
ISO/IEC의 온톨로지와 메타데이터 표준화 동향

2013. 06. 26.

한국과학기술정보연구원

김장원

목차

- ❖ 메타데이터와 온톨로지
 - ❖ 표준 제정 및 표준화 기구
 - ❖ 사례 분석
-

소개

❖ 메타데이터(Metadata) 정의

- ✓ 데이터의 데이터 (Wikipedia)

met·a·da·ta

/ˈmetəˌdɑːtə/

Noun

A set of data that describes and gives information about other data.

- ✓ 객체 혹은 사물에 관한 기술 (ISO/IEC)

if P is data and if $P \rightarrow Q$ represents the descriptive relationship such that P describes Q ,
then P is metadata about Q .

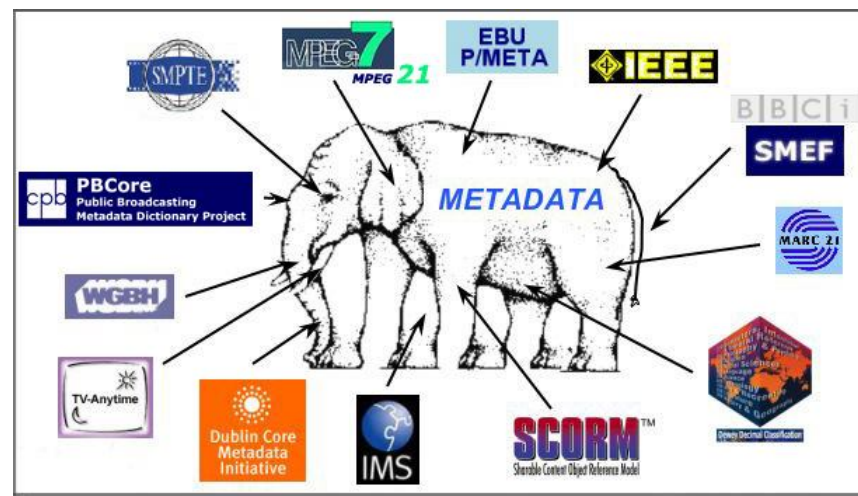
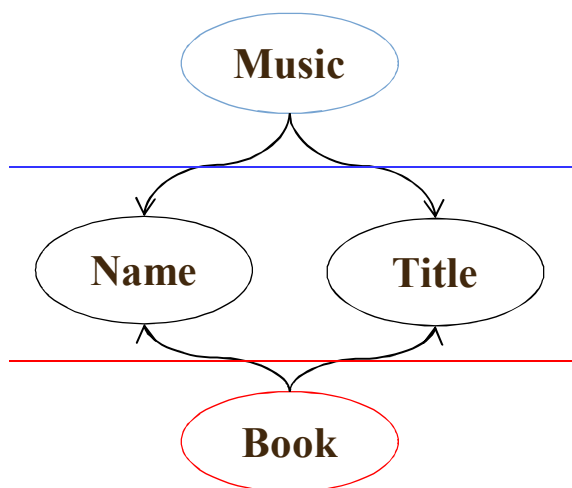
P is merely data, because metadata is always relative to the object of description.

P only becomes metadata once its descriptive relationship to Q is established.

소개

❖ 메타데이터 이슈

- ✓ 데이터의 메타데이터를 추출 시 문맥과 관계정보들을 고려하지 않고서는 정확한 메타데이터를 결정할 수 없다
 - 메타데이터는 일종의 데이터이다
 - 데이터들처럼 상호 교환이 가능하지만, 여전히 메타데이터이다
 - 메타데이터의 상호 교환 시 연관된 상황정보와 관계 정보들을 포함해야 한다



