DS 데이터 어노테이션 문서

이 매뉴얼은 DS 데이터 어노테이션 방법을 설명합니다.

Revision Date: 2017/08/13

SWRC, KAIST Phone. 010-9113-4736

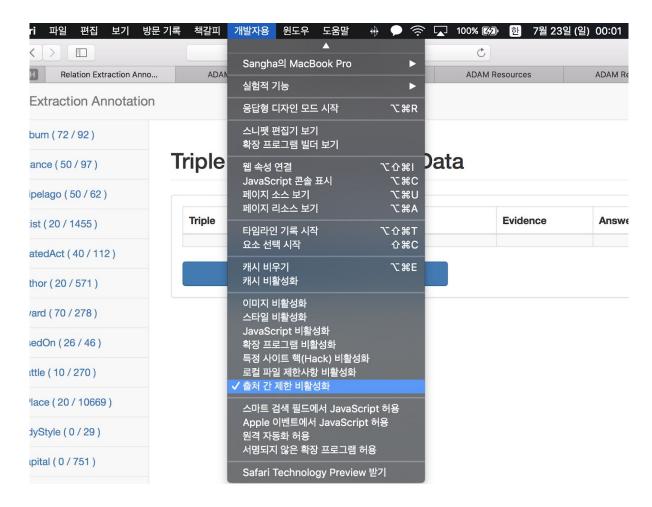
Email. namsangha@world.kaist.ac.kr

버전	내용	작성일	작성자
0.1	초안 작성	2017/08/11	남상하
0.2	프로그램 추가	2017/08/13	남상하

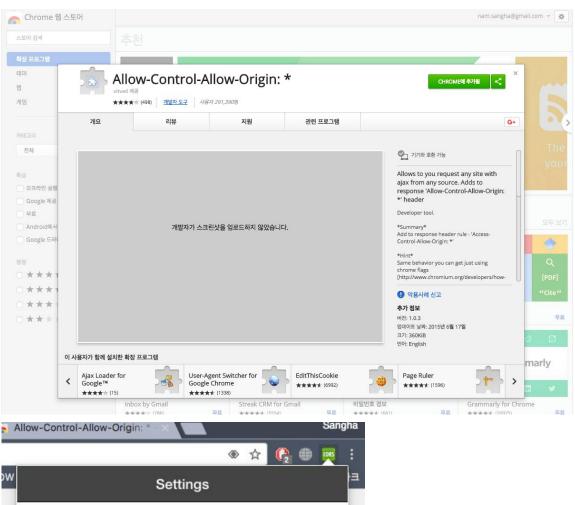
Prerequisite: 브라우저 세팅

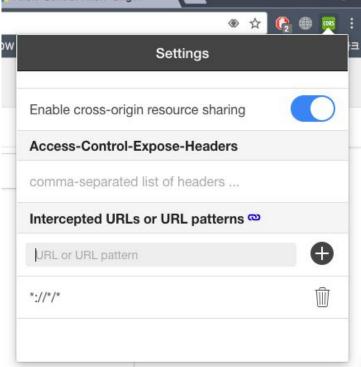
사용 가능 브라우저 : 구글 크롬, 사파리

- **사파리 사용자** : 스크린샷에서 보이는 것 처럼 '개발자용' - '출처 간 제한 비활성화' 체크 되게 한 후 사용하시면 됩니다.



- **크롬 사용자**: 아래 링크의 확장프로그램 설치 후 사용하시면 됩니다.
 - https://chrome.google.com/webstore/detail/allow-control-allow-origi/nlfbmbojpeacfghkpbjhddihlkkiljbi?hl=ko





