

KoDDAS Questions

1. KoDDAS 의 목적

- 지식 전파, 데이터 분석, 그리고 데이터 재사용을 위해 한인 디아스포라를 조직, 보존, 및 활용하는데 그 목적이 있음.

1. What is the purpose of KoDDAS?

- To organize, preserve and utilize Korean diaspora data for knowledge dissemination, data analysis, and data reuse.

2. 다른 DB 체계와 달리 KoDDAS 가 갖는 차별성 (즉, KoDDAS 만이 할 수 있는 중대한 공헌)

- 원시 자료에 주석을 추가하고 이를 조직하기 위해 전통적 관계형 데이터베이스를 구성할 뿐만 아니라, 패킷 온톨로지와 개방형 연계 데이터(LOD) 리포지토리 를 구축하기 위해 데이터 마이닝, 지식 조직화, 증강 지능, 및 데이터 모델링 방법론을 활용함.

- 전통적으로 인문사회학 분야의 복합 데이터는 연구자가 기초적인 수준에서 수동으로 처리하는 양상을 보여왔으나, KoDDAS 는 동일한 데이터를 반자동식 큐레이션 작업하는 시스템을 확립하는 것을 목표로 함.

2. How is KoDDAS different from other DB systems? (or What is the significant contributions of KoDDAS?)

- It utilizes data mining, knowledge organization, augmented intelligence, and data modeling methods to construct not only the conventional relational database but also faceted ontology and linked open data repository to annotate and organize raw data

- KoDDAS aims to establish a system for semi-automatically curating complex data typical in 인문사회학 that has traditionally been processed mostly manually and on the rudimentary level.

3. KoDDAS 산출물/결과

- 산출물: 관계형 데이터베이스, 패시 온톨로지, 개방형 연계 데이터(LOD) 리포지토리
- 서비스: 내용검색, 범주검색/탐색, LOD 다운로드, Q&A/통계/분석을 위한 데이터베이스 비정형 질의

3. What are the output/outcome of KoDDAS?

- Output: Relational database, Faceted ontology, Linked Open Data repository
- Service: Content search, Category search/browse, LOD download, Ad-hoc query of DB for Q&A, stats, and analysis

4. 증강 지능의 정의 및 원리

- 증강 지능(Augmented Intelligence)이란, 인간이 작업을 더 효과적/효율적으로 수행할 수 있도록 돕기 위해 인간 지능을 향상시키도록 설계된 인공 지능.

- 예. 이미지 => 인공지능 => 주석 (a:99+% 정확; b: 90-98% 정확; c: 기타)

=> 인간이 검토 (a: 한 사람이 확인, b: 두 사람이 검토/ 판단)

=> AI 에 피드백=> 주석 수정=> 인간이 검토(범주 a 및 b 내에 신규 항목)

4. What is the “augmented intelligence” and how does it work?

- Artificial intelligence designed to enhance human intelligence for the purpose of assisting human to perform tasks more effectively and efficiently.

- e.g., image => AI => annotations (a:99+% correct; b: 90-98% correct; c: rest)

=> human review (a: validation by 1 person, b: review/adjudication by 2 people)

=> feedback to AI => revised annotations => human review (new items in a & b categories)

5. 데이터 큐레이션

(일반적 정의) 데이터를 효율적/효과적 재사용할 수 있도록 데이터 수명주기 동안 실시하는 데이터 관리 작업.

(KoDDAS 정의) 메타데이터, 온톨로지, 및 데이터 모델을 생성할 목적으로 데이터 디지털화/문자화/주석처리, 데이터 마이닝, 지식 조직화, 및 데이터 모델링을 수반하는 데이터 수동/자동/반자동 처리 행위.

5. What is data curation?

(General) Management of data throughout its lifecycle to ensure efficient/effective data reuse

(KoDDAS) Manual/Automatic/Semi-automatic processing of data that involves data digitization/transcription /annotation, data mining, knowledge organization, and data modeling to generate metadata, ontology, and data model.

