

인간-기계 지식소통을 위한 자연어 QA 워크샵 -엑소브레인 인공지능

# 이중나선 기반 지식베이스 구축과 질의응답 시스템 구현

2015.08.21

(주)솔트룩스 / 함영경 팀장



한국정보과학회

KOREAN INSTITUTE OF INFORMATION SCIENTISTS AND ENGINEERS

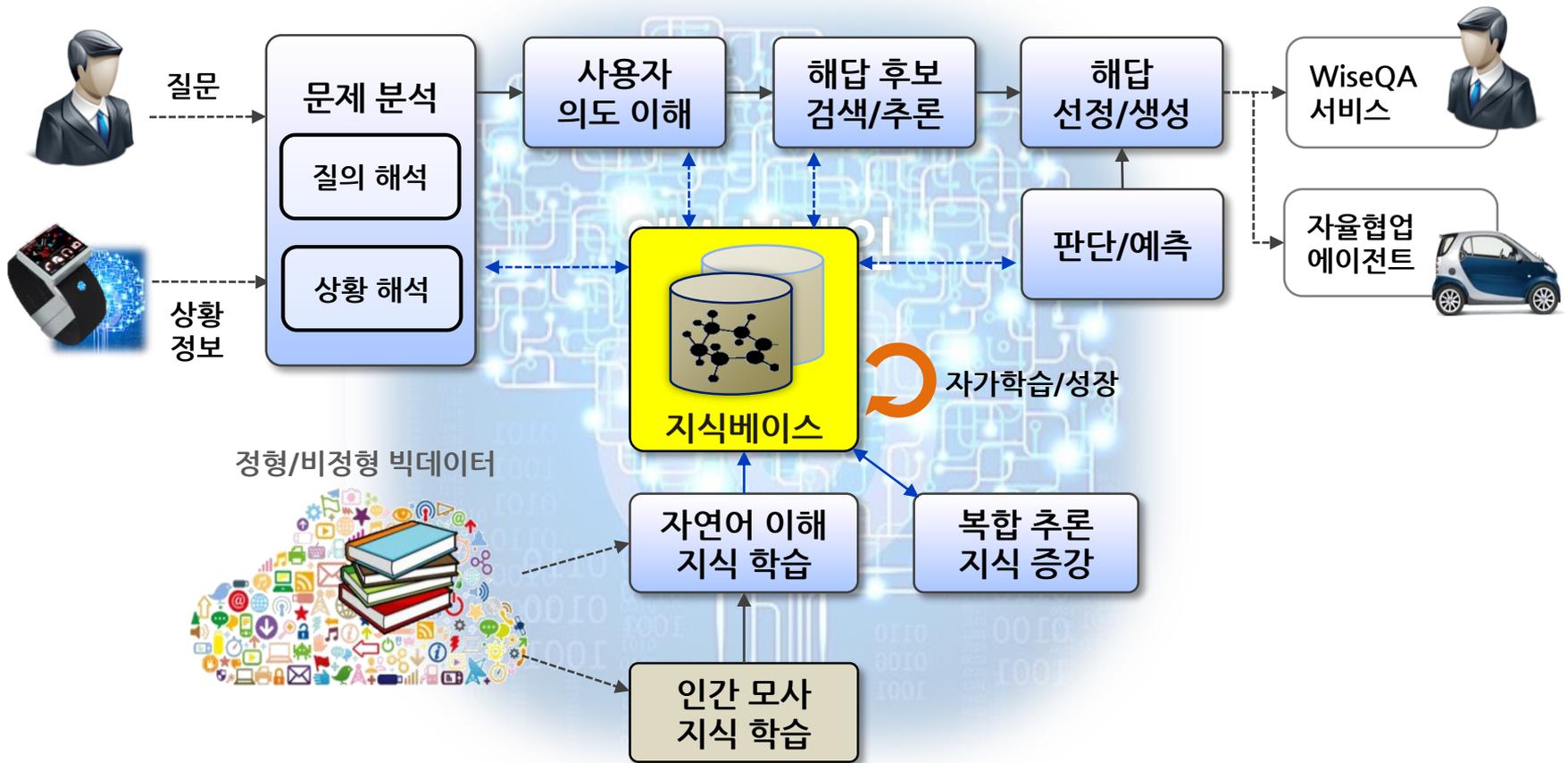


# 목 차

- Overview
- XB Ontology Schema
- XB 지식표현
- XB Ontology Instance
- XB QA

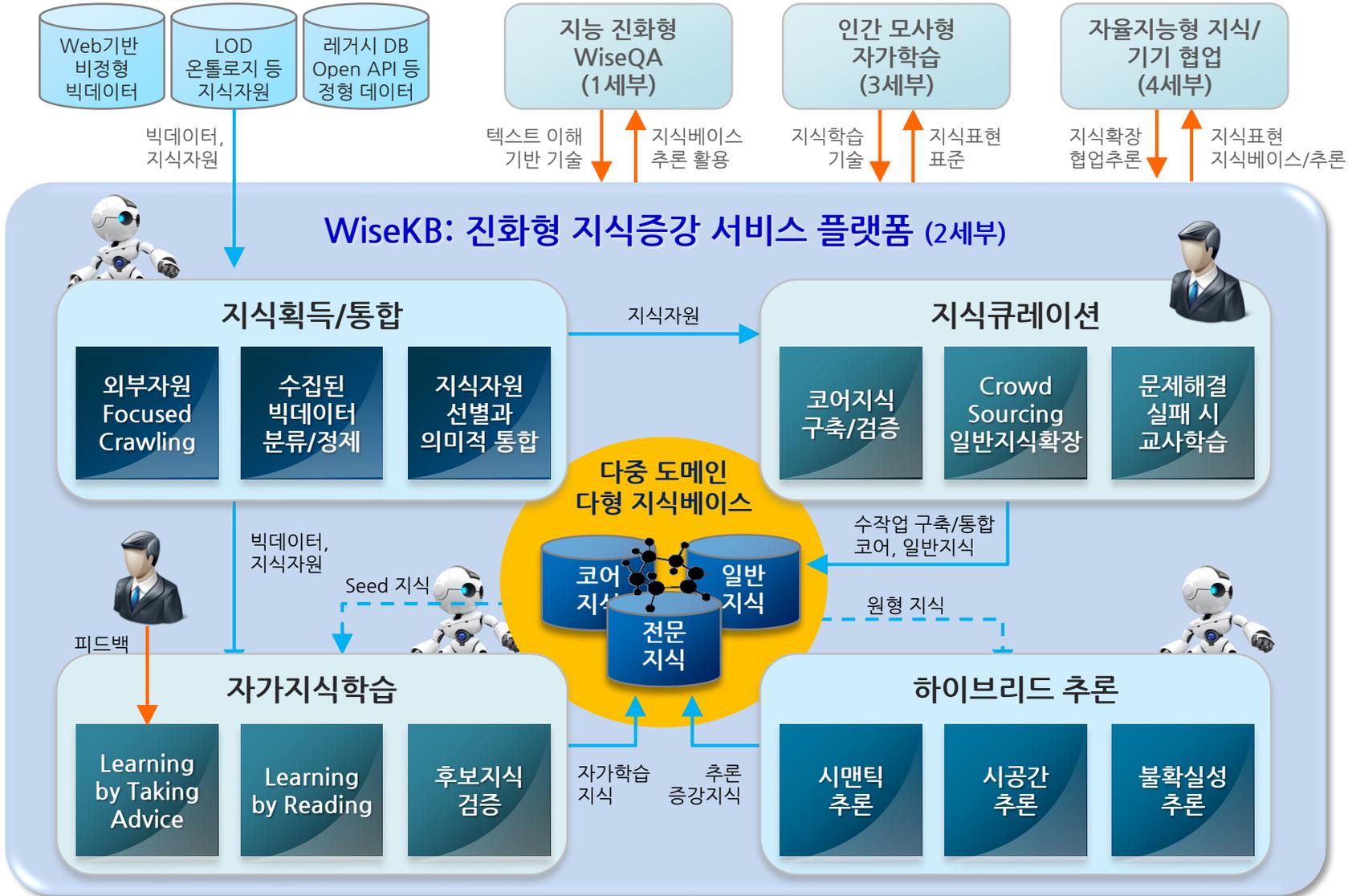
# Overview

## 엑소브레인은 인간처럼 학습, 판단 가능한 미래 AI 기술



# WiseKB 플랫폼 작동 개념

이중나선 기반 지식베이스 구축과  
질의응답 시스템 구현





# 지식학습과 하이브리드(복합) 추론

이중나선 기반 지식베이스 구축과  
질의응답 시스템 구현



이 사람은 중등 교사인 어머니와  
재혼한 군인 아버지 사이에서  
차녀로 태어났다.  
경상도가 고향으로 광복절에  
어머니를 잃었으며 올해 환갑이  
된 이 사람은 누구인가?

1세부

Whols(?x) :-  
hasMother(?x, ?mother),  
job(?mother, 중등교사),  
becomes(?x, 환갑),  
home(?x, 경상도),  
hasFPosition(?x, 차녀),  
hasFather(?x, ?father),  
job(?father, 군인),  
remarriage(?father, ?mother)  
...

고아라,  
전재호,  
한준위...

단순질의  
(그래프매칭)

복합추론

시맨틱추론

시간추론

공간추론

불확실  
확률추론

박근혜(90%)  
고아라(10%)  
전재호(5%)  
...



박정희



박근혜  
朴槿惠

한국 전 대통령 박정희 장녀 박근혜 대통령선  
거 참가 선포

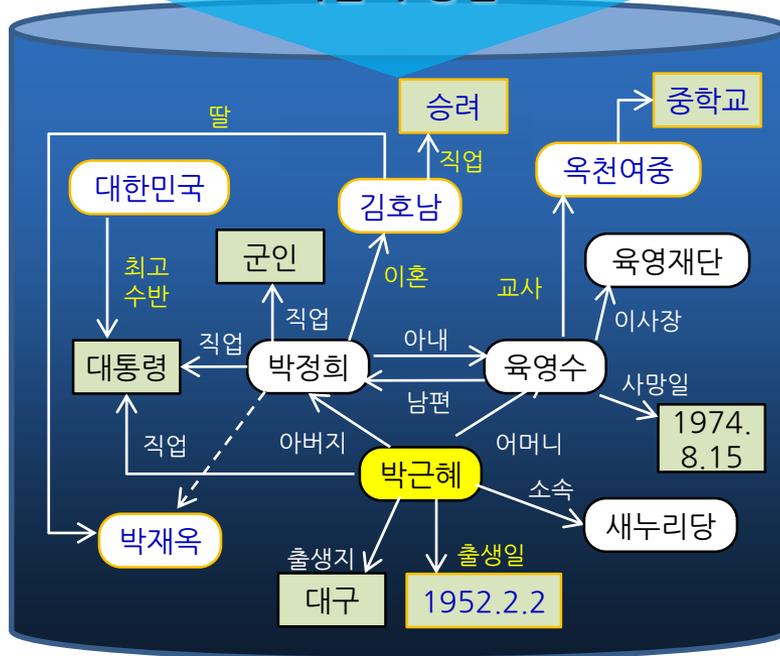
Krynews.co 2012-7-11 14:33:46 출처: 연합뉴스 기자:佚名 조회수: 8