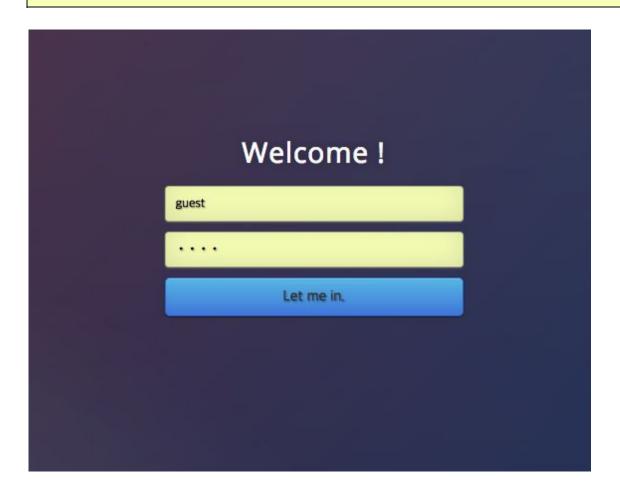
어노테이션 페이지 접근 및 로그인

사이트 주소 : http://wisekb.kaist.ac.kr/dsa/

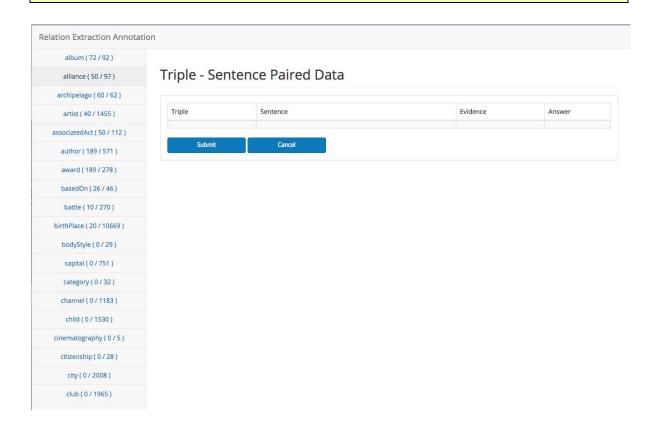
로그인은 본 센터로부터 허가받은 사람에 한해 ID/PW를 발급해드립니다. 본인이 신청한 ID/PW로 로그인 하시기 바랍니다.

* 다른 사람의 ID로 로그인 할 경우에 그 사람의 작업양으로 계산되니 계정 정보 유출에 주의하세요!

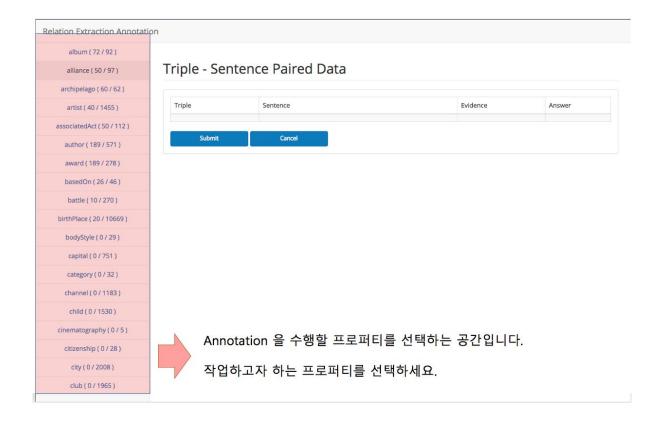
임시 ID/PW : guest/pass



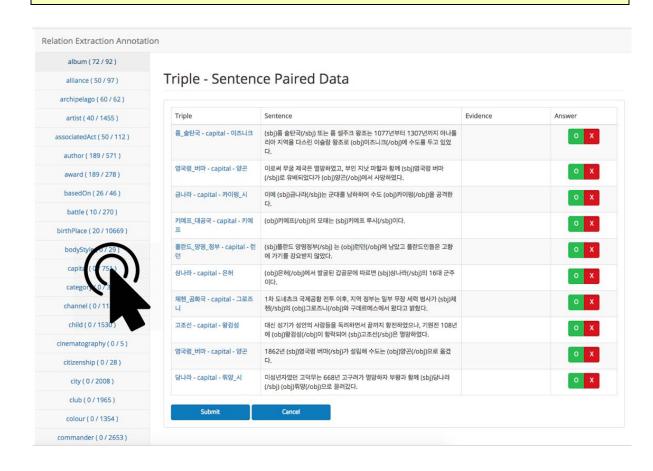
로그인 후, Annotation page 로 잘 들어온 모습입니다. 각각의 기능은 아래 스크린샷과 설명을 참조하세요.



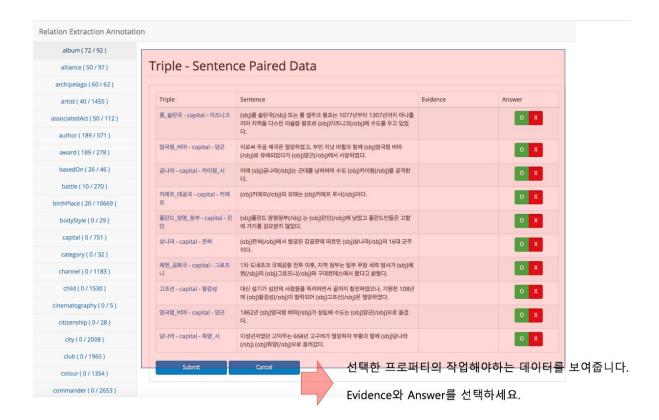
왼쪽에 위치한 서술어 리스트 중에서 어노테이션 작업을 수행할 서술어 하나를 클릭하세요.



작업할 서술어를 선택하면 해당 서술어를 기준으로 수집된 데이터를 10개 보여줍니다. Capital 이라는 프로퍼티를 선택한 예시 화면입니다.