Public Broadcasting Metadata Dictionary Project

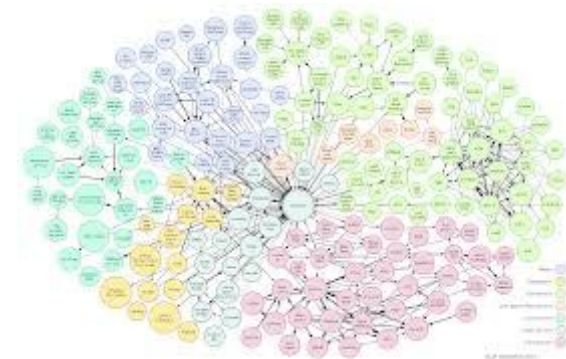
소개

❖ 온톨로지

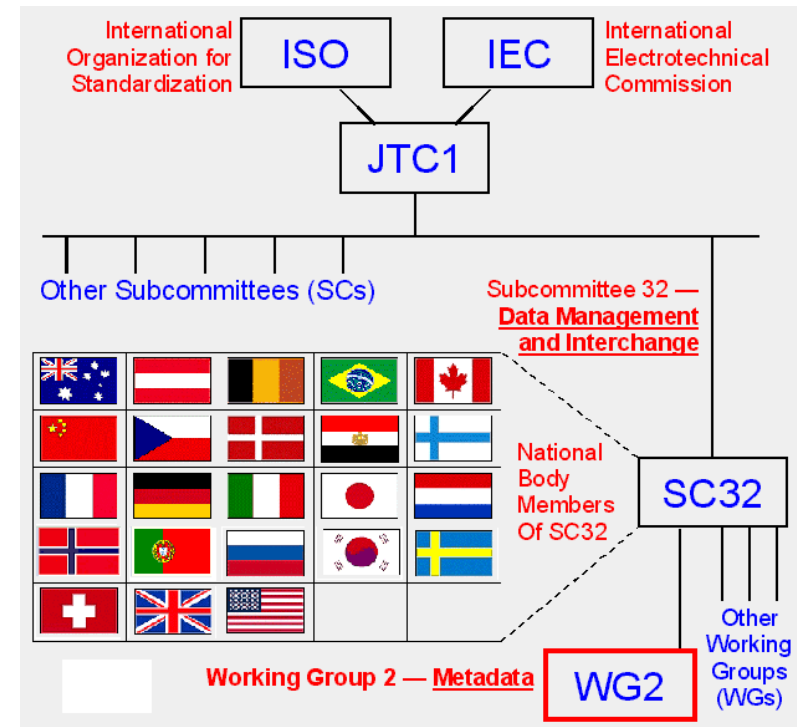
- ✓ 개념을 식별하고 표현하기 위한 철학적 접근 방법 (Wikipedia)

Ontology is the philosophical study of the nature of being, becoming, existence, or reality, as well as the basic categories of being and their relations.

- ✓ 지식과 정보를 표현하기 위한 체계
 - 명시적 개념화
 - 어휘의 공유를 통한 의미 표현
 - 자원(Resource)을 서술하기 위한 프레임워크 (RDF)
 - 자원들간의 의미와 관계를 표현



표준 기구



ISO/IEC JTC 1

❖ 소개

- ✓ 빠르게 변화하는 기술 추이에 맞춘 표준을 개발하고 ISO, IEC 두 기관에서의 중복을 피하고 공통의 세계 표준을 만들하고자 1987년 창설
 - MPEG, IC cards, 데이터베이스 질의 언어, 메타데이터 등 2만여 개의 표준 개발
 - 2,500의 국제 표준을 만들었으며, 37개의 나라에서 참여하며, 총 19개의 분과위원회로 구성 됨

❖ 비전

- ✓ 정보와 통신 기술에 대한 표준을 만드는 것으로, 다양하고 복잡한 기술을 전 세계적으로 통합하기 위한 표준화된 환경을 조성하고자 함
- ✓ 최근, 급속도로 변화하는 기술을 따라가기 위하여 De-facto 표준을 만드는 곳들의 표준안들을 적용 및 승인함
 - Publicly available specification (PAS submitters)
 - W3C, OMG

ISO/IEC JTC 1/SC 32

❖ 소개

- ✓ 데이터와 데이터 교환을 위한 위원회
- ✓ 총 4개의 작업반(Working Group)으로 구성
 - 작업반 1: 전자상거래(e-Business)
 - 전자상거래와 관련한 표준을 개정하는 곳으로 ebXML, 전자상거래 절차 등의 표준을 개발
 - 작업반 2: 메타데이터(metadata)
 - 메타데이터의 정의, 메타데이터를 등록하고 관리하기 위한 방법, 메타데이터 레지스트리, 메타모델의 상호 운용을 위한 메타모델 프레임워크, 온톨로지 생성 프레임워크 개발
 - 작업반 3: 데이터 질의 언어(Structured Query Language)
 - 데이터 질의 언어와 질의 언어를 기반으로 온톨로지에 질의 하기 위한 표준 등을 개발
 - 작업반 4: 데이터 질의 언어 멀티미디어와 응용(SQL/MM)
 - 데이터 질의 언어에서 정의하고 있는 사용자 정의 타입(User Defined Type)을 기반으로 마이닝, 지리정보, 메타데이터 질의 인터페이스 등을 개발