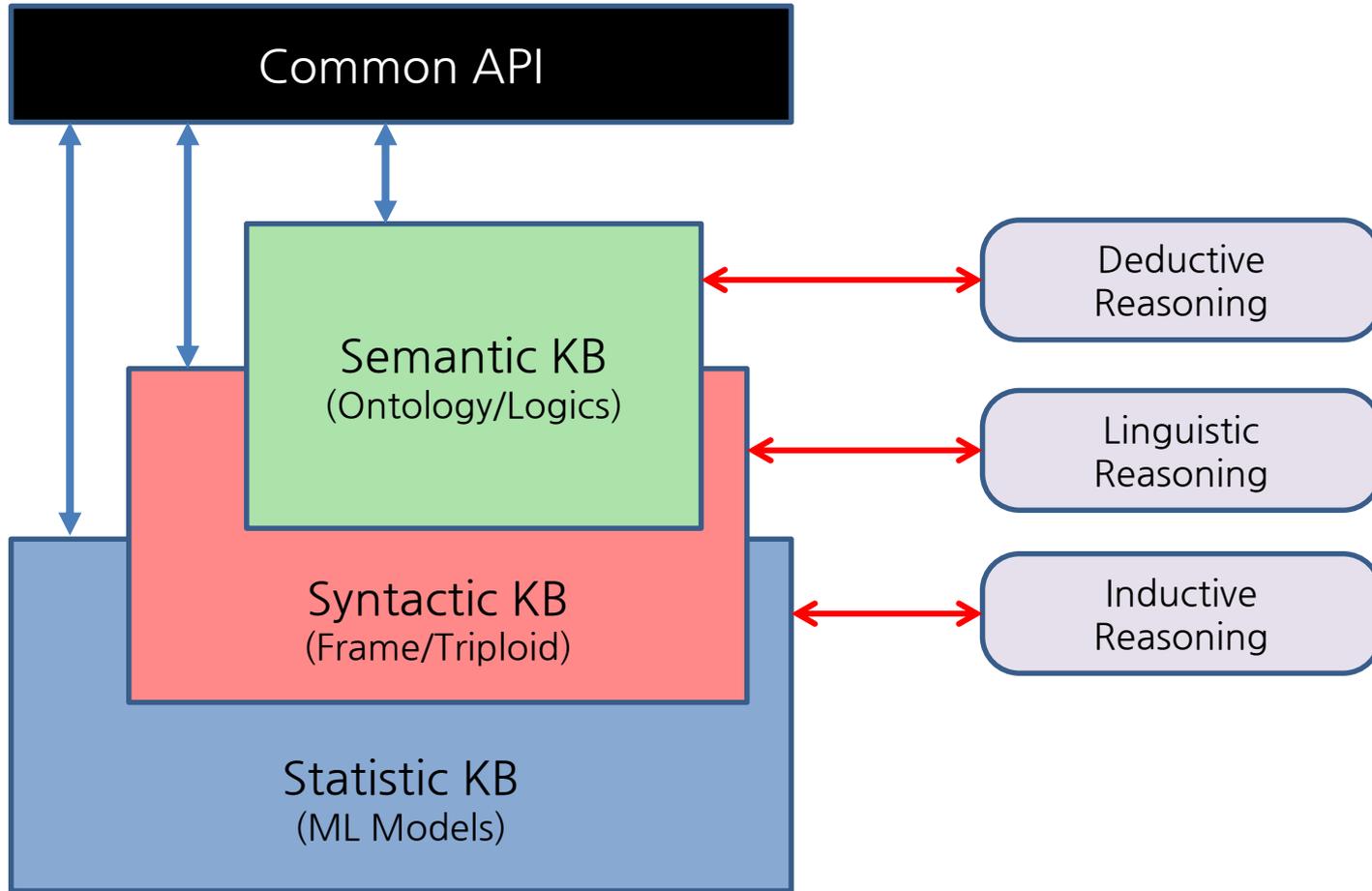
박근혜는 1952년 2월 2일 경상북도 대구시 삼덕동(현재의 대구광역시 중구 삼덕동 5-1번지, 5-2번지 일대)에서 전시(戰時) 대구 주재 육군본부 작전·교육국 작전차장 박정희 대령과 중등학교 교사 출신인 그의 부인 옥영수의 딸로 태어났다. 어머니 옥영수에게는 첫 소생이었으나 아버지 박정희는 이혼 경력과 전처소생의 장녀 박재옥이 있었으므로 박정희에게는 차녀가 된다. 여동생 박근령과 남동생 박지만이 있다.

출생일 1917년 11월 14일      국적 대한민국  
출생일 1952년 2월 2일 (61세)

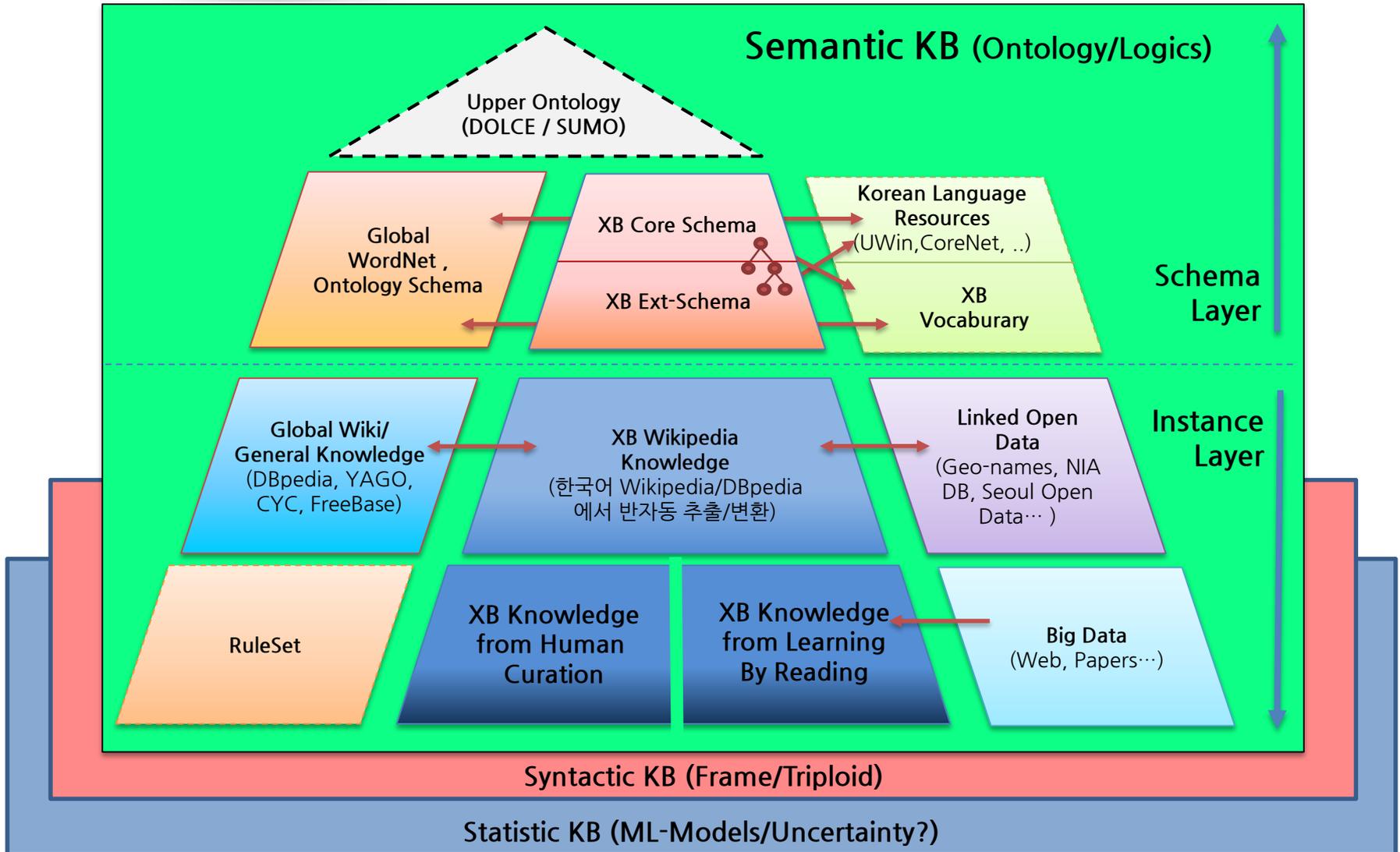
파편화된 지식의  
학습과 통합



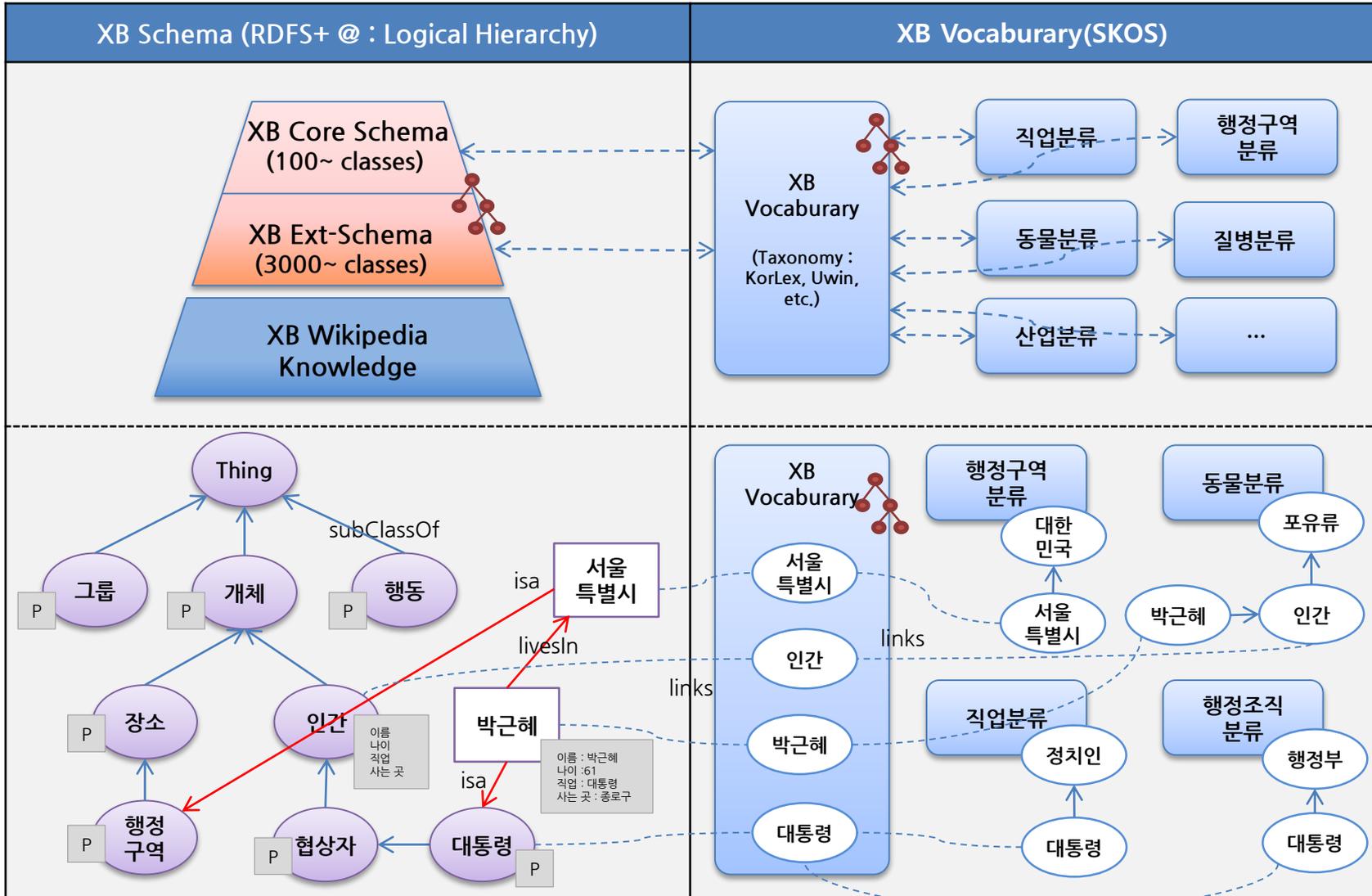
# XB Ontology Overview



# XB Ontology Overview



# XB Schema & XB Vocabulary





## 지식 큐레이션

- 고품질 코어 지식 구축과 개선
- 기존 지식 자원의 선별과 통합
- 지식 베이스 평가와 피드백

## 지식크라우드 소싱

- 게이미피케이션 적용 (모바일 퀴즈 게임)
- 일반지식의 확장과 검증

## 전문 지식의 검증

- WiseQA 실험 등 질의응답
- 도메인 전문가의 지식 검증



## 지식증강

- 복합상황추론
- 대규모 불완전 추론
- 추론 실패 대응형 문제 해결

## 자가지식학습

- Learning by reading
- Learning by taking advice

## 지식자원 수집, 통합 자동화

- 빅데이터 및 지식자원의 자동수집
- 부족 자원을 스스로 확보하는 기능
- 의미기반 자동 통합 기술의 구현



- XB 지식베이스 구축 현황

항목 구분	현재 규모 (2015.07.24 기준)	증감 (2015.01.13 기준)	비고
총 클래스 수	6,273 개	▲143	
총 프로퍼티 수	921 개	▲419	프로퍼티 상세화
총 인스턴스 수	11,492,006 개	▲약 1천만	백과사전 인스턴스 통합
총 트리플 수	1.2 억 개	▲약 1.05억	

