참조데이터 기반 데이터 연관성 도출 및 연관도 가시화 모델

KiSTi 한국과학기술정보연구원 Korea Institute of Science and Technology Information www.kisti.re.kr

특 허 명 데이터 매핑을 위한 참조데이터 모델
데이터셋 연관도 가시화 모델

Keyword 빅데이터, 참조데이터, 데이터셋

발 명 자 황윤영 / 신수미

기 술 성

기술 개요

- 본 발명은 사용자에게 재난분야 데이터셋을 제공하기 위한 데이터 연계성을 구축하고, 연계 기준의 되는 참조데이터 모델을 제공하는 기술임
- 데이터셋간의 연관도를 사용자에게 효과적으로 보여주기 위한 표현방법을 제공하는 기술임

기존 기술 문제점

- 데이터 활용을 위한 노력의 80%는 데이터 정제 및 전처리 작업에 소모되고 있으며, 데이터 사용자마다 소모성 작업을 반복하고 있는 문제의 해결이 필요함
- 공공데이터 포털에 제공되는 데이터를 타 데이터와 연계하고 의미관계를 분석할 수 있도록 지식화가 필요함
- 데이터셋은 메타데이터 수준의 비교로는 다양한 측면의 연관도 비교가 어렵고 실제 값 전체를 비교할 경우 유의미한 연관도를 찾기 어렵기 때문에 데이터셋간의 연관도를 측정하고자 하는 시도가 부족함

기술의 특징 및 우수성

▶ 기술의 특징

- 재난의 예방, 대응, 해결, 복구등을 위해 필요한 데이터셋을 발견하기 위해 비즈니스 프로세스, 관련지식, 재난관리 분류체계를 적용하여 데이터셋 연계성을 도출함
- 이질적이며 다양한 컬럼명을 가지는 데이터셋 간의 연관성을 높이기 위한 데이터 매핑에 기준이 되는 참조 데이터를 제공함
- 데이터셋의 동시 활용성을 높이고 새로운 데이터셋을 발견하기 위한 여러 기능에 적용할 수 있도록 데이터셋간의 연관도를 계산할 수 있는 방법과, 이 데이터셋간의 연관도 분석결과를 직관적으로 이해할 수 있는 가시화 방안을 함께 마련

▶ 기술의 우수성

- 사회 현안 해결형 데이터 활용 · 생태계 및 데이터 산업 육성 기반 구현을 위한 데이터 수집 · 처리 · 연계 · 분석 · 예측 등 관련 기술을 확보함
- 데이터셋간 연관도를 공간, 시간 항목, 주제 측면에서 직관적으로 확인할 수 있으며, 특히 공간, 시간 이라는 새로운 개념의 연관관계를 통해 획득할 수 있음

참조데이터 기반 데이터 연관성 도출 및 연관도 가시화 모델

상세설명

- 본 기술의 데이터 매핑을 위한 참조데이터 모델은 공통 데이터셋 추출부, 구축부 및 재난분야 분류정보 재난관련 프로세스 정보, LOD(Linked Open Data)데이터 구축부, 데이터셋 연관도 분석부, 연관 데이터셋 도출부를 포함
- 공통 데이터셋 추출부는 데이터셋의 연관도 분석을 위해, 항목별 데이터셋을 설명하는 메타데이터, 공간 관련 모든 항목별 및 값, 시간 관련 모든 항목명 및 값을 추출함
- 본 발명은 데이터셋간의 연관도를 사용자에게 효과적으로 보여주기 위한 표현방법은 제공
- 데이터셋의 주요 상관관계는 공간, 시간, 항목, 주제로 구분하였으며, 정의한 표현 방식에 따라 데이터셋의 연관도를 측정함
- 가시화한 분석모델은 데이터셋 객체 컬러 요소 및 데이터셋간의 연결선은 높이, 굵기, 컬러의 요소를 포함

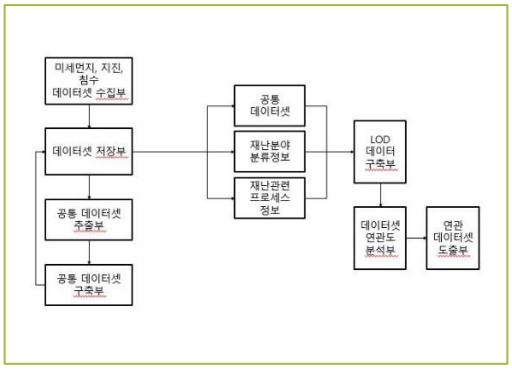

그림1 데이터 매핑 장치의 구성도

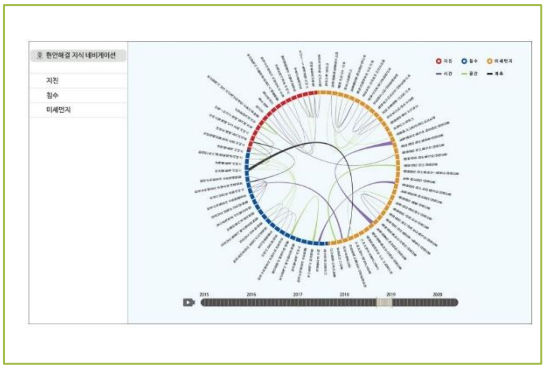

그림2 정해진 표현방식에 따라 데이터셋의 연관도를 가시화한 예

기술완성도 (TRL)

