

QA에 사용되는 Knowledge Embedding 응용 및 활용방안

By Donggeun choi



INDEX

- 01 Profile
- 02 What is Knowledge Base?
- 03 What is Knowledge Base Embedding?
- 04 Kinds of Knowledge Base Embedding
- 05 Application

01. Profile



최동근 (Donggeun Choi)

History

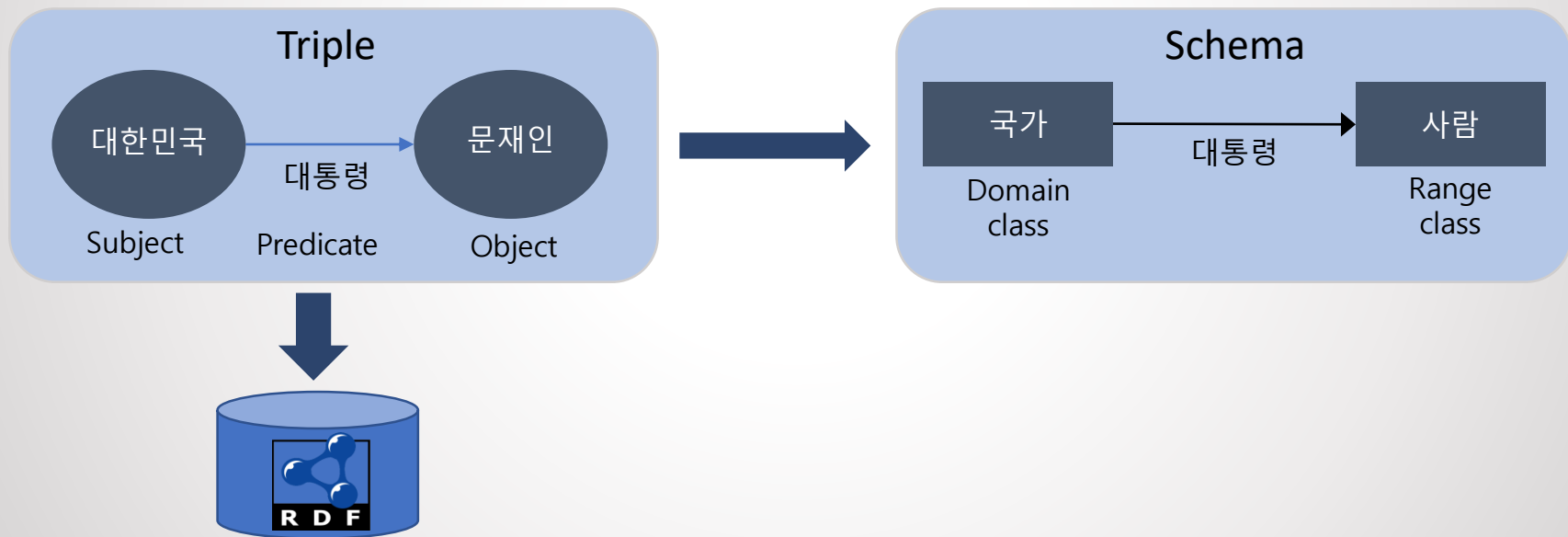
- (現) Saltlux Inc. (2017.08~)
- Master degree in IDS Lab. hanyang Univ. (2015.09~2017.08)
- Bachelor's degree in Hongik Univ. (2011.03~2015.08)

Project

- 자율지능 디지털 동반자 프레임워크 및 응용 연구개발 (2017~)
- WiseKB : 빅데이터 이해 기반 자가학습형 지식베이스 및 추론 기술 개발 (Exobrain. 2013~)
- 서버접근통제 영상정보 보안시스템 (2016)
- Big Data Discovery 솔루션 개발 (2016)
- APM의 성능정보 빅데이터를 이용한 Machine Learning 기반 장애예측 기술 개발 (2015)
- SK하이닉스 Big Data 기반 대용량 실시간 분석 시스템 구축 (2단계)(2015)

02. What is Knowledge Base?

지식 베이스는 전문가 시스템의 구성 요소의 하나로서, **인공지능 에이전트**가 사용될 분야와 관련된 지적 활동과 경험을 통해서 축적한 전문 지식 그리고 문제 해결에 필요한 사실과 규칙 등이 저장되어 있는 **데이터베이스**이다.