# **XB Ontology Schema**

## Step1. XB Schema - Class Hierarchy 설계

- 1.1. KorLex Synset Hierarchy 추출, 상위 level concept 식별 (약 3,000개 클래스)
- 1.2. DBpedia ontology classes, ETRI NE\_Schema와 mapping - 호환/유효성 검증
- 1.3. 200개 수준의 Core Class Set 선정 및 검증 (1차년도 property 설계 대상)
- 1.4. Core Class에 대한 logical restriction 설정 (disjointWith)

## Step2. XB Schema - Property 설계

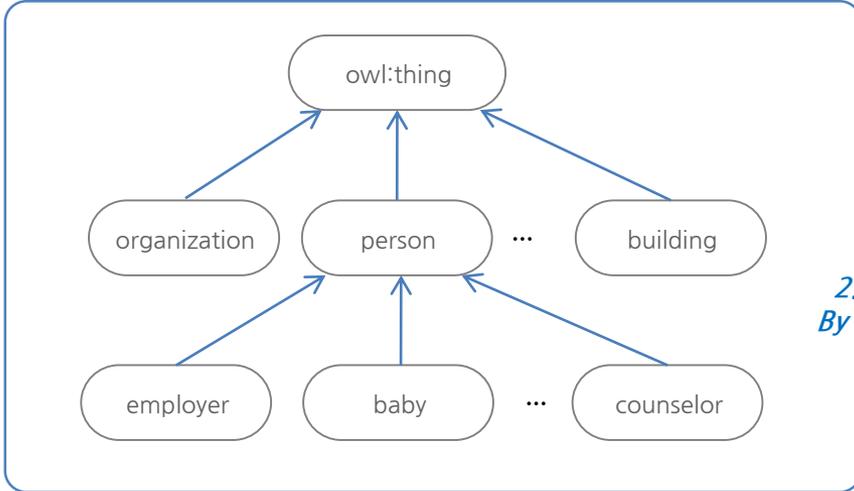
- 2.1. YAGO2 중심으로 DBpedia, ETRI NES 참조 property 설계 (Core Class 대상)
- 2.2. 한국/한국어 고유 개념 property 설계 (한국어 WP 참조, Extended Class 포함)
- 2.3. FOAF, SIOC, DC, GoodRelations 등 잘 알려진 vocabulary와 호환성 검증
- 2.3. Property restriction/axiom 설정, 검증 (domain/range, inverseOf, transitive...)

## Step3. XB Schema 검증

- 3.1. 한국어 DBpedia/Wikipedia entry 대상으로 class/property 유효성 검증
- 3.2. Uwin, CoreNet, YAGO Schema Mapping 및 consistency check

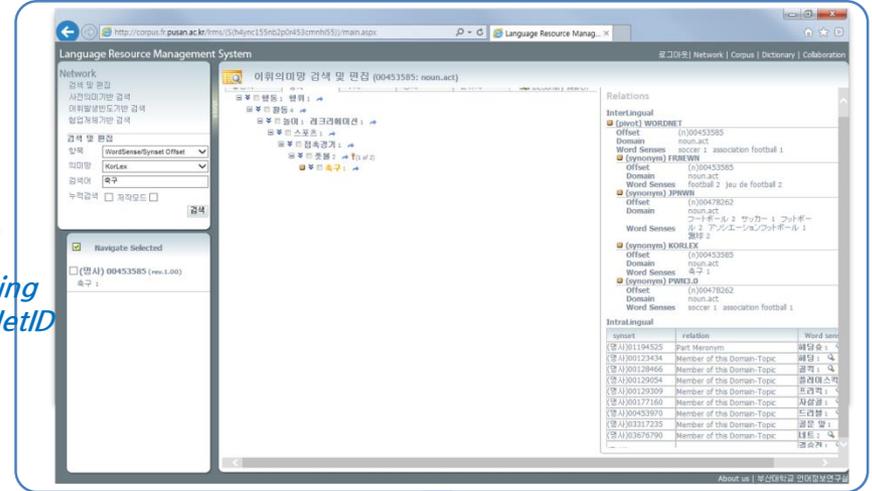
# XB Schema 구성 - 클래스 설계

## YAGO2 Class Hierarchy



1. Simpletax - Class 추출

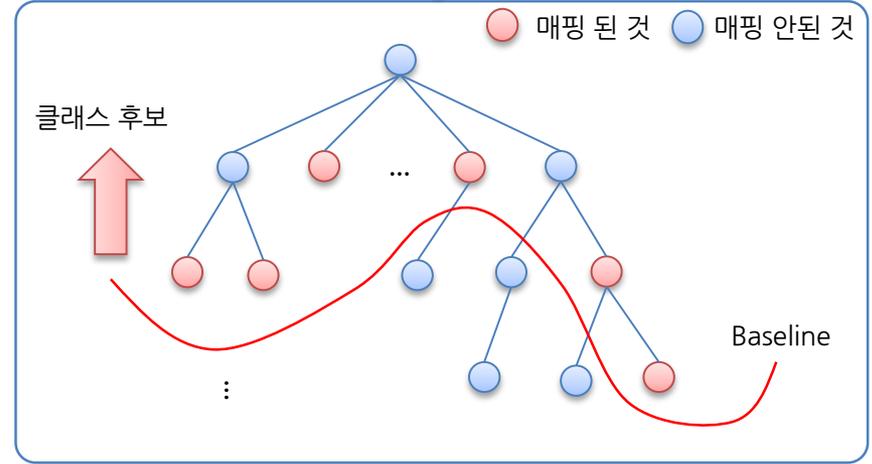
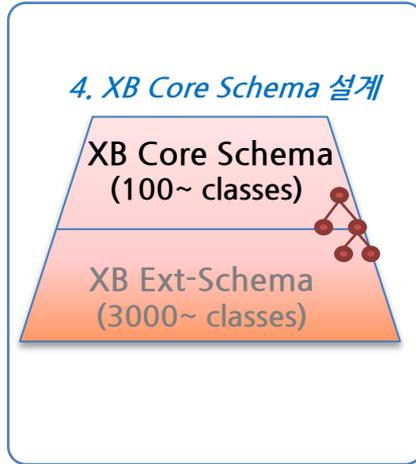
## KorLex Class Hierarchy



2. Mapping By WordNetID

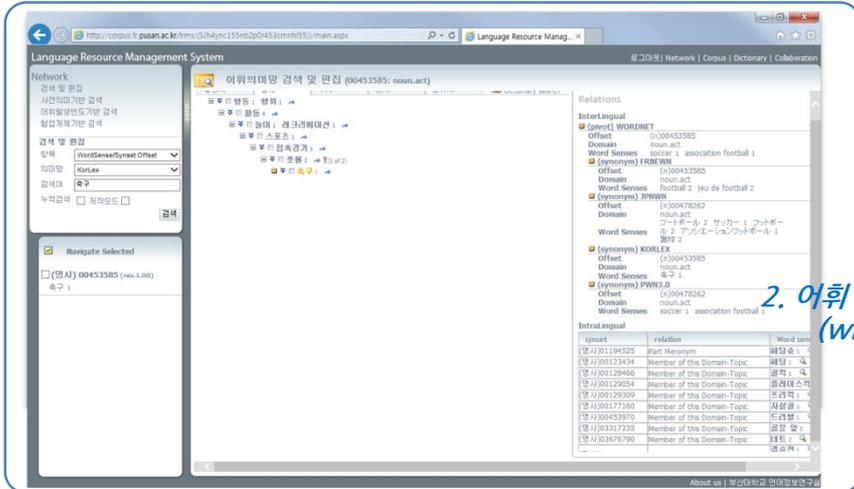
- ETRI NE 분류 참조  
- DBpedia 온톨로지 참조

5. 검증 및 보완



# XB Schema 구성 - 클래스 설계

## KorLex Class Hierarchy



1. Korlex 모든 어휘 추출(약 10만 개)

class	label	frequency
xbc:edge_03145955	둘레	34536878
xbc:edge_03145955	가	34536878
xbc:edge_03145955	언저리	34536878
xbc:edge_03145955	가장자리	34536878
xbc:edge_03145955	테두리	34536878
xbc:cotton_03002171	면포	26160189
xbc:cotton_03002171	면직물	26160189
xbc:cotton_03002171	목면	26160189
xbc:cotton_03002171	무명	26160189
xbc:cotton_03002171	면	26160189
xbc:man_03577697	알	21328216
xbc:man_03577697	말	21328216
xbc:cage_02832381	우리	21101479
xbc:cage_02832381	장	21101479
xbc:inside_08063710	안	18591973
xbc:inside_08063710		
xbc:person_0000		

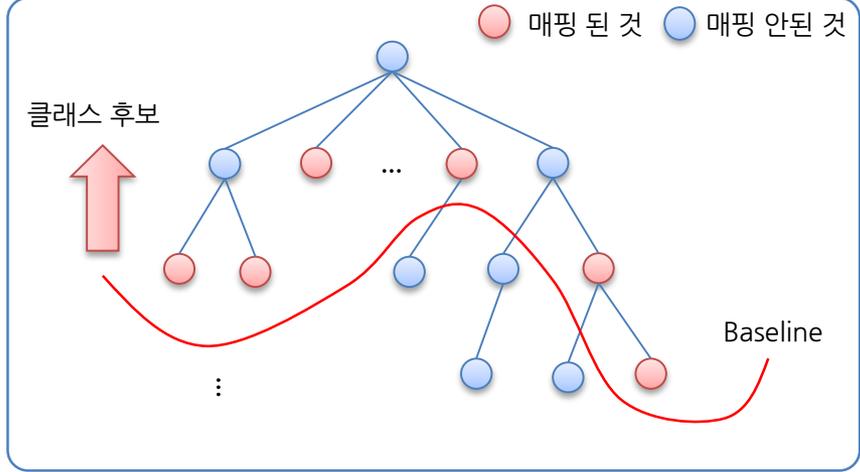
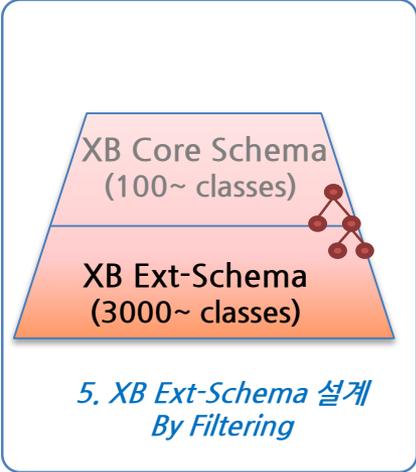
2. 어휘 출현 문서수 분석  
(with O2 Infra)

\* 문서수가 너무 적거나, 많은(불용어) 키워드는 제외

3. 출현 문서수 기준 필터링 (10만 ~ 1000만)

- XB Core Schema ,  
ETRI NE 분류,  
DBpedia 온톨로지  
참조

6. 검증 및 보완



# XB Schema 구성- Property 설계

## YAGO2 Properties

<\_hasTypeCheckPattern>  
 <\_timeToLocation>  
 <actedIn>  
 <byTransport>  
 <created>  
 <dealsWith>  
 <diedIn>  
 <diedOnDate>  
 <directed>  
 <edited>  
 <endedOnDate>  
 <endsExistingOnDate>  
 <exports>  
 <extractionSource>  
 <extractionTechnique>  
 <graduatedFrom>  
 <happenedIn>  
 <happenedOnDate>  
 <hasAcademicAdvisor>  
 <hasAirportCode>  
 <hasAnchorText>  
 <hasArea>  
 <hasBudget>  
 ...