6. 패싯 분류의 정의 (기존 분류체계와의 차이점)

- 패싯(facet) 및 아이솔레이트(isolate)로 불리는 시멘틱 분류셋으로 구성된 유연한 분류체계

- facet: 어떤 클래스의 특성/특징(예. 시간, 장소)에 해당하며 데이터베이스 테이블의 속성과 유사함.

- isolate: 있음직한 패싯의 값 (예. 시간: 기원전, 기원후)

- 이점: 클래스 계층 구조 및 다중 탐색경로를 유연하게 배치함.

6. What is faceted classification (and how is it different from traditional classification)?

- Flexible classification system composed of semantic category sets called facets & isolates)

- Facet: property/characteristics of a class (e.g., time, place) similar to attributes in a database table

- isolates: possible values of a facet (e.g., time: B.C, A.D)

- advantages: flexible ordering of class hierarch and multiple path of navigation

e.g., color (red) -> type (convertible) -> motor (electric) -> brand (Kia)

7. 개방형 연계 데이터 (LOD)의 정의와 필요성/응용사례

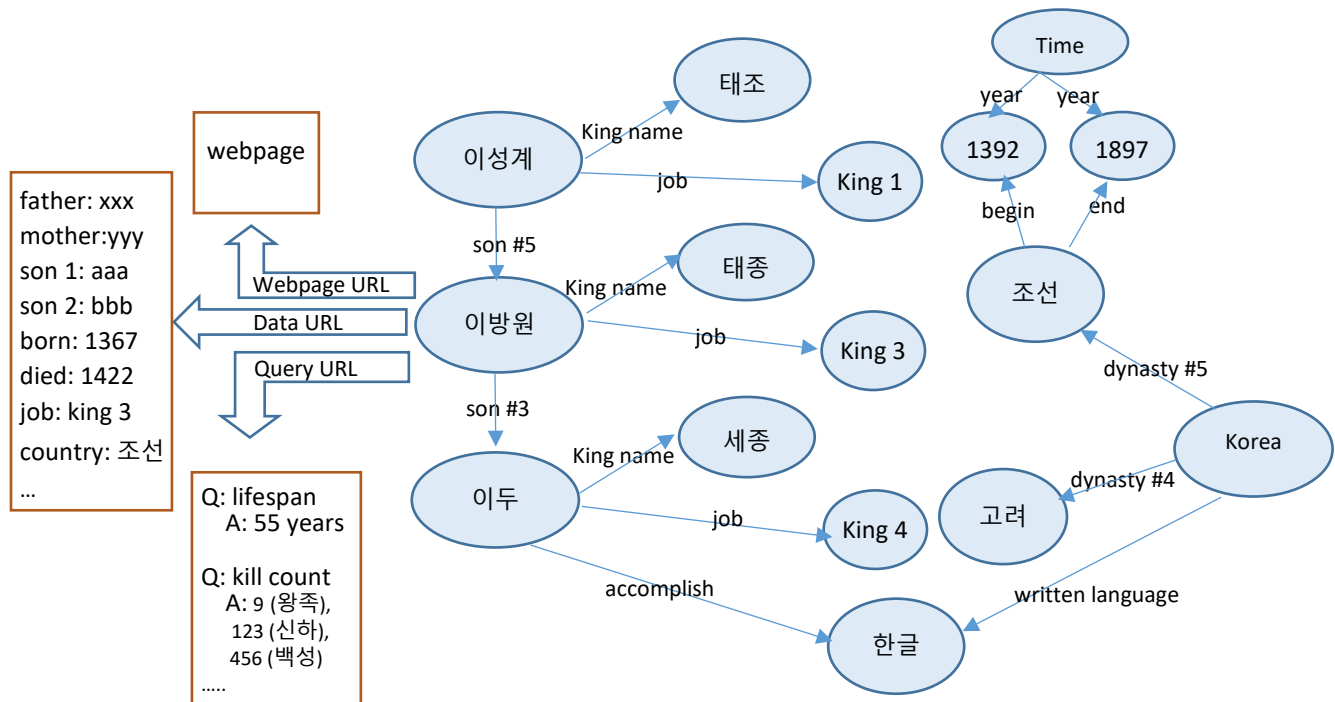
- 데이터베이스와 같은 구조화된 정보를 말하며, 개방형 웹에서 상호 연결되어 있는 RDF 양식을 주로 따름