01. What is Knowledge Base?

>> 지식베이스를 활용한 인공지능

- 지식을 수집
- 컴퓨터가 처리 및 이해할 수 있는 형태로 구축 및 저장
- 추론, 검색 등과 같은 처리
- 사용자의 의사결정, 질의응답과 같은 지능형 처리를 지원

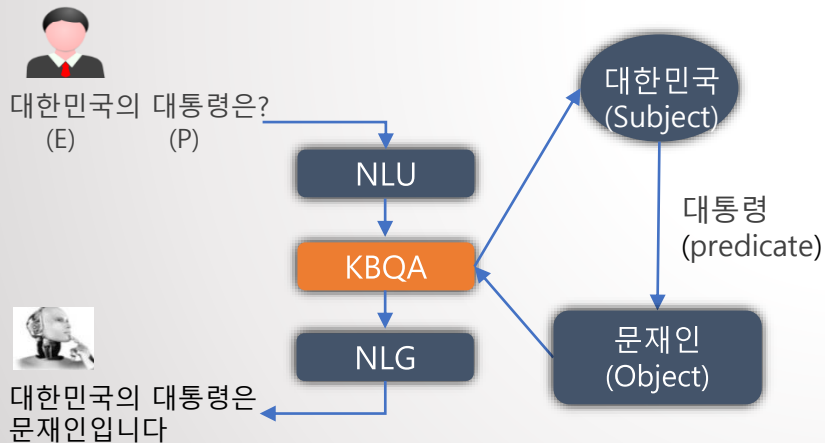


01. What is Knowledge Base?

>> 지식베이스를 활용한 인공지능 → QA(Question Answering) 시스템

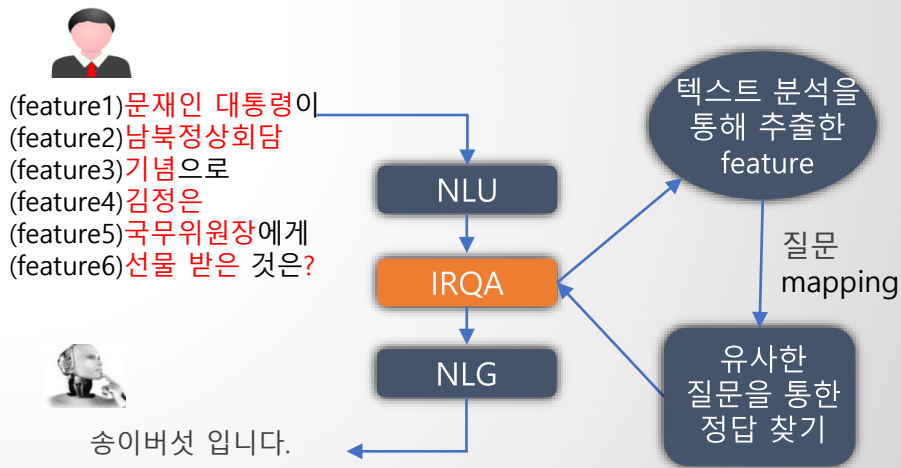
■ KBQA(Knowledge Based Question Answering) 구성

- 지식베이스 기반 질의처리 모듈로써 **구조화가 용이한 도메인에 적합한 QA**
 - 사실(facts)기반 질의 처리와 템플릿 기반 질의 처리
 - 복잡하거나 복합적인 질의(**multi-hop**) 처리



■ IRQA(Information Retrieval Question Answering) 구성

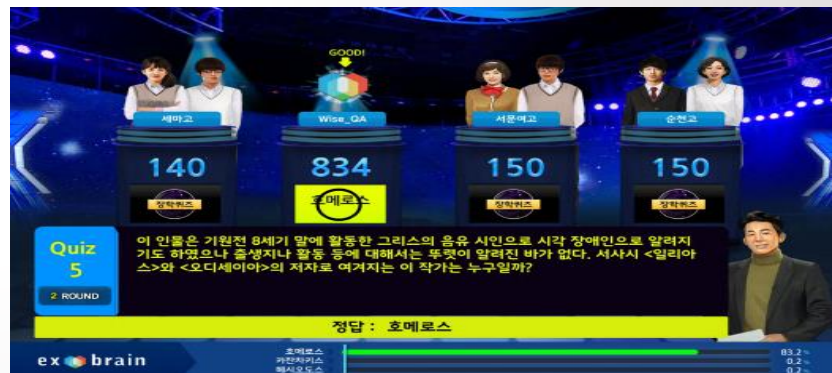
- 질문/답 기반 질의처리 모듈로, **구조화가 용이하지 않은 도메인에 적합한 QA**
 - 텍스트 분석을 통해 **질문과 유사한 질문을 찾아** 랭킹 된 답변 도출



02. What is Knowledge Base?



2011년

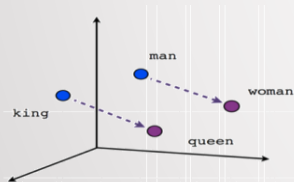


2016년

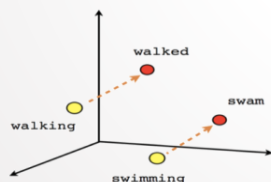
03. What is Knowledge Base Embedding?