2. Filtering  
By Manual

### - DBpedia 온톨로지 참조

Properties on Activity:

Name	Label	Domain	Range
abbreviation (edit)	abbreviation	owl:Thing	xsd:string
abstract (edit)	has abstract	owl:Thing	xsd:string
access (edit)	access	owl:Thing	xsd:string
accessDate (edit)	access date	owl:Thing	xsd:gYear
activeYearsEndDate (edit)	active years end date	owl:Thing	xsd:date
activeYearsEndYear (edit)	active years end year	owl:Thing	xsd:gYear
activeYearsStartDate (edit)	active years start date	owl:Thing	xsd:date
activeYearsStartYear (edit)	active years start year	owl:Thing	xsd:gYear
administrativeStatus (edit)	administrative status	owl:Thing	xsd:string
agglomerationPopulationTotal (edit)	agglomeration population total	owl:Thing	xsd:nonNegativeInteger
alphaCode (edit)	alpha code	owl:Thing	xsd:string
album (edit)	album	owl:Thing	Album
alias (edit)	alias	owl:Thing	xsd:string
alongside (edit)	alongside	owl:Thing	Person
animal (edit)	animal	owl:Thing	Animal
apcPresident (edit)	apc president	owl:Thing	Person
apoapsis (edit)	apoapsis	owl:Thing	Length
appointer (edit)	appointer	owl:Thing	Person
area (edit)	area	owl:Thing	Area
areaMetro (edit)	area metro	owl:Thing	Area

### 3. 검증 및 보완 (DBpedia 온톨로지 Class와 매핑 가능 한 XB Schema Class를 기준으로)

### - 한국어 위키피디아 참조



조선 세종  
朝鮮 世宗

위키피디아 인포박스

조선의 제4대 국왕

본명 이도(李祹)

재위 1418년 9월 9일 (음력 8월 10일) ~ 1450

### 1. XB Schema - Core Class와 매핑된 Class에 정의된 Properties 식별

v · d · e · h (숨기기)

인물 정보 틀의 종류

가공인물 정보 틀	가공인물 정보 · 동명인물 정보 · 인화이나 동명인물 정보 · 해리 포터 동명인물 정보
원주민 정보 틀	원주민 정보 · 연쇄 살인자 정보
스포츠인물 정보 틀	스포츠 인물 정보 · NBA 선수 정보 · 각주기 정보 · 각주 선수 정보 · 농구 선수 정보 · 바둑 기사 정보 · 베드민턴 선수 정보 · 스키 선수 정보 · 아이스 선수 정보 · 장기 기사 정보 · 종합격투기 정보 · 축구 선수 정보 · 축구 선수 정보 · 테니스 선수 정보 · 펜싱 선수 정보 · 프로레슬러 정보
역사인물 정보 틀	군주 정보 · 귀족 정보 · 소고 정보 · 왕족 정보 · 일본 무사 정보
예술인물 정보 틀	연예인 정보 · 영화인 정보 · 음악가 정보 · 희극인 정보 · 상무 정보 · 모델 정보 · 예술가 정보 · 요리사 정보 · 문학가 정보 · 작가 정보 · 프로그래머 정보 · 인화가 정보
전문직 정보 틀	건축가 정보 · 탐험가 정보 · 우주인 정보
정치인 정보 틀	간신 정보 · 공작자 정보 · 국가원수 정보 · 정치인 정보 · 총리 정보
종교인 정보 틀	가톨릭 성직자 정보 · 교황 정보 · 대주교 정보 · 성직자 정보 · 성인 정보 · 승려 정보 · 주교 정보 · 총대주교 정보 · 추기경 정보 · 천주교 대주교 정보
학자 정보 틀	경제학자 정보 · 공학자 정보 · 과학자 정보 · 철학자 정보
기타인물 정보 틀	인물 정보

위키피디아 틀 속성 정보(예 - 인물정보)

본명	출생명	태어날 때 지어진 이름, "이름"과 다름 때
출생일	출생일	태어난 날: {{출생일과 나이 1930 11}}
출생지	출생지	태어난 장소: 국가, 행정 구역, 도시.
사망일	사망일	죽은 날: {{사망일과 나이 1995 10 9}} 생일이 알려지지 않았을 경우).
사망지	사망지	죽은 장소: 국가, 행정 구역, 도시.
사망원인	사망원인	죽음의 원인.
발견지	body_discovered	신체가 발견된 장소. (죽은 장소와는 다를 수 있음)
매장지	restingplace	매장된 장소, 재류 무민 장소 등.
매장지좌표	resting_place_coordinates	매장된 장소, 재류 무민 장소 등의 좌표.

### 4. 한국어 위키피디아 정의 속성 추가 및 보완 (예 : 국호, 본관, 매장지, ...)

- 장학퀴즈, 적격 질의(Competency Question) 기반 검증
  - 예상되는 질의 유형들에 대한 다양한 예상 질문 및 답변 목록
  - 기존 장학퀴즈 기출문제 유형별 분석

문제	후보 클래스	후보 프로퍼티
1911년 일본 경찰이 민족운동을 탄압하기 위하여 신민회(新民會) 600여명을 검거하여 고문하고 105명을 기소한 사건. 이 사건으로 신민회(新民會) 조직은 와해되었다. 이 사건은 무엇인가?	시간 국가 (일본) 직업 단체 사건 사회운동	발생년도 원인 결과 피해
월드컵 역사상 처음으로 공동 개최한 17회 FIFA 월드컵의 두 나라를 나열하시오.	대회 월드컵 국가 (일본, 대한민국) 연맹	개최회차 개최국



후보 클래스	클래스 매핑	후보 프로퍼티	프로퍼티 매핑
시간	xbc:time_14406534	발생년도	xbp:isHappenedOnDate
국가 (일본)	xbc:Japan_08369977	원인	xbp:haveCauses
직업	xbc:bull_09249702	결과	xbp:hasResult
단체	xbc:event_00025950		
사건	xbc:movement_07951694		
사회운동	xbc:association_07560389		
대회	xbc:contest_07000331	개최회차	xbp:hasNumberOf
월드컵	xbc:federation_07802717	개최국	xbp:hasHeldCountry
국가 (일본, 대한민국)	xbc:South_Korea_08400201		
연맹	xbc:Japan_08369977		

## Step1. 한국어 Wikipedia instance mapping 및 importing

- 1.1. 한국어 Wikipedia category/infoBox 기반 class mapping 알고리즘 설계
- 1.2. Property mapping rule 생성, Semantic disambiguation rule 확보
- 1.3. 한국어 Wikipedia 인스턴스 변환 및 importing
- 1.4. restriction (domain/range 외) 위반, consistency 오류 점검

## Step2. 영어 DBpedia/YAGO, 일어DBpedia instance mapping 및 importing

- 2.1. 한국어 인스턴스에 연결된 영(일) DBpedia/YAGO로부터 부족 property 보완
- 2.2. 한국어 wiki에 없는 entry에 대해 영(일) DBpedia/YAGO instance 추출, 변환
- 2.3. Importing 오류, consistency 오류, 언어간 충돌 오류 등의 확인과 교정

## Step3. Curation Center를 통한 수동 구축과 검증

- 3.1. Wikipedia, 백과사전 등 비정형 데이터로부터 지식 추출
- 3.2. 자동 추출된 기존 XB 지식의 검증

- 클래스 매핑 방안 - Wikipedia 분류기반 클래스 매핑
  - 형태소 분석 + XRE 분석 활용

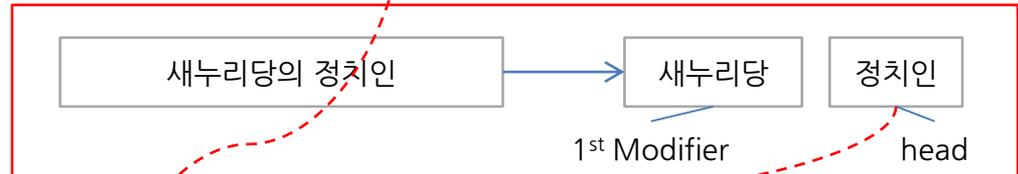
## 진영

위키백과, 우리 모두의 백과사전.

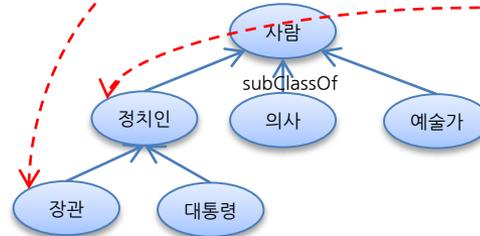
진영(陳永, 1950년 10월 23일 ~ , 전남 담양)은 법조인 출신의 대한민국 정치인이다. 여왕 진씨 고창 교흥 문중으로<sup>[1]</sup> 한국전쟁 중에 담양 외가에서 태어났다. 보건복지부 장관 내정자 기자회견에서 출신지 표기를 고창으로 요청했다.<sup>[2]</sup>

분류: 대한민국의 제17대 국회의원 | 대한민국의 제18대 국회의원 | 대한민국의 제19대 국회의원 | 1950년 태어남 | 살아있는 사람 | 대한민국의 법조인 | 새누리당의 정치인 | 대한민국의 보건복지부 장관 | 박근혜 정부의 장·차관급 공직자 | 경기고등학교 동문 | 서울대학교 법과대학 동문 | 담양군 출신 | 여왕 진씨

### 자연언어처리 (의존성 분석 / 개체명 인식 등)



```
Function wiki2wordnet(c)
Input: Wikipedia category name c
Output: WordNet synset
1 head=headCompound(c)
2 m =Modifier(c)
3 head=stem(head)
4 If there is a WordNet synset s for m + head
5 return s
6 If there is a WordNet synset s for head
7 return s
8 fail
```



진영 isa 장관

- 클래스 매핑 방안 - Wikipedia Infobox 타입(템플릿)기반 클래스 매핑

## 진영

위키백과, 우리 모두의 백과사전.