(예)

7. What is LOD (Linked Open Data) and why is it needed (what are its applications)?

- LOD: structured information (e.g., datasets) typically in RDF format that are linked to one another in the open Web

(Example)



8. RDF 정의

- RDF(Resource Data Framework)는 웹상의 데이터(혹은 자원)의 상호교환을 목적으로 표현되는 정보를 기술하는 프레임워크/표준/모델을 말하며 3 가지 데이터(subject, predicate, object) 에 기반함

- subject = 자원(예. 엔티티)
- predicate = 자원의 특성/속성, subject 및 object 간 관계(특질)
- object = 값(예. 수치)

예문: 태종 이방원은 조선의 3 대 왕으로 출생일은 1367 년 6 월 13 일이며 사망일은 1422 년 5 월 30 일이다. 그는 조선을 건국한 태조 이성계의 5 번째 아들이었으며, 한글 창제로 유명한 조선의 4 대왕 세종대왕의 부친이기도 하다.

RDF 예제:

(이방원, job, 3th King of 조선), (이방원, king name, 태종) => 추론(3rd King of 조선, king name, 태종)
(이방원, born, 6/13/1367), (이방원, died, 5/30/1422) , (이방원, 3rd son, 이두)
(이성계, job, 1st King of 조선), (이성계, king name, 태조), (이성계, 5th son, 이방원) => 추론(이방원, father, 이성계)
(이두, job, 4th King of 조선), (이두, king name, 세종), (이두, accomplishment, 한글)
(Korea, written language, 한글),

8. What is RDF?

- Resource Data Framework: triple-based (subject, predicate, object) framework/standard/model to describe things for data interchange on the Web.
- subject = resource (i.e., entity)
- predicate = traits/aspects of the resource; relationship between subject and object (i.e., attribute)
- object = value (i.e. value)

Text example:

이방원, 3rd king of 조선 known as 태종, lived from 6/13/1367 to 5/30/1422. He was the 5th son of 태조 이성계, who was the first king of 조선, and the father of 세종, who was the 4th king of 조선 most famous for creating 한글.

RDF example:

(이방원, job, 3th King of 조선), (이방원, king name, 태종) => derive (3rd King of 조선, king name, 태종)
(이방원, born, 6/13/1367), (이방원, died, 5/30/1422) , (이방원, 3rd son, 이두)
(이성계, job, 1st King of 조선), (이성계, king name, 태조), (이성계, 5th son, 이방원) => derive (이방원, father, 이성계)

(이두, job, 4th King of 조선), (이두, king name, 세종), (이두, accomplishment, 한글)
(Korea, written language, 한글),

9. 메타데이터의 정의

- ‘데이터에 대한 데이터’를 말하며, 데이터 그 자체가 포함하고 있지 않은 정보를 제공함.
- 예. 책의 제목, 저자, 주제 등.

9. What is metadata?

- Data about data that provide information not contained that data
- e.g., title, author, subject (of a book)

10. 데이터마이닝의 정의

- 패턴 식별이나 정보 조직, 데이터 분석 등에 활용할 목적으로 많은 데이터(자료)중에서 정보를 추출하는 일련의 과정 혹은 방법

10. What is data mining?

- Process/method of extracting information from typically a large amount of data for pattern identification, information organization, and data analysis

11. 온톨로지 정의

- 개념/범주로 구성된 하나의 집합과, 특정 범위(domain)를 표현하는 집합 내 구성원 관계를 일컬음.
- 온톨로지는 본질적으로 동적이나, 텍소노미(Taxonomy)는 온톨로지의 정적 부분집합으로 간주될 수 있음.

11. What is ontology?

- a set of concept/categories and relations among them that describe a domain
- Taxonomy can be thought of as a static subset of ontology, which is dynamic