표준화 동향

- ❖ 다양한 기술의 등장으로 새로운 표준 아이TEM 도출을 위한 SWG-P (Special Working Group Planning) 회의 개최
 - ❖ 빅데이터, 클라우드 컴퓨팅, 소셜 네트워크 분석, 차세대 분석 기법, 인터넷 사물 통신 등의 분야에서 표준 아이TEM을 도출하려 하고 있음
 - ❖ 소셜 네트워크, 빅데이터 환경에 이르러 데이터 생성에 대한 페러다임의 변화 발생
 - ❖ 빅데이터 큐레이션이 중요하게 되며 이를 위한 기계적인 방법에 대한 표준이 중요 이슈로 등장
-

표준화 동향



표준화 이슈

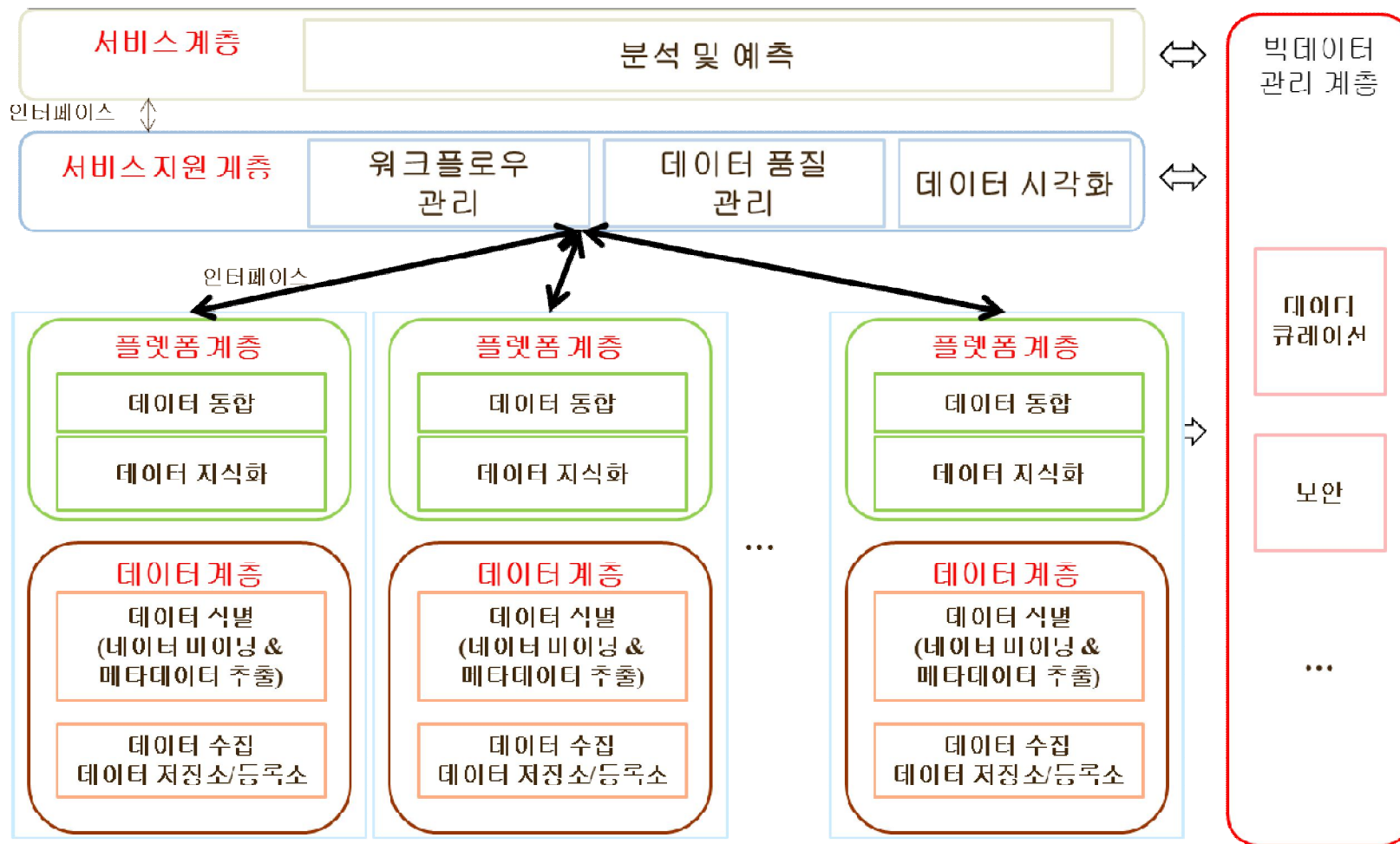
❖ 빅데이터와 관련 한 개념 및 용어 정의

✓ 데이터 큐레이션

- 정의
 - Active management of data over its lifecycle
 - Ensuring that data is trustworthy, discoverable, accessible, reusable, and fit for use