진영(陳永, 1950년 10월 23일 ~ , 전남 담양)은 법조인 출신의 대한민국 정치인이다. 여당 진씨 고향 교흥 문중으로<sup>[1]</sup> 한국전쟁 중에 담양 외가에서 태어났다. 보건복지부 장관 내정자 기자회견에서 출신지 표기를 고향으로 요청했다.<sup>[2]</sup>

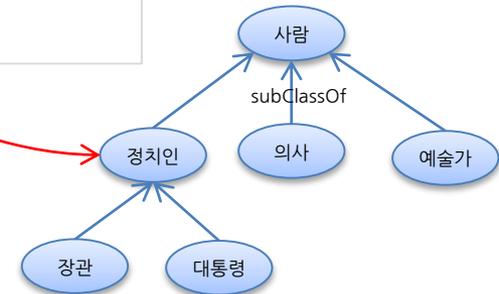
진영	
기본 정보	
국적	<span><span><span><span></span></span><span> </span></span>대한민국</span>
출생	1950년 10월 23일 (62세) 전라남도 담양
학력	서울대학교 법과대학 법학 학사 워싱턴 주립 대학교 법과대학원 법학 석사
국회의원 정보	
	새누리당
의원 선수	3선(17,18,19대)
정당	새누리당
지역구	서울 용산구
웹사이트	

Infobox

```

{{정치인 정보
|이름      = 진영
|그림      =
|국적      = {{국기나라|대한민국}}
|출생      = {{출생일과 나이|1950|10|23}}
|출생지    = [[전라남도]] [[담양]]
|사망      =
|학력      = [[서울대학교]] 법과대학 법학 학사 <br/> [[워싱턴 주립 대학교]] 법과대학원 법학 석사
|종교      =
|정당      = [[새누리당]]
|지역구    = [[서울특별시|서울]] 용산구
|의원 선수 = 3선(17,18,19대)
|경력      =
|홈페이지 = http://www.chinyoung.kr/
}}
    
```

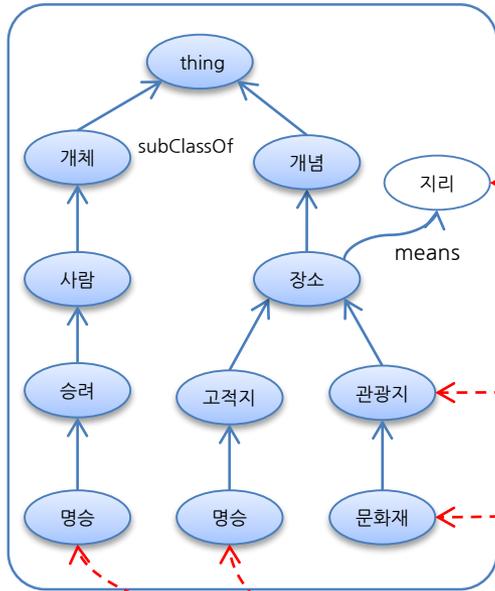
진영 isa 장관



```

Function wiki2wordnet(c)
Input: Wikipedia infobox type c
Output: WordNet synset
1 If there is a WordNet synset s for c
6 return s
7 fail
    
```

- 클래스 매핑 시 의미모호성 해소 방안 - 위키 카테고리 기반 클래스 후보 식별



## 소금강

위키백과, 우리 모두의 백과사전.

소금강(小金剛)은 강릉시 및 오대산 국립공원에 위치한 명승지이며, 1970년 11월 18일 대한민국 명승 제1호로 지정되었다. 지리적으로 오대산의 동쪽 기슭에 자리 잡고 있다.

- 목차 [숨기기]
- 1 지명 유래
  - 2 소금강의 자연 경관
  - 3 행사
  - 4 관련 인물

### 지명 유래 [편집]

옛날부터 맑은 폭포와 수려한 기암괴석, 빼어난 풍광을 자랑하여 그 모습이 마치 작은 금강산 같다고 하여 소금강이란 이름이 붙여졌다. 또한 학이 날개를 편 듯한 형상이라 하여 청학산이라고도 불린다.

분류: 강원도 (남)의 명승 | 강원도 (남)의 관광지 | 강릉시의 문화재 | 강릉시의 지리 | 오대산

## 명주 청학동 소금강 (溟州 靑鶴洞 小金剛)

대한민국의 명승

지정번호 명승 제1호  
소재지 강원 강릉시 연곡면 부연동길 753-13 등  
비고 1970년 11월 23일 지정



소금강 isa 명승  
소금강 isa 문화재

하나의 어휘가 여러 개의 클래스와 매핑 된 경우, XB Schema Hierarchy에서의 위치 및 출현 빈도를 고려하여 의미모호성 해소

- 클래스 매핑 시 의미모호성 해소 방안 - 인포박스 타입 기반 클래스 후보 식별

**박근혜**  
朴槿惠



박근혜 (2013년)

**대한민국의 제18대 대통령**

임기 2013년 2월 25일 ~ 2018년 2월 24일

국무총리 정홍원 (2013년 ~)

국적 **대한민국**

출생일 1952년 2월 2일 (61세)

출생지 **대한민국 경상북도 대구시 삼덕동**

정당 새누리당

학력 서강대학교 전자공학 학사

종교 무교<sup>[1]</sup> (로마 가톨릭교회 세례명: 율리이나, 불교 법명: 대자행(大慈行), 선덕화(善德華))<sup>[2]</sup>

부모 부 박정희, 모 육영수

서명 

웹사이트 박근혜 - 트위터<sup>[3]</sup>

인물 정보 템플릿

WikiProperty	XB KB property
국적	nationality
이름	personName
...	...

Property 매핑 테이블 예

WikiProperty	XB KB property
인물:이름	personName
기업:이름	companyName

Infobox 타입별 의미모호성 해소 규칙 예

기업 정보 템플릿

**삼성전자**  
Samsung Electronics Co., Ltd.



산업 분야 전자제품 제조 등

창립 1969년 1월

시장 정보 한국: 005930 [📄](#)

업종: SMSN [📄](#), 런던: SMSD [📄](#)

프랑크푸르트: SSU [📄](#)

프랑크푸르트: SSUN [📄](#)

국가 **대한민국**

본사 경기도 수원시 영통구 삼성로 129

핵심 인물 이건희 (회장)  
이재용 (부회장)  
권오현 (대표이사 부회장)

매출액 201,110,000(백만원) (2012년 12월)

주요 주주 삼성전자 자사주 14.15%  
삼성생명보험 7.53%  
국민연금공단 7.20%<sup>[1]</sup>  
씨티은행(DR) 6.74%  
캐피탈 리서치 앤 매니지먼트 컴퍼니 5.03%  
삼성물산 4.06%  
이건희 3.43%  
삼성화재해상보험 1.26%  
홍라희 0.74%

- Property 매핑 규칙 기반의 인스턴스 및 정보 생성

- category에 대해 확장된 정규 표현식 (XRE)을 기반으로 XB Schema에 정의된 property와의 매핑 규칙 생성
- XRE(eXtended Regular Expression) : 형태소 분석 + Grammar Structure

WikiCategory	정규표현식	XB KB property
1952년 태어남	[0-9]*년 태어남	birthYear
중국의 산	[/WS]*의 산	locatedIn

정규표현식 기반 property 매핑 규칙 예

- Modifier의 개체명 분석을 통한 의미모호성 해소

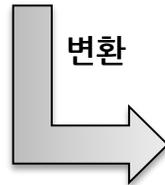
WikiCategory	NamedEntity	XB KB property
숙명여대 출신	(OGG_EDUCATION) 출신	graduateOf
강원도 출신	(LC_PLACE) 출신	placeOfBirth

개체명 기반 의미모호성 해소 규칙 예

- 변환 규칙 예시(인물 출생일/사망일)

```
[person_date:  
  (?uri xbp:wasBornOn ?birthdate),  
  (?uri xbp:wasDiedOn ?deathdate)  
  
  <- (  
    ?uri.bind('xbr:' + fn:replace(label, ' ', '_')),  
    ?birthdate.bind(xbf:parse_date(출생일)),  
    ?deathdate.bind(xbf:parse_date(사망일))  
  )  
  
  <- view.mysql_01(  
    field('all'),  
    table(xbv:infobox(q1, '인물 정보'))  
  )  
]
```

규칙 예시



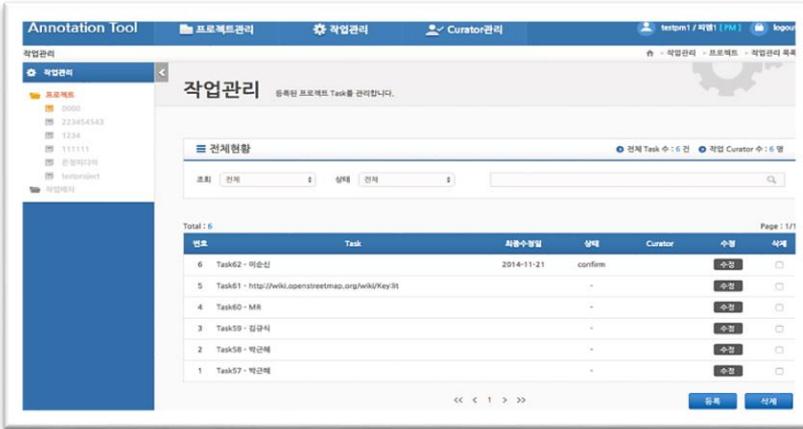
변환

```
xbr:유재석  rdf:type  xbc:person_00006026 ;  
  xbp:hasName      "유재석" ;  
  xbp:wasBornIn   xbr:대한민국 ;  
  xbp:wasBornOn   "1972-08-14" .
```

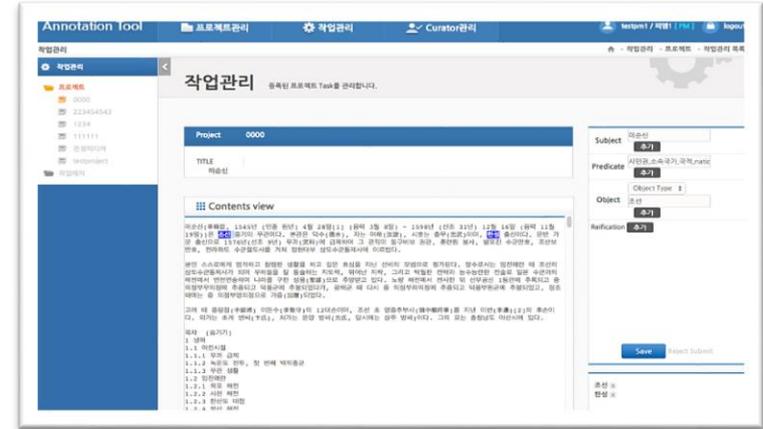
```
xbr:이순신  rdf:type  xbc:person_00006026 ;  
  xbp:hasName      "이순신" ;  
  xbp:wasBornOn   "1545-04-28" ;  
  xbp:wasDiedOn   "1598-12-16" ;
```

변환 결과 예시

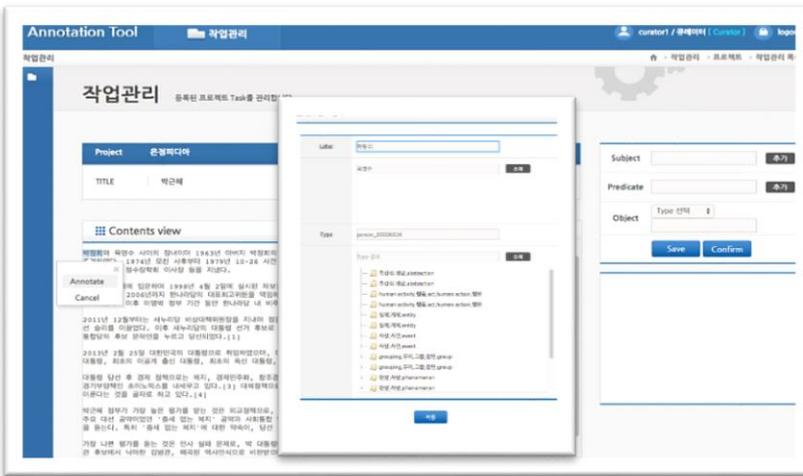
- 큐레이션 도구 개발



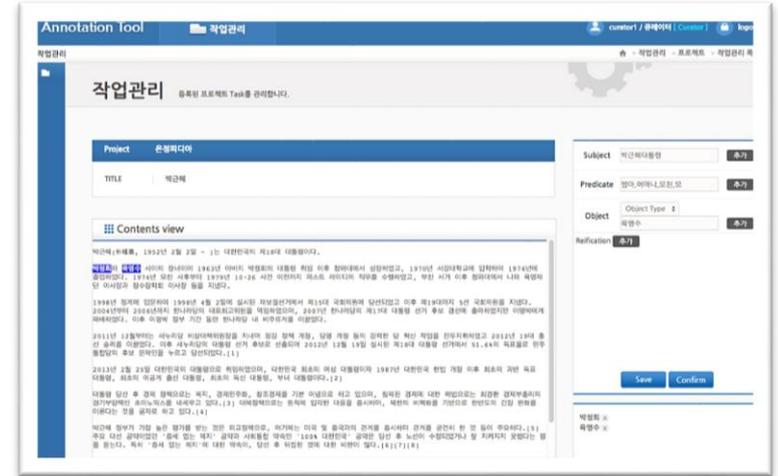
<PM-작업 관리 화면>



<PM-큐레이터 작업검토 및 수정 화면>



<큐레이터-큐레이션 작업 화면 1>



<큐레이터-큐레이션 작업 화면 2>

## Step1. Linked Open Data, WikiData, Open CYC 등 연계, 확장

- 1.1. GeoNames, Music Brainz, Open Library 등 LOD 연계, XB instance 확장
- 1.2. WikiData와 Open CYC 등 개방된 지식베이스 연계 및 XB instance 확장
- 1.3. 기존 instance와 충돌, inconsistency, 중복데이터 문제 해결 (sameAs 등)
- 1.4. 데이터 통합 프로세스, 데이터 생명주기 관리, 품질 관례 체계 확보