기술완성도 : TRL2 (연구개발 진행중)

TRL1	TRL2	TRL3	TRL4	TRL5	TRL6	TRL7	TRL8	TRL9
기술원리 발표	기술컨셉 설정	기술컨셉 증명	Lab Scale 시제품개발	구현환경 적용실험	Full Scale 시제품개발	유사 상용품 개발	상용품 완성	상용품 실시

활 용 분 야

활용분야 및 적용제품

활용분야

- ◆ 재난 해결 및 대응에 필요한 데이터 검색
- ◆ 업무처리에 필요한 데이터 검색
- ◆ 정보관계 분석 기업

적용제품

- ◆ 업무메뉴얼에 따른 필요데이처 제공 서비스
- ◆ 데이터 연관도 기반 가시화 서비스
- ◆ 데이터셋 공개 웹/모바일 서비스

참조데이터 기반 데이터 연관성 도출 및 연관도 가시화 모델

산업동향(기술 동향 및 트랜드 등)

- 최근 국가 수준에서 축적된 연구개발 정보들에 대해 다양한 경로로의 지식 연계성 확보를 통해 그 활용을 보다 극대화하고자 하는 노력이 확대되고 있음
- 국가적으로 보유 또는 생산한 지식을 체계적으로 축적하고, 이를 다양한 지식서비스나 채널을 통해 제공함으로써, 지식의 재생산과 순환이 역동적으로 이루어지도록 하는 체계에 대한 중요성이 강조됨
- 공공데이터를 쉽게 탐색하고 활용할 수 있도록 체계화하거나 목록화하고, 이에 대한 사용자 친화적 기능을 제공하여 공공데이터의 확산을 지원하는 데이터 접근성이 강조되고 있음

시장전망(목표시장 규모 및 전망)

- 국내 빅데이터 시장은 2023년까지 연평균 성장률 11.2%를 기록하며 2조 5,692억원의 규모에 달할 것으로 전망. 전망 기간 동안 전체 시장에서 IT 및 비즈니스 서비스가 연평균 16.3% 성장하며 가장 큰 비중을 차지할 것으로 예상
- 빅데이터 및 분석 시장은 이를 성장 동인으로 규모와 가능성이 꾸준히 발전하도 있는 가운데 빅데이터 기반의 고급 분석 및 인공지능 시스템 구축을 위한 데이터의 필요성 증가 요인에 의해 관련 시장은 향후 지속적 성장 예측
- 또한, 기업은 빅데이터 및 분석 솔루션을 통해 다양한 형태의 데이터 인사이트를 실시간으로 발굴하고, 복잡한 분석 및 비즈니스 의사 결정을 자동으로 해결함으로써 지속적인 수익 창출과 경쟁 우위를 확보할 수 있을 것으로 기대

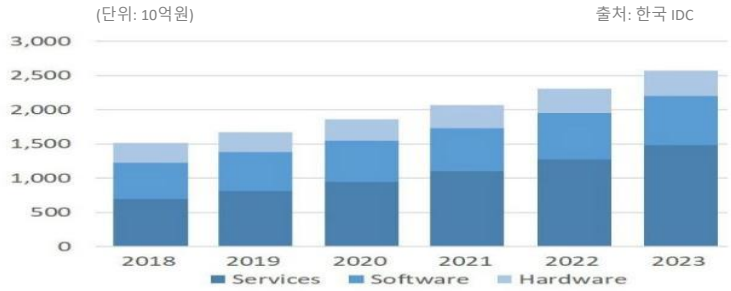

그림3 국내 빅데이터 분석 시장 전망

지재권현황

권리현황	특허출원번호	발명의 명칭
출원 출원	10-2021-0096184 10-2021-0094083	데이터매핑을위한참조데이터모델 데이터셋연관도가시화모델

문 의 처

기술이전

KiSTi www.kisti.re.kr 한국과학기술정보연구원 Korea Institute of Science and Technology Information

담 당 자 ooo
연 락 처 042-869-0915
이 메 일 kwsim@kisti.re.kr

기술문의

KiSTi www.kisti.re.kr 한국과학기술정보연구원 Korea Institute of Science and Technology Information

담 당 자 이재민 책임연구원
연 락 처 042-
이 메 일 @kisti.re.kr