 - ◆ E.g. data preparation for analytics
- 큐레이션의 형태
 - ◆ 수동 큐레이션 (Manual Curation)
 - ◆ 반자동 큐레이션 (Semi Curation)
 - ◆ 데이터 정제, 레코드 중복 방지, 군집화
 - ◆ 가벼운 큐레이션(Sheer Curation)
 - ◆ 좋은 데이터와 디지털 자산들을 통해 창의적이며 우선적으로 사용될 수 있는 좋은 예를 보여주기 위한 것으로 데이터의 공유, 공개 행위를 통해 목적을 달성함

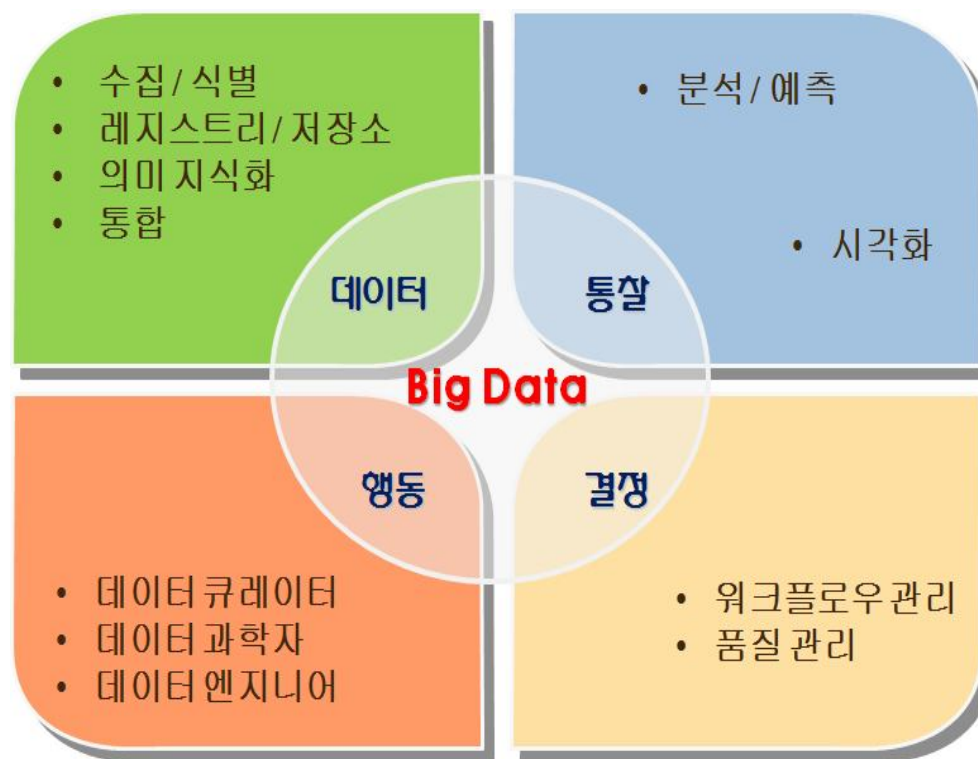
표준화 이슈

❖ 빅데이터 참조 모델 정의



표준화 이슈

❖ 빅데이터 환경에서의 데이터 생명 주기



표준화 이슈

❖ 메타데이터 이슈

- ✓ ISO/IEC JTC 1/SC 32 Plenary Meeting (2013. 5, 경주)
- ✓ 빅데이터, 클라우드 컴퓨팅 환경에서 상황정보와 관계를 반영할 수 있는 메타데이터를 정의하고 이를 추출하기 위한 기술에 대한 표준안 제안 (한국, 캐나다, 폴란드)