## Step2. 국내 DB, 개체명, 한국어 어휘 확장

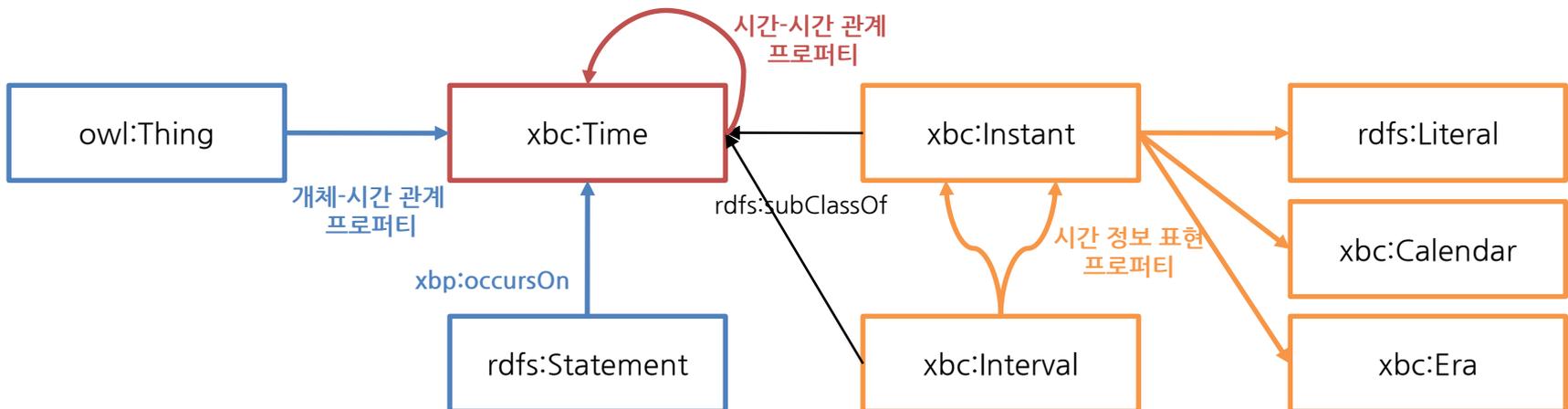
- 2.1. 인명DB, 지명DB, 기관DB, 상품DB 등 국내 DB 어휘 확장 (NIA, 서울시 등)
- 2.2. 표준국어대사전(incl. UWin), 자동 추출된 개체명 등 한국어 어휘 확장
- 2.3. 확장 어휘 검증 및 품질 관리

## Step3. Learning by Reading 결과 통합, 검증

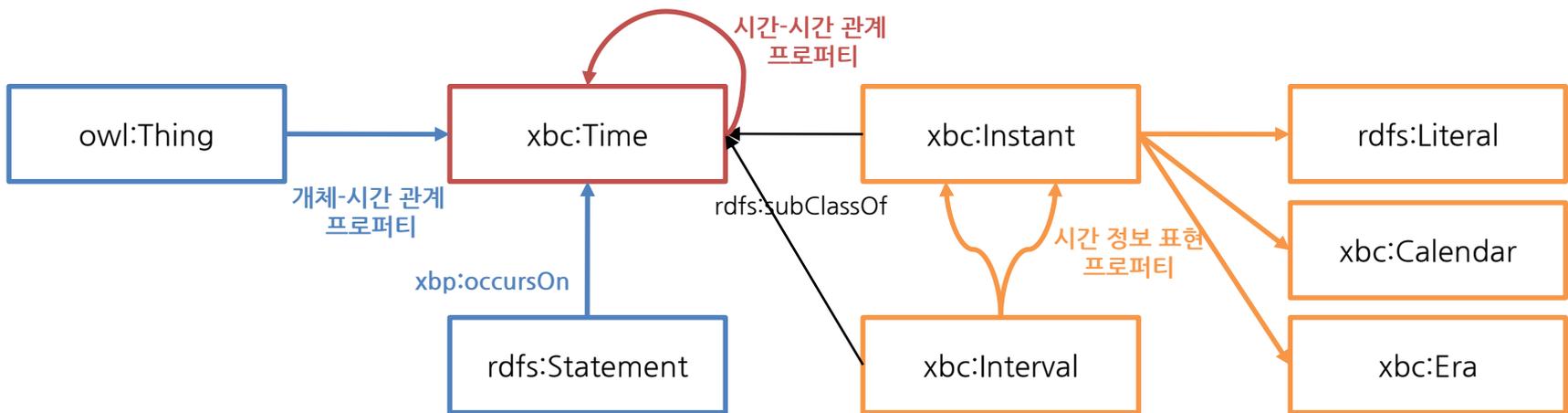
- 3.1. Learning by Reading을 통해 학습된 지식의 연계와 2차 검증 (consistency 등)
- 3.2. 지식 통합, 확장 프로세스, 생명주기 관리 체계 확보

# **XB 지식 표현**

- 시간 정보 표현 스키마
  - 시간 자체를 표현하기 위한 클래스와 프로퍼티
- 개체-시간 관계 프로퍼티
  - 사람, 조직, 이벤트 등의 개체와 관련한 시간을 표현하기 위한 프로퍼티
  - xbp:XXXOn과 같이 On으로 끝남 (예) xbp:startedOn, xbp:bornOn
  - 시간 정보 Reification을 위한 프로퍼티인 xbp:occursOn도 포함됨
- 시간-시간 관계 프로퍼티
  - 시간과 시간 간의 관계를 표현하기 위한 프로퍼티
  - xbp:tm\_XXX와 같이 tm\_으로 시작됨 (예) xbp:tm\_starts, xbp:tm\_equals

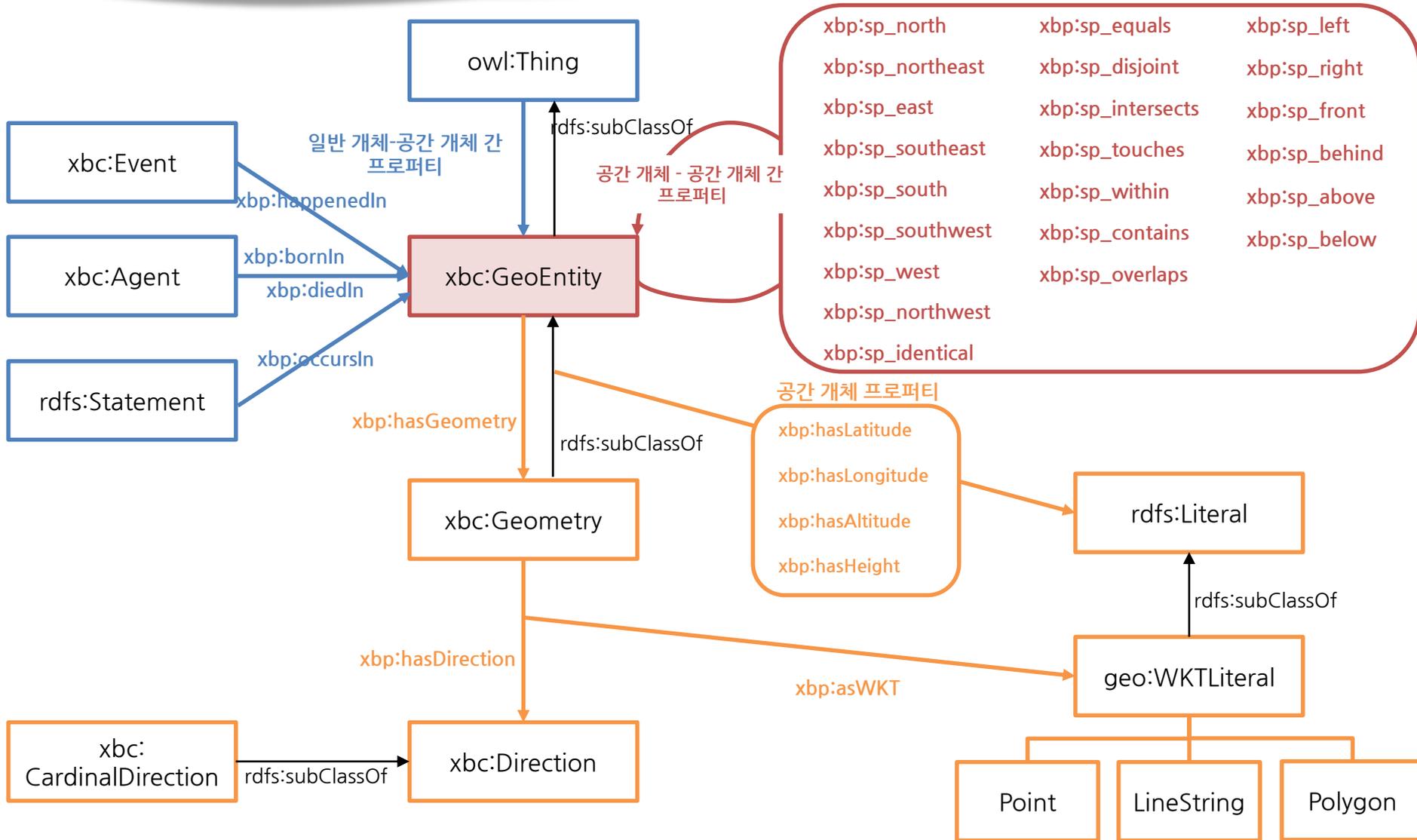


- xbc:Time
  - 시간을 표현하는 추상적인 클래스
- xbc:Instant
  - xbc:Time의 서브 클래스로 특정 시간적 Point를 표현함
  - 시대, 달력, 연, 월, 일, 시, 분, 초 등의 값을 가짐
- xbc:Interval
  - xbc:Time의 서브 클래스로 특정 시간적 범위를 표현함
  - xbc:Instant 타입의 시작시간과 종료시간을 가짐
- xbc:Era
  - 조선시대, 조선 세종시대 등을 표현하기 위한 클래스



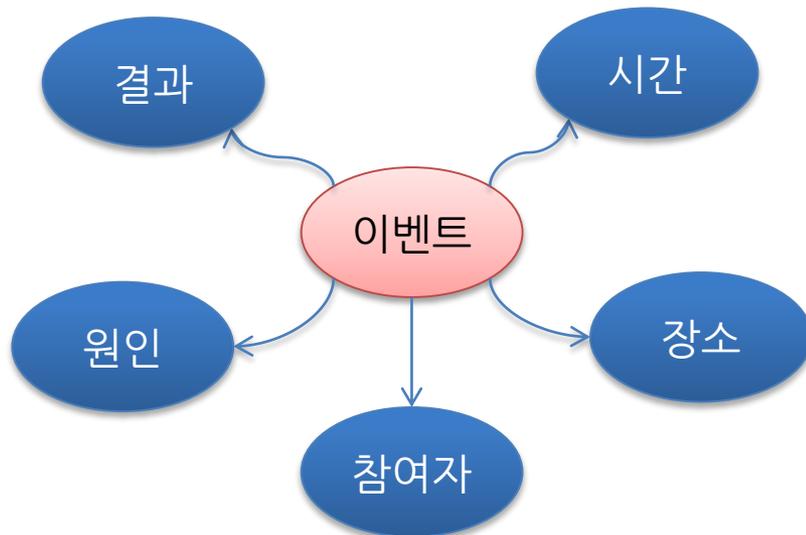
- 공간 개체 스키마
  - 공간 개체를 표현하기 위한 클래스와 프로퍼티
- 일반 개체-공간 개체 간 프로퍼티
  - 사람, 조직, 이벤트 등의 개체와 관련한 공간 개체를 표현하기 위한 프로퍼티
  - xbp:XXXIn과 같이 In으로 끝남 (예) xbp:startedIn, xbp:locatedIn
  - 공간 정보 Reification을 위한 프로퍼티인 xbp:occursIn도 포함됨
- 공간 개체-공간 개체 간 프로퍼티
  - 공간 개체와 공간 개체 사이의 관계를 표현하기 위한 프로퍼티
  - xbp:sp\_XXX와 같이 sp\_로 시작됨 (예) xbp:sp\_west, xbp:sp\_within, xbp:sp\_left