Word Embedding

Man	Woman	
1	0	0
0	1	0
0	0	0
⋮	⋮	⋮
0	0	0
0	0	1
0	0	0



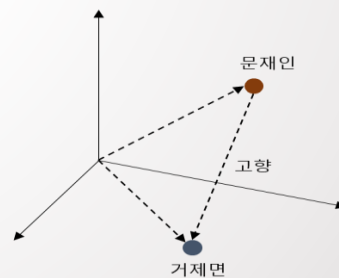
Male-Female



Verb tense

Knowledge Embedding

문재인	거제면	
1	0	0
0	1	0
0	0	0
⋮	⋮	⋮
0	0	0
0	0	1
0	0	0



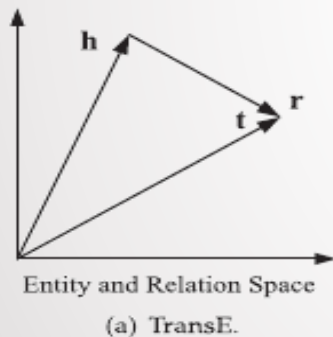
04. Kinds of Knowledge Embedding

- Translational Distance Models
 - 거리 기반의 scoring function을 이용한 모델
 - 한 개의 fact(triple)에서 두 entity간의 특정 관계를 해석한 후에 거리를 측정
- Semantic Matching Models
 - 유사도 기반의 scoring function을 이용한 모델
 - Vector로 표현된 entity와 relation 의미적으로 맞는지 측정

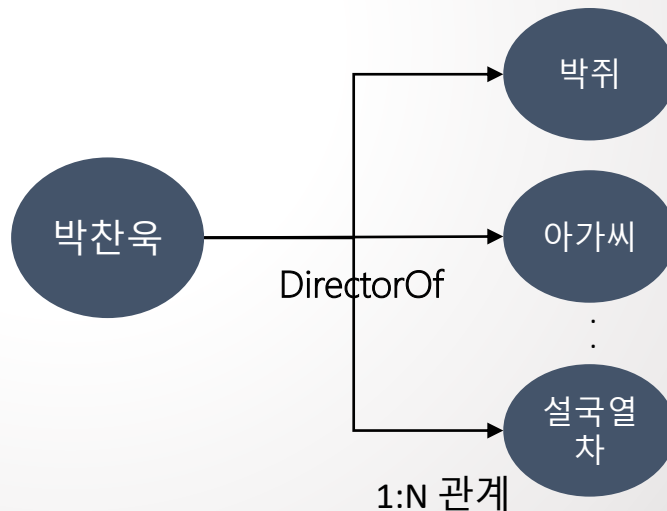
04. Kinds of Knowledge Embedding

>> Translational Distance Models – TransE (2013)

- 기본 가정 : $h + r \approx t$ (h : head entity, r : relation, t : tail entity)
- 가장 단순한 translate based Embedding
- **1:N, N:1, N:N 관계에서 약점**
- Score function : $f(h, t) = -||h + r - t||_2^2$



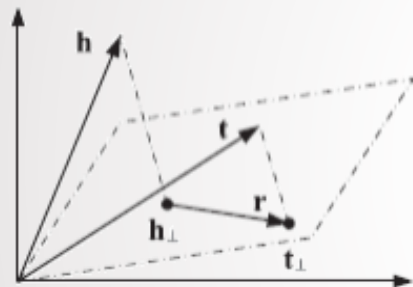
Entity and Relation Space
(a) TransE.



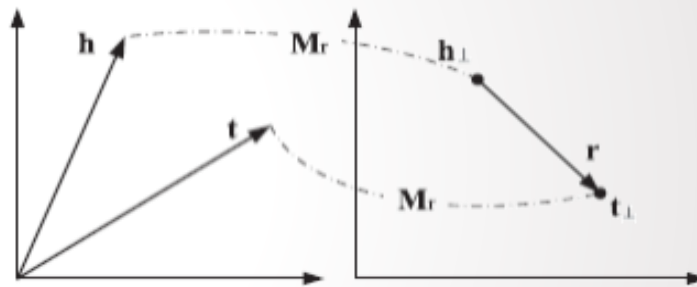
04. Kinds of Knowledge Embedding

>> Translational Distance Models – TransE (2013)

- 기본 가정 : $h + r \approx t$ (h : head entity, r : relation, t : tail entity)



Entity and Relation space
(b) TransH.



Entity Space Relation Space of r
(c) TransR.