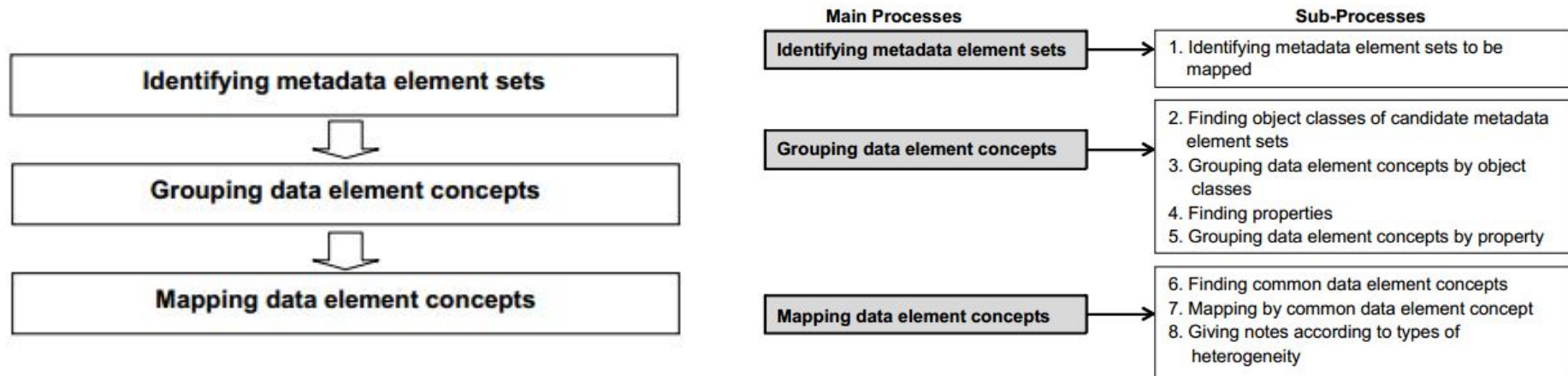
❖ 온톨로지 이슈

- ✓ 빅데이터 큐레이션을 통해 새로운 지식을 추출함에 있어 체계적이며 다양한 관점의 데이터를 추출하기 위하여 데이터 지식화가 필요
- ✓ 텍스트로부터 의미 있는 개체를 추출 및 식별을 하고 식별된 개체들 간의 의미적 연관성을 도출하여 체계적인 형태로 관리 하기 위하여 온톨로지를 사용하기 위한 방법 및 절차의 표준 도출

표준화 사례

❖ ISO/IEC 20943-5: Metadata Mapping Procedure

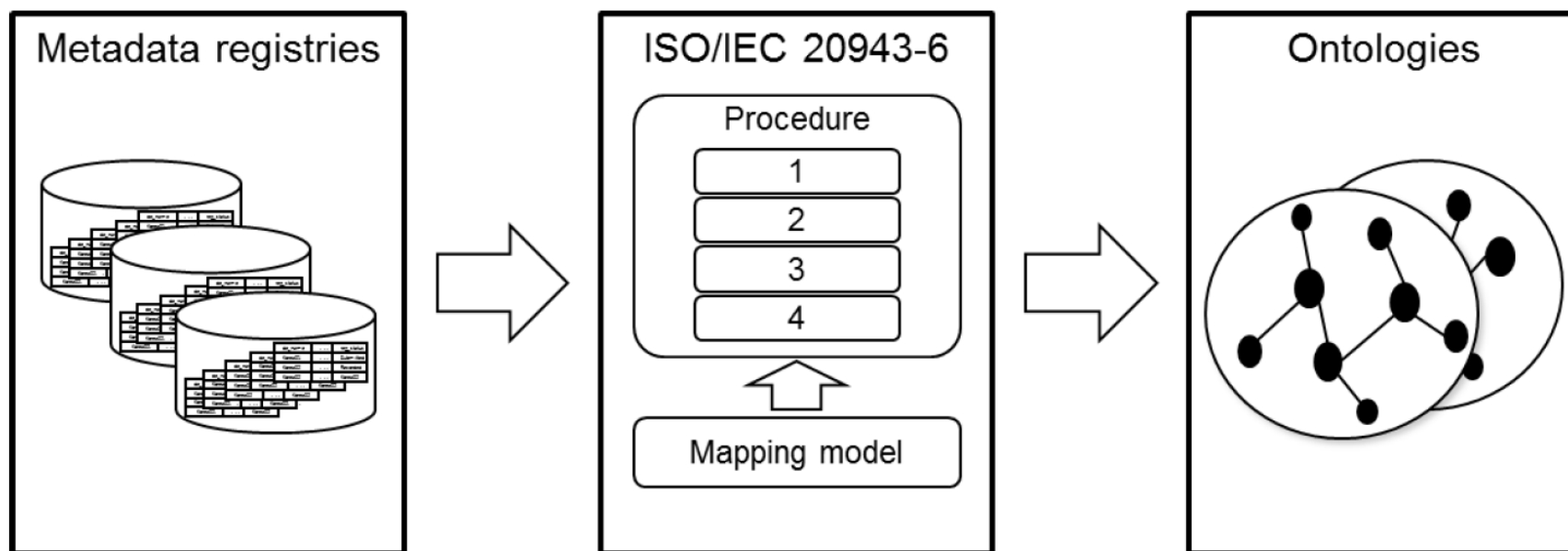
- ✓ 표준 제안 기관 (한국 – KISTI, KODB)
- ✓ 내용
 - 메타데이터의 의미적 일치를 위한 절차의 표준
 - 동일한 의미를 가짐에도 불구하고 다른 이름을 가지는 데이터 요소들에 대한 매핑을 하기 위한 표준