# 지식 표현 - 공간 정보



- xbc:GeoEntity
  - 공간 개체를 표현하는 클래스
- xbc:Geometry
  - xbc:GeoEntity의 서브 클래스로 점, 선, 면 등 기하학적 개체를 표현하는 클래스
- xbc:Direction
  - xbc:GeoEntity 자신의 절대적인 방향을 표현하는 클래스
- xbc:CardinalDirection
  - xbc:Direction의 서브 클래스로 동, 서, 남, 북과 같은 이산화된 방향을 표현하는 클래스

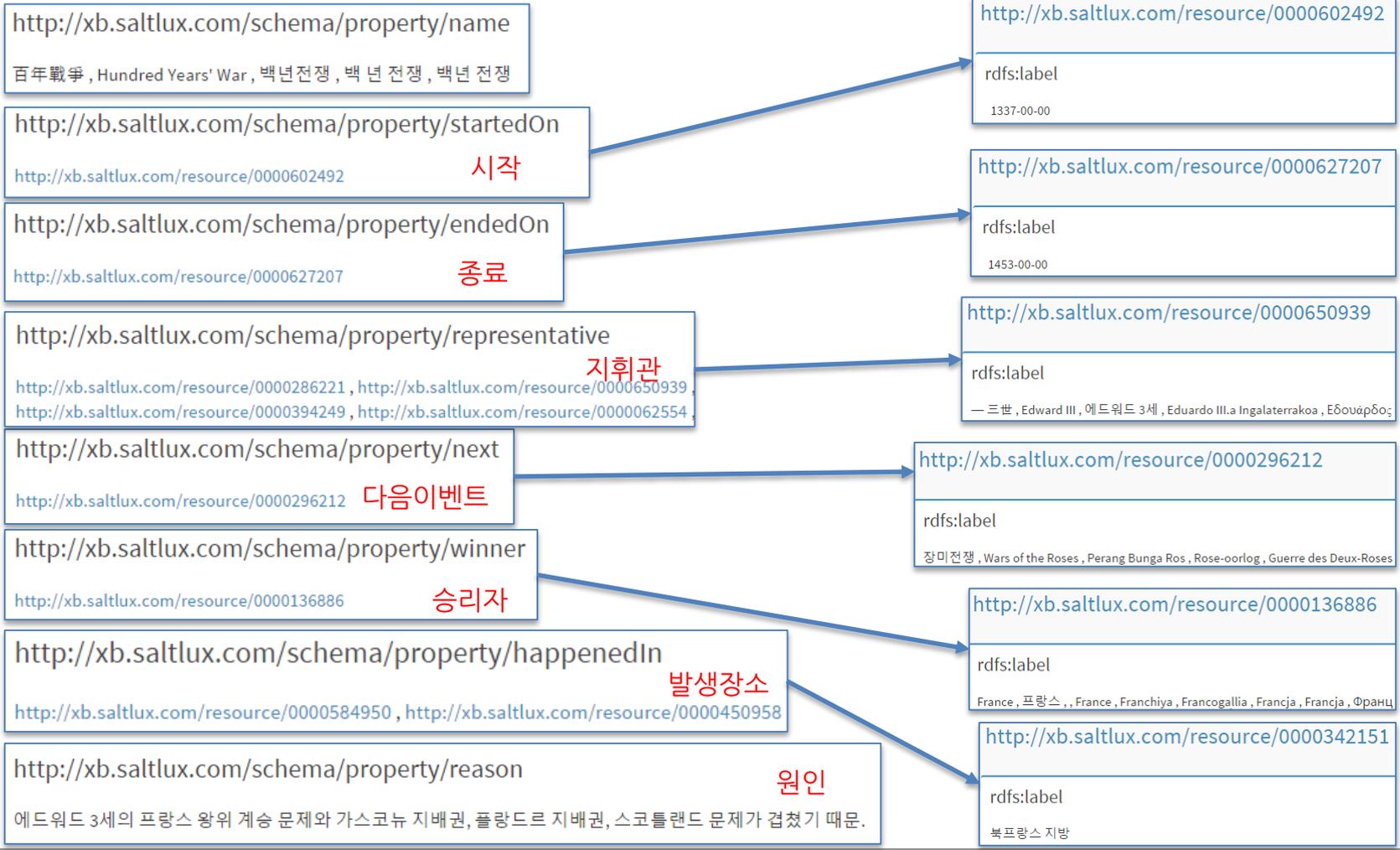
- 이벤트 지식 표현
  - 시간: 시작, 종료, 기간, 주기 등
  - 장소
  - 참여자: 인물, 조직 등
  - 원인
  - 결과: 피해자, 피해금액, 다른 이벤트 등
  - 연관 이벤트: 하위, 다음 등



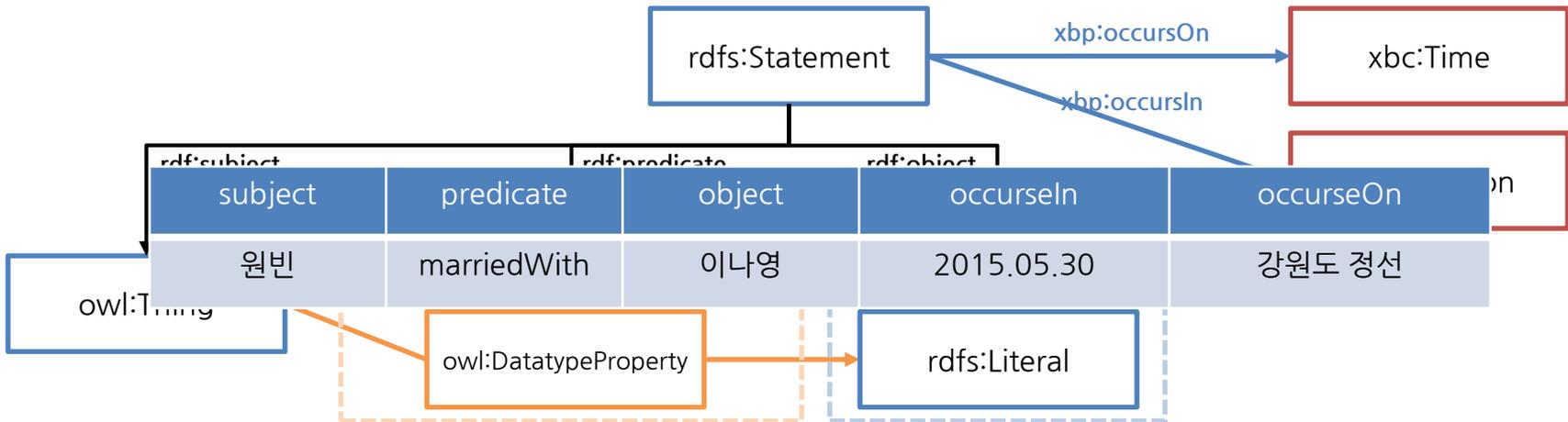
## <xbc:event\_00025950>의 하위클래스

accident_06852933		
happening_06836234		
accompaniment_06837114	film_festival_00492511	
affair_06992125	final_07010653	fire_06854630
social_event_06841042	match_07014432	convergence_06961171
affliction_06856683	finish_06901957	conversion_06903775
trouble_06841386	flash_06958577	creation_06879256
animation_06208045	footrace_07003961	cross_country_07004632
movie_06205452	forest_fire_06855369	damage_06966762
appearance_06872562	formality_06999527	dance_06993581
athletic_contest_07000745	funeral_06996019	death_06904072
contest_07000331	game_00432385	decrease_06904458
automobile_race_0700275	game_show_06212025	dedication_06996868
race_07002329	horse_race_07005250	deformation_06906613
beginning_06843132	impact_06888434	destiny_06880285
bicycle_race_07003107	increase_06905247	destruction_06884710
birth_06871123	initiation_06997715	ending_06843539
change_06848633	interruption_06916077	development_06969552
birthday_party_06992985	lightning_06959424	disappearance_06885918
party_06992505	miracle_06841204	discharge_06859110
boat_race_07003380	misfortune_06856319	easing_06905654
broadcast_06210603	movement_06861102	electrical_discharge_06859251
show_06210248	musical_06589822	emergence_06875415
case_06860210	play_06589599	episode_06212699
ceremony_06995452	mutation_06970967	episode_06842237
collision_06854335	noise_06975545	error_06851542
contact_06889092	origin_06874664	exercise_06998972
concert_06469093	outbreak_06980307	experience_06837915
performance_06467811	outburst_06979976	explosion_06859884
conflagration_06855245	periodic_event_0684839	explosion_sound_06938889
	playing_06469847	sound_06919436
	playoff_07009111	failure_06868584
		festival_06995261

- 백년전쟁: <<http://xb.saltlux.com/resource/0000375362>>



- 부가 정보 표현 스키마
  - 트리플에 대한 시공간 부가 정보를 표현하기 위한 클래스와 프로퍼티
  - rdfs:Statement 클래스를 사용하여 reification 표현
    - 기본 프로퍼티 : rdf:subject, rdf:predicate, rdf:object
- 시간 표현 프로퍼티
  - xbp:occursOn 프로퍼티를 사용하여 시간 표현
- 공간 표현 프로퍼티
  - xbp:occursIn 프로퍼티를 사용하여 공간 표현



# **XB Ontology Instance**



최신, 실시간, 불확실성, 개별

오래된, 주기적, 안정성, 전체



**역사** (숨기기)

- 1월 1일 1 시카고 지역사회 활동가 버락 오바마
- 2월 1일 1 시카고 거항
  - 2.1 주 의원 (1997년 ~ 2004년)
  - 2.2 2004년 미국 상원 선거 출마
  - 2.3 미국 상원 (2004년 ~ 2008년)
  - 2.3.1 입법
  - 2.3.2 상원 위원회
- 3 2008년 선거 출마
- 4 대통령 당선
- 5 1월 20일 취임
- 6 2009년 대통령을 증경하는 오바마 대통령
- 7 주요 정책과 비판
  - 7.1 이라크 철수문제