- 특정 relation hyperplane으로 projecting
- $h_{\perp} = h - w_r^T h w_r$, $t_{\perp} = t - w_r^T t w_r$
- $f(h, t) = -||h_{\perp} + r - t_{\perp}||_2^2$

- 특정 relation space로 projecting
- $h_{\perp} = M_r^1 h$, $t_{\perp} = M_r^2 t$
- $f(h, t) = -||h_{\perp} + r - t_{\perp}||_2^2$

04. Kinds of Knowledge Embedding

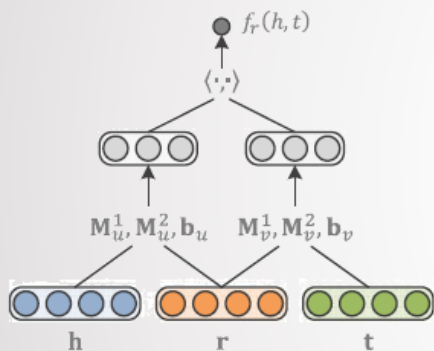
>> Translational Distance Models - 성능

Tasks	Predicting Head(Hits@10)				Predicting Tail(Hits@10)			
관계 종류	1-to-1	1-to-N	N-to-1	N-to-N	1-to-1	1-to-N	N-to-1	N-to-N
TransE	43.7	65.7	18.2	47.2	43.7	19.7	66.7	50
TransH	66.8	87.6	28.7	64.5	65.5	39.8	83.3	67.2
TransR	78.8	89.2	34.1	69.2	79.2	37.4	90.4	72.1

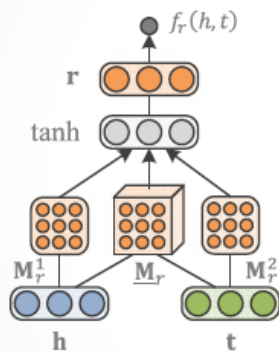
(TopK accuracy : 후보로 추출되는 상위K개의 정확도, 상위K개의 후보 중 정답이 있으면 맞는 것으로 판단)

04. Kinds of Knowledge Embedding

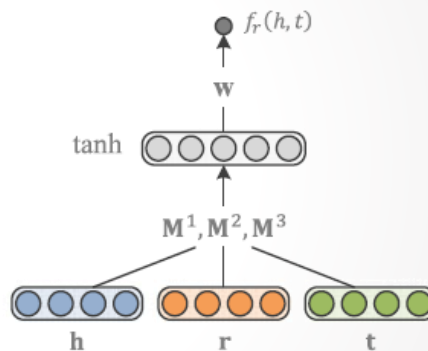
>> Semantic Matching Models – Matching with Neural Networks



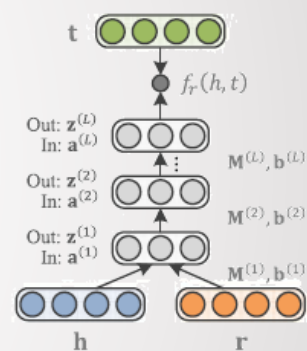
(a) SME.



(b) NTN.



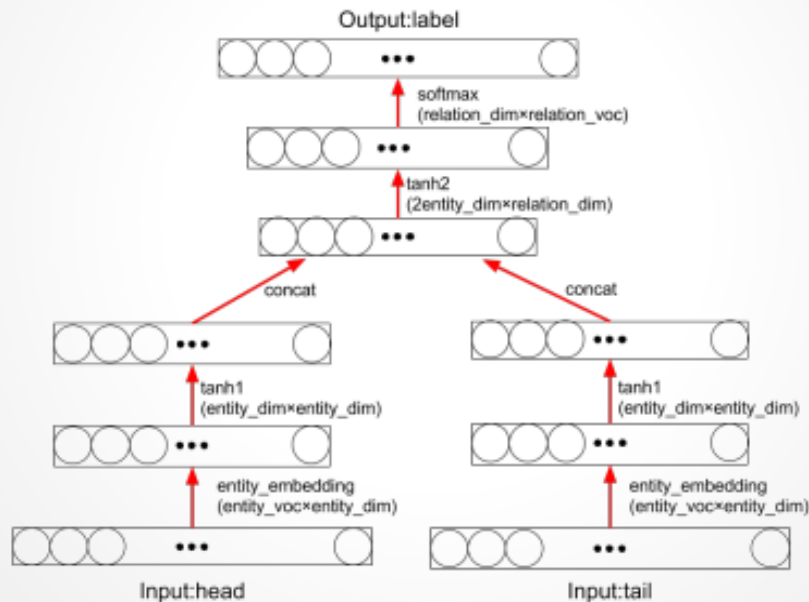
(c) MLP.



(d) NAM.

