519종을 수집하여 데이터의 표준화 및 정제를 하는 시스템과 대기 관련 25개의 센서데이터를 수집하는 시스템을 대상으로 한다.

1.2 Apache Nifi

Apache Nifi는 시스템간 데이터 전달을 효율적으로 처리, 관리, 모니터링하기에 최적화된 시스템으로 대량의 데이터를 수집, 처리할 수 있는 오픈소스 시스템이다. 데이터를 처리를 위해 Dataflow의 설정할 수 있는 UI제공으로 쉽게 개발이 가능하고, 시스템간의 이동과 내용을 확인할 수 있다. 데이터 연계는 다양한 DB와 Connection이 가능하고 HDFS와 같은 분산환경 파일시스템에 데이터를 로드가 가능하다.[3]

이러한 NiFi의 기능을 활용하여 오픈지식플랫폼에서 데이터수집 기능에 적용하고자 한다.

III. The Proposed Scheme

지역현안 해결 솔루션의 시스템의 구성은 공공데이터 및 현안데이터의 수집하여 정제하는 시스템과 데이터를 활용하여 현안을 해결하는 시스템으로 구성하였고, 데이터 활용을 위해 수집, 저장·관리, 데이터 제공하는 오픈지식플랫폼으로 구성되어 있다.

오픈지식플랫폼은 데이터 수집 및 정제시스템과 IoT데이터 수집 시스템에서 데이터를 연계 수집하고, 이를 메타데이터와 대용량 데이터를 위해 하둡시스템에 저장 관리를 한다. 이렇게 수집된 데이터를 현안해결에 활용할 수 있도록 지식그래프로 구축한다. 구축된 데이터셋과 지식그래프는 개방형 데이터 솔루션에 OpenAPI 및 데이터 다운로드를 통해 제공하게 된다.



Fig. 1. System Architecture

1세대 지능형 공공데이터에서는 자동 수집 기능을 통해 공공데이터 및 지역현안 데이터의 csv, excel의 파일과 OpenAPI로 수집되는 JSON의 파일을 RDB에 저장하여 표준화 및 정제를 한다. 3세대에서는 IoT장치에서 수집된 데이터는 JSON형태로 내부 시스템에 저장한다. 이러한 데이터는 아래 그림과 같이 데이터 연계 모듈을 통해 데이터의 특성에 따라 데이터 연계 및 저장구조에 저장하는 프로세스를 수행한다.

메타데이터는 NiFi를 통해 mysql RDB에 저장하고 대용량의 데이터셋은 Hadoop의 HDFS에 변환하여 저장하여 관리한다. IoT데이터는 IoT 수집서버에서 API를 통해 HDFS와 검색색인 시스템에 저장하여 데이터를 저장 관리한다.

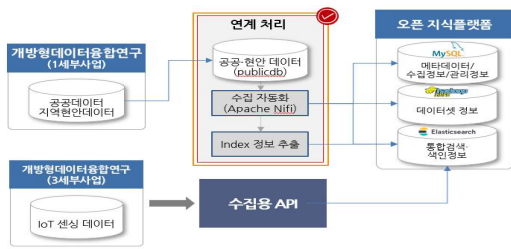


Fig. 2. Data collection method

Apache NiFi는 대량 데이터로 확대 시 분산 클러스터를 이용하여 부하분산이 가능하며, 외부 시스템의 DB연결과 OpenAPI를 통해 데이터 수집이 가능하다.

NiFi와 API를 이용하여 데이터를 연계하는 방식은 현재의 시스템에서 데이터 구조에 따른 유연하게 데이터 연계가 가능하고, 부하분산을 위한 클러스터 활용 및 직접 데이터 수집 방식을 지원이 가능하다.

IV. Conclusions

지역현안 해결을 위한 시스템은 공공데이터셋 및 IoT데이터의 수집시스템의 오픈지식플랫폼의 데이터 연계를 위해 NiFi와 API를 활용하였다. 오픈소스인 NiFi를 채택하여 데이터의 형태와 시스템에 환경의 변화에도 유연하게 적용이 가능하고, 부하분산을 위한 클러스터 활용 및 외부 시스템으로부터 직접 데이터 수집을 할 수 있는 확장성 갖추고 있고 있다.

ACKNOWLEDGEMENT

이 연구는 국가과학기술연구회의 지원을 받은 DDS융합연구단의 “AI기술을 활용한 공공데이터 기반 지역현안 솔루션 개발 및 실용화 -안전·안심사회 실현을 위한 실증연구 중심으로-”(CRC-19-01-KISTI)사업으로 수행되었습니다.

REFERENCES

- [1] Ministry of the Interior and Safety. ACT ON PROMOTION OF THE PROVISION AND USE OF PUBLIC DATA. , <https://www.law.go.kr/LSW/eng/engLsSc.do?menuId=2§ion=lawNm&query=공공데이터&x=0&y=0#liBgcolor2>
- [2] Public Data Portal, <https://www.data.go.kr/en/ugs/selectPublicDataUseGuideView.do>
- [3] Apache Nifi , <http://nifi.apache.org>