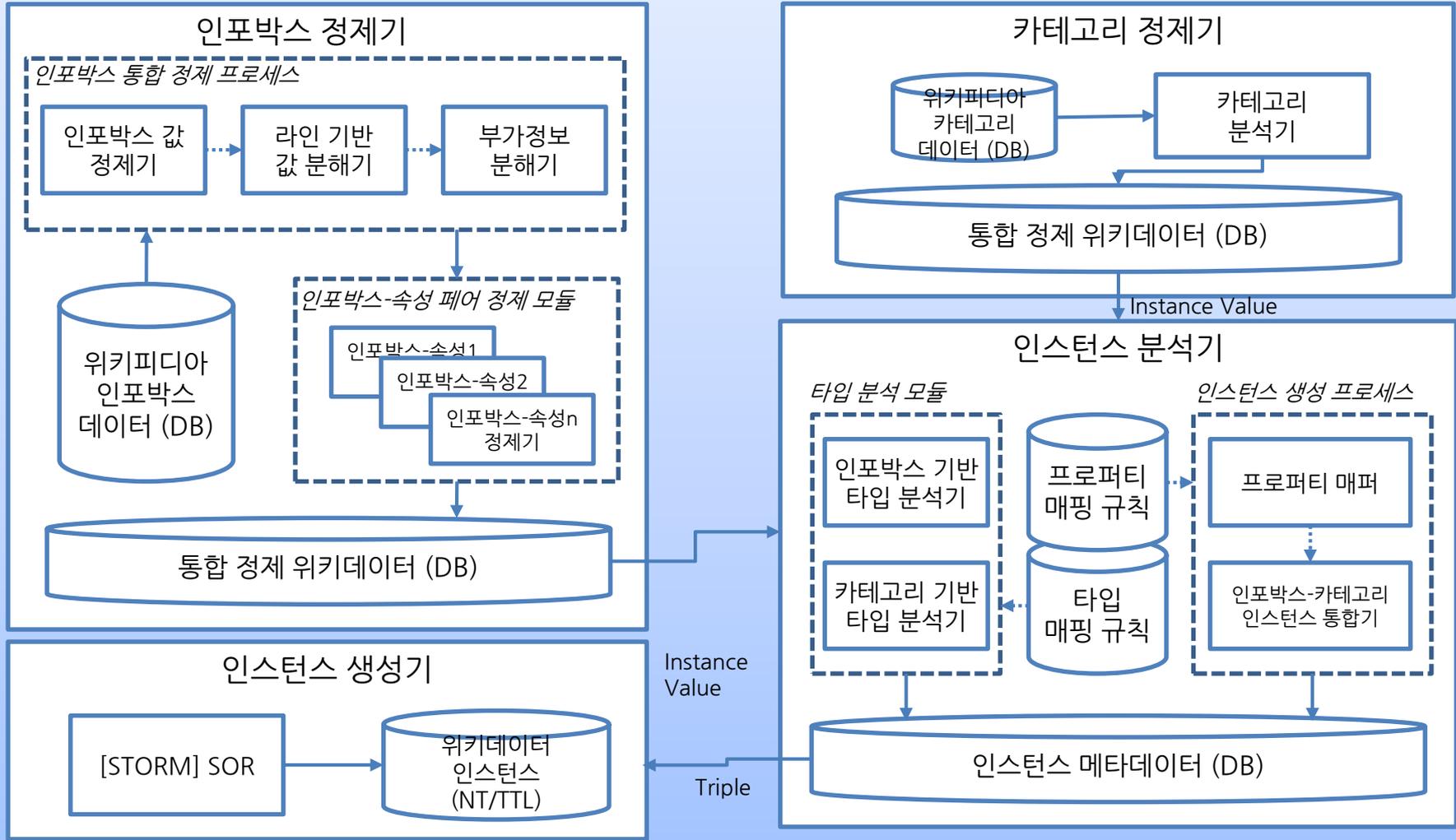
{{대통령 정보  
 |이름 = 버락 오바마 {{노벨상 딱지}}  
 |원래 이름 = Barack Hussein Obama II  
 |명칭 = 대통령  
 |그림 = President Barack Obama.jpg  
 |크기 = 250px  
 |설명 = 버락 오바마  
 |국가 = 미국  
 |대수 = 44  
 |취임일 = 2009년 1월 20일  
 |퇴임일 = 2017년 1월 20일 (예정)  
 ... }}

[[분류:버락 오바마]]  
 [[분류:1961년 태어남]]  
 [[분류:살아있는 사람]]  
 [[분류:미국의 정치인]]  
 [[분류:미국의 대통령]]  
 ...

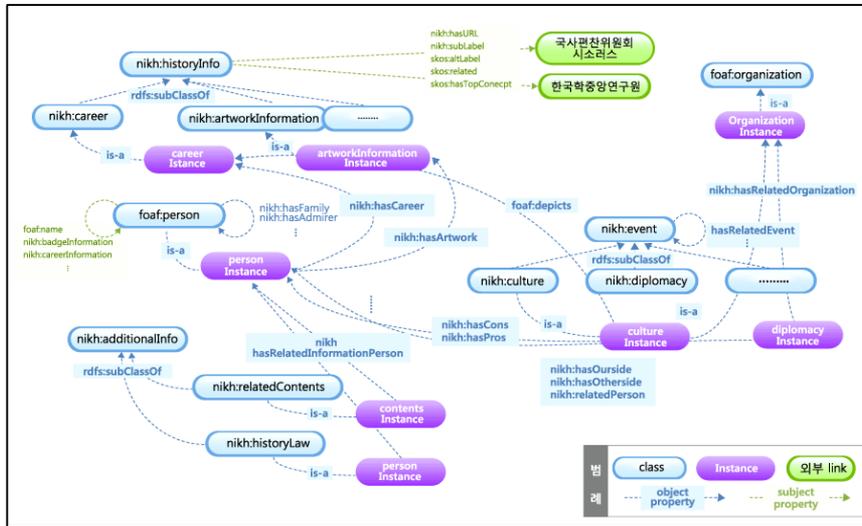
분류: 버락 오바마 | 1961년 태어남 | 살아있는 사람 | 미국의 정치인 | 미국의 대통령 | 케냐의 미국인 | 아프리카계 미국인 | 아일랜드계 미국인 | 영국계 미국인 | 독일계 미국인 | 하와이 주 출신 | 할리우드 대학교 | 윌슨 하버드 로스쿨 | 윌슨 | 시카고 대학교 교수 | 2008년 미국 대통령 선거 | 스포츠맨들계 미국인 | 2009년 미국인 | 노벨 평화상 수상자 | 미국의 노벨상 수상자 | 그레이의 수상자 | 노벨평화상 출신 | 현재 국가 지도자 | 아일랜드계 미국인 | 미국의 개신교도



## 위키피디아 인스턴스 자동 생성기



	클래스	프로퍼티	트리플
한국사LOD	30 개	157 개	약 4.5만 개
한국어DB피디아	148 개	9,558 개	약 825만 개
GeoNames	18 개	28 개	약 8,000만 개
LinkedGeoData	1,200 개	225 개	약 4,000만 개



한국사 LOD Schema Class	엑소브레인 Schema Class	
person	person	
additionalInfo	불필요	
country	Nation	
historyLaw	Law	
historyPolicy	Policy	
historySystem	프로퍼티 의미	한국사 LOD Schema Property
relatedContents	저작물종류	artworkKind
event	서훈내용	badgeContent
culture	시작일	beginDate
diplomacy	출생일	birthDate
economy	출생지	birthPlace
policy	봉호	bonghoName
politics	부명	buName
socialEvent	개명	changedName
war	이명	childhoodName
	초명	choName
	시간정보	date
	사망일	deathDate
	사망지	deathPlace
	상세설명	detailedExp
	한국민족문화대백과사전 카테고리	encyKoreaCategory
	종료일	endDate

〈한국사LOD 스키마 분석 예〉

**KB QA**

# XB QA를 위한 SPARQL 자동 생성 모듈

- 자연어 질문에 대한 의미적 정답제약 정보를 포함한 의사질의 (Pseudo Query)를 XB 지식베이스에 기반한 SPARQL로 자동 변환
- 표준질의 중 400개의 주요 질문 분석을 통해 변환 모델 생성



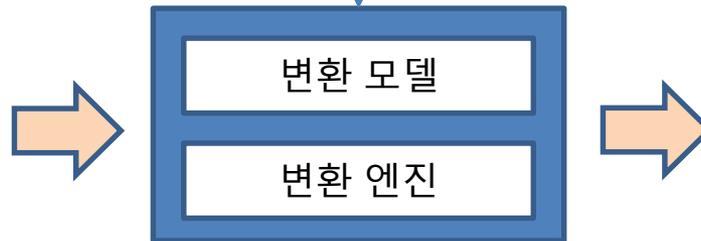
질문분석

```
{
  "qid": "202",
  "sgid": "1",
  "question": "에언자 마호메트에게 전해진 알라의 계시와 계율을 집대성한 이슬람교의 경전",
  "satlist": {
    "AFW_RELIGION"
  },
  "latlist": [
    "경전"
  ],
  "negation": "false",
  "SubQSemFrameList": [
    {
      "target_variable": "X",
      "negation": "false",
      "conditions": [
        {
          "Predicate": {
            "pred_type": "PT_ATYPE",
            "pred_string": "PT_ATYPE",
            "negation": "false",
            "norm_superlative": ""
          },
          "Arguments": [
            {
              "arg_type": "AT_VARIABLE",
              "arg_value": "X",
              "norm_superlative": "",
              "lang_code": "",
              "syn_rel": "",
              "sem_role": "",
              "ne_list": []
            }
          ]
        }
      ]
    }
  ]
}
```

1세부 Pseudo Query

NLQ (자연어 질의)	SPARQL
1911년 일본 경찰이 민족운동을 탄압하기 위하여 신민회원(新民會員) 600여명을 검거하여 고문하고 105명을 기소한 사건 이 사건으로 신민회(新民會) 조직은 와해되었다. 이 사건은 무엇인가?	<pre>SELECT ?e WHERE {   ?e rdfs:type xbc:culture_00025950.   ?e xbp:hasHappenedReason ?hr FILTER (REGEX (STR(?hr), "민족운동") &amp;&amp; REGEX (STR(?hr), "탄압")).   ?e xbp:hasResult ?r FILTER (REGEX (STR(?hr), "신민회") &amp;&amp; REGEX (STR (?hr), "와해")).   ?e xbp:isHappenedOn ?t FILTER REGEX (STR(?t), "1911"). }</pre>
다음 중 문명과 해당 문명이 발생했던 강을 잘못 짝지어진 것은? 1. 메소포타미아 문명 - 티그리스 강 2. 인더스 문명 - 인더스 강 3. 이집트 문명 - 미시시피 강 4. 황하 문명 - 황하	<pre>SELECT ?s ?h WHERE {   ?s rdfs:type xbc:culture_07787480 FILTER REGEX (STR(?s), "메소포타미아 문명").   ?s xbp:isStartedIn ?h . } SELECT ?s ?h WHERE {   ?s rdfs:type xbc:culture_07787480 FILTER REGEX (STR(?s), "인더스 문명").   ?s xbp:isStartedIn ?h . } SELECT ?s ?h WHERE {   ?s rdfs:type xbc:culture_07787480 FILTER REGEX (STR(?s), "이집트 문명").   ?s xbp:isStartedIn ?h . } SELECT ?s ?h WHERE {   ?s rdfs:type xbc:culture_07787480 FILTER REGEX (STR(?s), "황하 문명").   ?s xbp:isStartedIn ?h . }</pre>

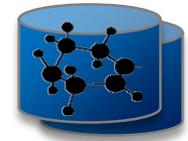
변환 모델 학습



자동 변환기

주요 자연어 질의  
(400여 개) 분석

지식베이스



```
{
  "PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  SELECT ?X (COUNT(?X) AS ?CNT)
  WHERE {
    ?X a
    <http://xb.saltlux.com/schema/class/sacred_text_06030797
    >?X <http://xb.saltlux.com/schema/property/content>
    <http://xb.saltlux.com/resource/kowiki/에언자> .
  }
  UNION
  ?X a
  <http://xb.saltlux.com/schema/class/sacred_text_06030797
  >?X <http://xb.saltlux.com/schema/property/content>
  <http://xb.saltlux.com/resource/kowiki/무함마드> .
  }
  UNION
  ?X a
  <http://xb.saltlux.com/schema/class/sacred_text_06030797
  >?X <http://xb.saltlux.com/schema/property/content> ?I
  FILTER contains(str(?I), "마호메트") .
  }
  UNION
  ?X a
  <http://xb.saltlux.com/schema/class/sacred_text_06030797
  >?X <http://xb.saltlux.com/schema/property/content>
  <http://xb.saltlux.com/resource/kowiki/알라> .
  }
  GROUP BY ?X ORDER BY DESC(?CNT) LIMIT 100
}
```

**Q&A**

# 감사합니다.

---

[ykham@saltlux.com](mailto:ykham@saltlux.com)