04. Kinds of Knowledge Embedding

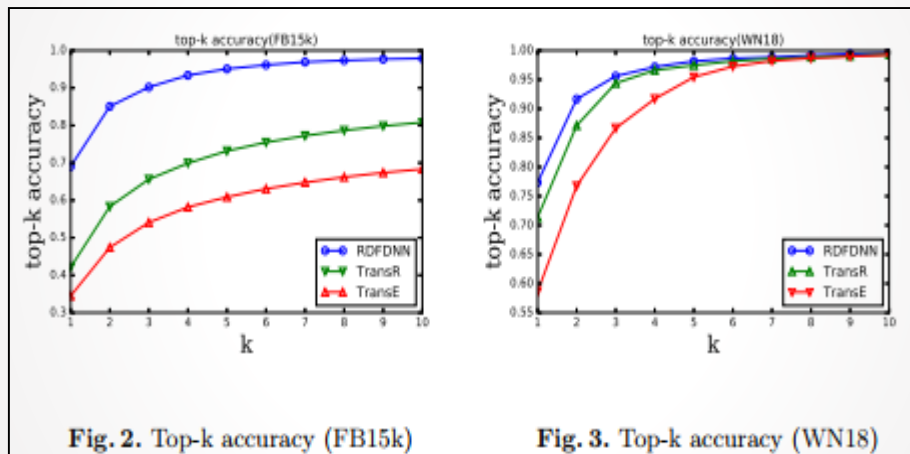
>> Semantic Matching Models – RDFDNN



RDFDNN Structure

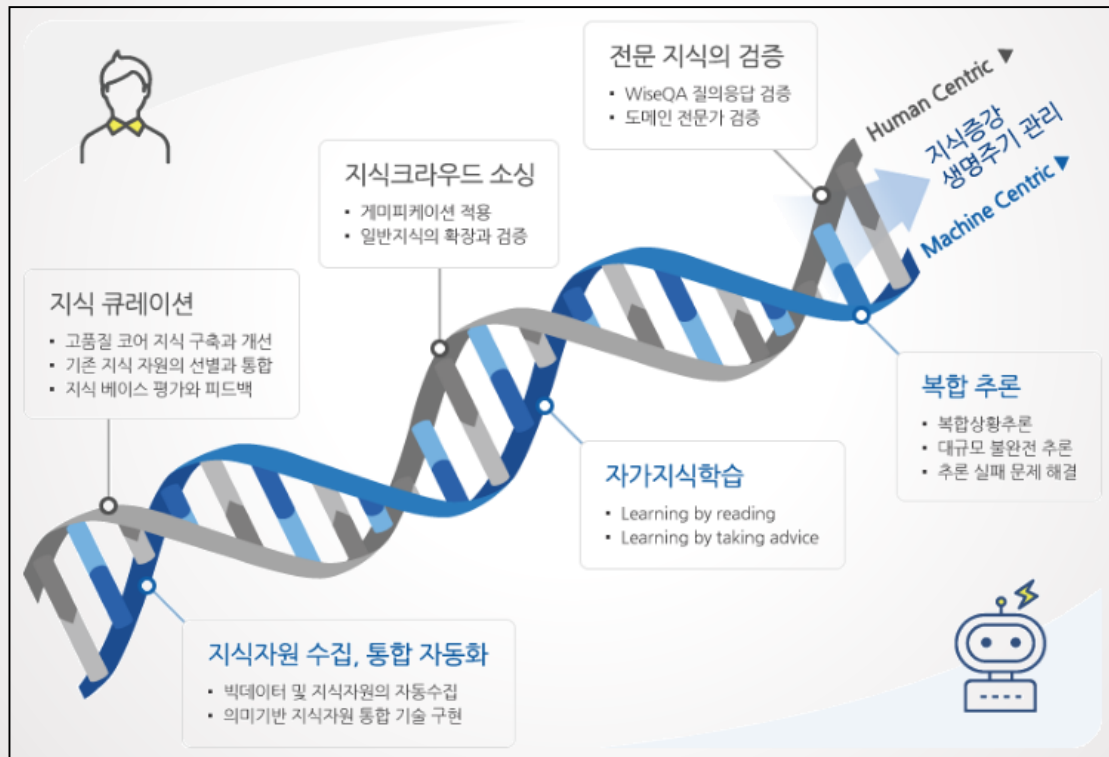
04. Kinds of Knowledge Embedding

>> Semantic Matching Models – RDFDNN



05. Application

>> 지식베이스 정제 및 확장 – 속성 추천 시스템



05. Application

>> 지식베이스 정제 및 확장 – 속성 추천 시스템

문재인 → Subject

위키백과, 우리 모두의 백과사전.

문재인(文在寅, Object1 1953년 1월 24일 ~)은 Object2 대한민국의 제19대 Object3 대통령이다. 본관은 남평(南平),^[2]

경희대학교 재학 시절 학생운동을 이끌며 Object1 박정희 유신 독재에 항거하다가 1975년 서대문구치소에 투옥됐고 대학에서 제적당했다. 출소 후에는 신체 검사도 받지 않은 상태로 군에 강제 징집되었다. Object2 특전사를 제대한 후 복학해 다시 학생운동을 이끌며 Object3 전두환 군부 독재에 항거하다가 1980년 청량리구치소에 투옥됐으나, 옥중에서 사법시험에 합격하면서 조영식 경희대 총장의 신원보증 아래 극적으로 풀려났다. 1982년 사법연수원을 최우수 성적^[3]으로 수료했으나 학생운동 전력으로 판사 임용이 거부되자 부산으로 내려가 Object1 노무현 변호사와 합동법률사무소를 운영하며 인권변호사로 활동했다.