표준화 사례

❖ ISO/IEC 20943-6: Framework for Generating Ontology

- ✓ 표준 제안 기관 (한국 – Korea Univ.)
- ✓ 내용
 - 온톨로지를 생성하기 위한 프레임워크

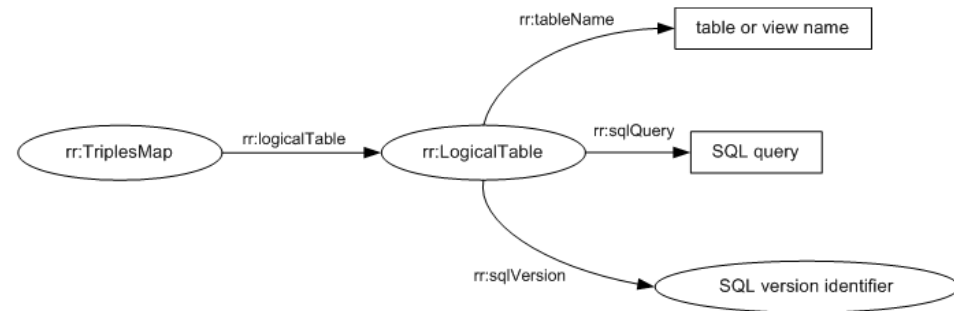
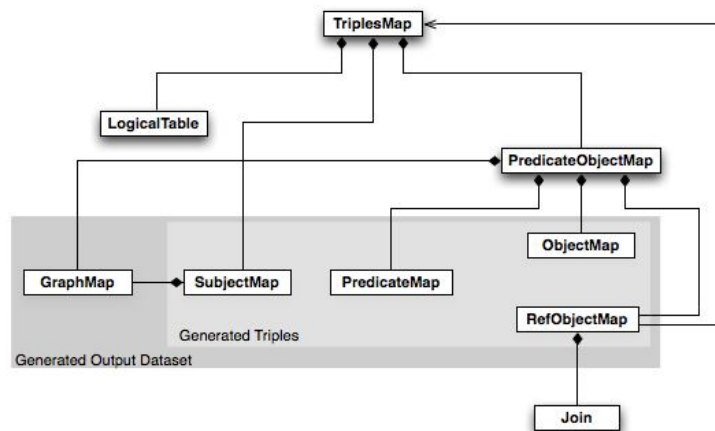


표준화 사례

❖ W3C - R2RML (RDB to RDF Mapping Language)

✓내용

- 관계형 데이터베이스로 부터 RDF 데이터 집합으로 맵핑방법을 표현하기 위한 언어
- 관계형 데이터베이스로 부터 RDF로의 매핑은 개념적 수준에서의 매핑으로 관계형 데이터베이스의 스키마에 표현되는 어휘들을 RDF의 어휘로 매핑



감사합니다.