2003년에 참여정부의 초대 Object2 대통령비서실 민정수석을 역임했다. 2004년, 자리에서 물러나 Object3 히말라야 산맥으로 트래킹을 떠났으나, 도중에 Object1 노무현의 탄핵 소추 소식을 듣고 즉시 귀국하여 변호인단의 간사를 맡았다.^[4] 2005년 다시 청와대에 들어가 대통령비서실 시민사회수석, 민정수석, 정무특보를 거쳐 참여정부 마지막 Object2 대통령비서실장을 지냈다.

문재인은 정계입문 제의가 올 때마다 사의를 표했지만, 2012년 제19대 총선에서 부산 사상구에 출마하여 당선되면서 본격적으로 정치활동을 시작했다. 동년 대선 출마를 선언했으며 Object1 손학규, 김두관, 정세균 등과 겨루어 전국 순회경선 13회 전승을 거두며 Object2 민주통합당 제18대 대선 후보로 확정되어 출마하였으나

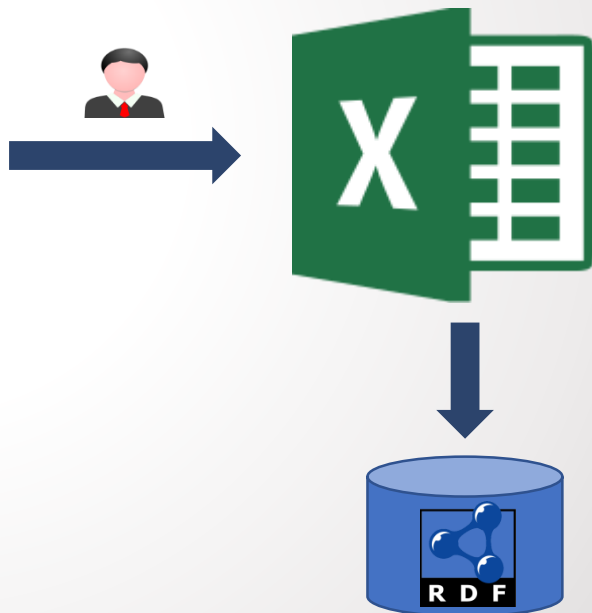
05. Application

>> 지식베이스 정제 및 확장 – 속성 추천 시스템

Subject : 문재인

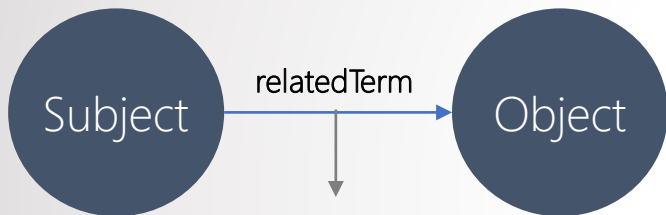
<http://kb.adams.ai/schema/property/relatedTerm>

김상곤	김현철
신하철, Shin Hae Chul	도널드 트럼프, Donald Trump
표창원, Pyo Chang-won	폴 라이언, Paul Ryan
문정수	안철수, Ahn Cheol-soo
박영선, Park Young-sun	대한민국, South Korea
미국, United States of America	일본, Japan
중국, China	트위터, Twitter
대통령, president	백두산, Baekdu Mountain
과천시, Gwacheon	박근혜, Park Geun-hye
민주당, Democratic Party	서울특별시, Seoul
조선민주주의인민공화국, North Korea	아베 신조, Shinzō Abe
자유한국당, Liberty Korea Party	노무현, Roh Moo-hyun
박정희, Park Chung-hee	시진핑, Xi Jinping
안동시, Andong	싱가포르, Singapore
네팔, Nepal	종로구, Jongno District
워싱턴 D.C., Washington, D.C.	한반도, Korean Peninsula
한국전쟁, Korean War	대한민국 제19대 국회의원 선거, South Korean legislative election
대한민국 제20대 국회의원 선거, South Korean legislative election	국민의당
민주당, Democratic Party	더불어민주당, New Politics Alliance for Democracy
정의당, Progressive Justice Party	통합진보당, Unified Progressive Party
박정호, Park Chong-ho	기타



05. Application

>> 지식베이스 정제 및 확장 – 속성 추천 시스템



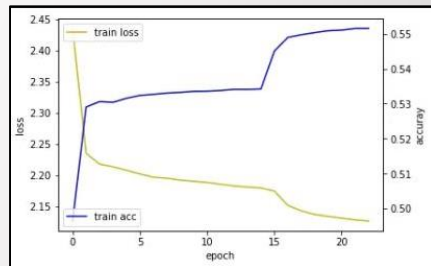
어떤 관계?

- relatedTerm과 연결된 Subject와 Object Entity 추출
- 추출한 Entity들이 가지고 있는 triple을 학습 데이터로 사용
- Entity 수 : 132만개, Triple 수 : 600만개, relation 수 : 748개

```
In [5]: model.fit([s_train,o_train],p_train,epochs=1)
```

Epoch 1/1

6070446/6070446 [=====] - 14950s 2ms/step - loss: 2.1252 - acc: 0.5511



05. Application

>> 지식베이스 정제 및 확장 - 속성 추천 시스템

subject_label	object_label	top1 후보	top2 후보	top3 후보	top4 후보	top5 후보	check	상
# .it	인터넷	topic	0.888 field	0.0142 type	0.012 sp_touches	0.011 race	0.0106	1
# .it	국가 코드 최상위 도메인	field	0.612 topic	0.0586 previous	0.023 work	0.019 bornIn	0.0187	2
# ?	경구개 접근음	bornIn	0.273 locatedIn	0.1723 belongsTo	0.11 diedIn	0.043 sp_touches	0.0375	3
# EXPTIME	복잡도 종류	field	0.847 topic	0.0526 work	0.012 prize	0.01 genre	0.0075	1
# 가네무라 사토루	일본	nationality	0.308 bornIn	0.2329 nation	0.19 belongsTo	0.063 locatedIn	0.0569	1
# 가네무라 사토루	사이타마 세이부 라이온스	employer	0.485 belongsTo	0.283 memberOf	0.116 education	0.025 opponent	0.0109	3
# 가네무라 사토루	호카이도 닛햄 파이터스	belongsTo	0.384 employer	0.2631 memberOf	0.189 position	0.043 opponent	0.0392	3
# 가네무라 사토루	야구	topic	0.464 sport	0.3209 field	0.093 participate	0.033 genre	0.0255	2
# 겨울연가	한국방송공사 유희라마	participant	0.124 locatedIn	0.0532 character	0.051 field	0.048 performer	0.0451	4
# 공위	국화과	topic	0.472 broader	0.1727 job	0.171 type	0.028 field	0.0248	2
# 공위	식물	broadier	0.67 topic	0.1477 type	0.076 job	0.058 party	0.0335	3
# 권오용	고려대학교	employer	0.347 education	0.1447 belongsTo	0.125 manager	0.105 channel	0.0991	2
# 권오용	영주시	locatedIn	0.558 bornIn	0.1221 belongsTo	0.069 mainLocation	0.041 education	0.0308	2
# 김건우	LG 트윈스	belongsTo	0.529 employer	0.2132 memberOf	0.18 opponent	0.025 party	0.0076	3
# 김구	이계인	participant	0.4 performer	0.1964 supportingPerformer	0.124 mainPerformer	0.097 productionCrew	0.0449	2
# 김구	사극	genre	0.709 performer	0.034 participant	0.031 sport	0.019 previous	0.0193	1
# 뉴원플랜드 메모리얼 대학교	세인트존스	locatedIn	0.077 field	0.0592 play	0.051 previous	0.049 belongsTo	0.0452	1
# 대구불교진흥원	장미	symbol	0.906 flower	0.0347 bloodtype	0.012 activity	0.009 participate	0.0061	2
# 대구불교진흥원	은행나무	symbol	0.978 flower	0.0074 degree	0.004 win	0.003 activity	0.0013	1
# 덕벌초등학교	은행나무	symbol	0.978 flower	0.0073 degree	0.004 win	0.003 activity	0.0013	1
# 덕벌초등학교	목련	symbol	0.975 flower	0.0104 degree	0.003 win	0.002 activity	0.0016	2
# 동두천중앙고등학교	동두천시	locatedIn	0.348 bornIn	0.2261 belongsTo	0.061 mainLocation	0.058 sp_touches	0.0477	1
# 리틀리	귀네스 말트로	performer	0.236 participant	0.1439 mainPerformer	0.136 position	0.124 supportingPerformer	0.0711	3
# 마르코 폴로	베네치아 공화국	bornIn	0.269 nationality	0.1703 diedIn	0.129 race	0.079 combatant	0.0707	2
# 마르코 폴로	베네치아	bornIn	0.505 diedIn	0.2528 buriedIn	0.055 livedIn	0.052 happenedIn	0.0303	1,3
# 역전포항	대한민국의 국가어항	locatedIn	0.077 field	0.0592 play	0.051 previous	0.049 belongsTo	0.0452	2
# 역전포항	하일먼	locatedIn	0.077 field	0.0591 play	0.051 previous	0.048 belongsTo	0.0452	1
# 미야자키 미호	도쿄 도	bornIn	0.563 locatedIn	0.1146 diedIn	0.063 belongsTo	0.059 livedIn	0.028	1
# 미야자키 미호	눈물 서프라이즈!	play	0.213 work	0.1522 field	0.058 topic	0.049 bornIn	0.0487	2
# 미야자키 미호	입술에 Be My Baby	play	0.219 belongsTo	0.0875 bornIn	0.074 work	0.067 employer	0.0545	4
# 미야자키 미호	닛토엔젤	belongsTo	0.182 play	0.1526 employer	0.145 memberOf	0.064 education	0.0569	4
# 박성준	통영시	locatedIn	0.396 bornIn	0.1977 mainLocation	0.079 belongsTo	0.074 sp_touches	0.0352	2
# 보리스 파스테르나크	작가	job	0.918 topic	0.0322 genre	0.024 broader	0.024 participate	0.0017	1
# 보리스 파스테르나크	국립 모스크바 대학교	education	0.245 employer	0.1647 belongsTo	0.117 manager	0.08 waypointOf	0.0547	1
# 보리스 파스테르나크	노벨 문학상	field	0.312 language	0.221 awarded	0.184 prize	0.101 operatingSystem	0.0302	4

05. Application

>> 지식베이스 정제 및 확장 – 속성 추천 시스템

프로퍼티 추천								
평가 데이터 업로드(엑셀)		업로드 방식 다운로드						
번호	인스턴스#1	인스턴스#2	추천#1	추천#2	추천#3	추천#4	추천#5	추천#6
1	adr:0000608360(손흥민)	adr:000099551(마우리시오 포체티노)	adp:employer	adp:belongsTo	adp:memberOf	adp:participant	adp:artist	adp:composedOf
2	adr:0000251657(애플)	adr:0000523688(삼성전자)	adp:belongsTo	adp:employer	adp:composedOf	adp:memberOf	adp:opponent	adp:employ
3	adr:0000079756(구글)	adr:0000373738(소프트웨어)	adp:topic	adp:sp_touches	adp:job	adp:field	adp:type	adp:border
4	adr:0000079756(구글)	adr:0000104192(안드로이드)	adp:awarded	adp:ideology	adp:operatingSystem	adp:state	adp:sport	adp:degree
5	adr:0000339821(미녀와 야수)	adr:0000311716(월트 디즈니 컴퍼니)	adp:producer	adp:distributor	adp:participant	adp:productionCompany	adp:editor	adp:productionCrew
6	adr:0000061019(이선희)	adr:0000259471(조용필)	adp:colleague	adp:friend	adp:relative	adp:employer	adp:contribute	adp:opponent
7	adr:0000557137(김경수)	adr:0000552525(김해 김씨)	adp:belongsTo	adp:employer	adp:locatedIn	adp:manager	adp:education	adp:memberOf
8	adr:0000327173(문재인)	adr:0032276200(2018년 남북정상회담)	adp:work	adp:awarded	adp:majorWork	adp:prize	adp:debutWork	adp:write
9	adr:0000450792(유재석)	adr:0000587173(수유중학교)	adp:play	adp:belongsTo	adp:next	adp:bornIn	adp:previous	adp:education
10	adr:0011457761(트와이스)	adr:000006082(K-pop)	adp:participate	adp:genre	adp:activity	adp:win	adp:diedBy	adp:sport
11	adr:0011457761(트와이스)	adr:0000534748(결 그룹)	adp:field	adp:prize	adp:work	adp:perform	adp:appear	adp:awarded
12	adr:0000147702(류현진)	adr:0000305944(메이저 리그 베이스볼)	adp:belongsTo	adp:memberOf	adp:employer	adp:composedOf	adp:participate	adp:party
13	adr:0000147702(류현진)	adr:0000618735(투수)	adp:job	adp:topic	adp:participate	adp:broader	adp:activity	adp:genre
14	adr:0000219349(대한민국)	adr:0000179924(남산)	adp:partner	adp:famousPlace	adp:majorAirport	adp:sp_touches	adp:celebrity	adp:famousArtifact
15	adr:0000219349(대한민국)	adr:0000620664(덕수궁)	adp:partner	adp:famousPlace	adp:majorAirport	adp:sp_touches	adp:celebrity	adp:famousArtifact
16	adr:0000286964(중국)	adr:0000316399(만리장성)	adp:celebrity	adp:participant	adp:famousPlace	adp:productionCrew	adp:majorAirport	adp:sp_touches
17	adr:0011820697(딥 러닝)	adr:0000160014(기계 학습)	adp:belongsTo	adp:employer	adp:memberOf	adp:education	adp:locatedIn	adp:composedOf

05. Application

> > Triple Classification

- 지식베이스 내의 Triple(fact)가 True인지 False인지 분류

> > Entity ranking

- QA에서 NLU 결과인 Semantic resource들의 우선 순위 문제 해결 필요
- EX) 지식베이스 내에 '문재인' 의 Entity가 여러 개 존재할 경우 어떤 '문재인' Entity가 우선 순위가 높을지

> > Question Answering

- bAbI Task와 같은 Reasoning QA에 활용

THE END