Triple - Sentence Paired Data 아래 부분에 나온 10개의 데이터에 대한 어노테이션 작업을 수행하시면 됩니다.



개별 데이터에 대한 작업 방법은 다음과 같습니다.

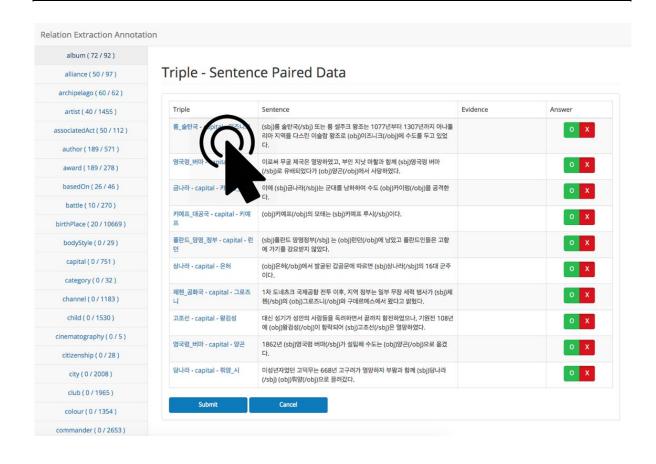
- 1. 문장과 트리플을 읽고, 서로 동일한 의미인지 아닌지에 대해 판단한다.
- 2. 만약 동일하다고 판단된다면,
 - a. 문장에서 나타난 확실한 단서에 대해 드래그를 한다.
 - b. 그 다음 Answer 의 O 버튼을 누르면 자동으로 Evidence 칸이 채워진다.
- 3. 만약 동일하지 않다고 판단되면.
 - a. Answer 의 X 버튼을 누른다.



실수로 잘못 태깅하였을 경우에는 가볍게 X 버튼을 눌러주세요. 자동으로 Evidence 가사라집니다.

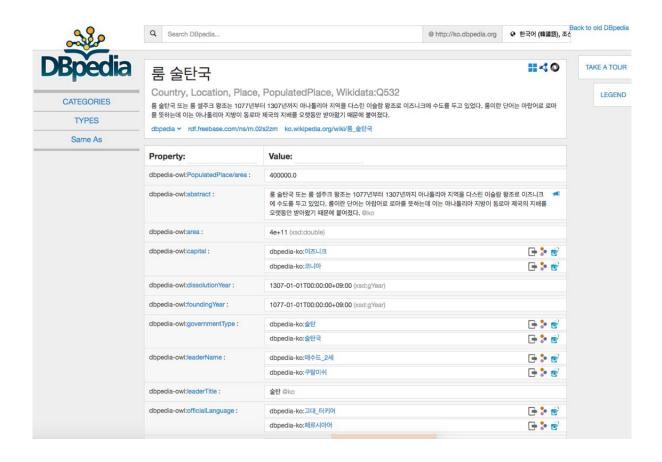
영국령_버마 - capital - 양곤	이로써 무굴 제국은 멸망하였고, 부인 지낫 마할과 함께 (sbj)영국령 버마 (/sbj)로 유배되었다가 (obj)양곤(/obj) <mark>에서 사망하였다.</mark>	에서 사망하였다.	O X
영국령_버마 - <u>capital</u> - 양곤	이로써 무굴 제국은 멸망하였고, 부인 지낫 마할과 함께 (sbj)영국령 버마 (/sbj)로 유배되었다가 (obj)양곤(/obj)에서 사망하였다.		O X
			Click

트리플의 의미가 궁금하다면, 트리플의 각 요소들을 클릭하세요. 각 요소들에 대한 Linked page 로 연결해줍니다.



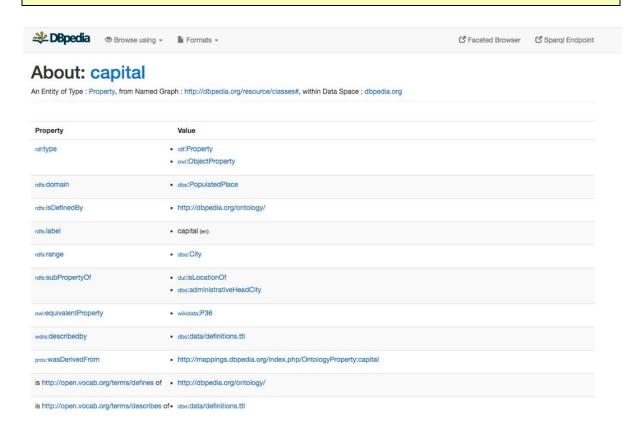
'룸_술탄국' subject 를 클릭했을 시 화면

* 개체 각각에 기술된 정의문을 살펴보시면 이해하기 편합니다.



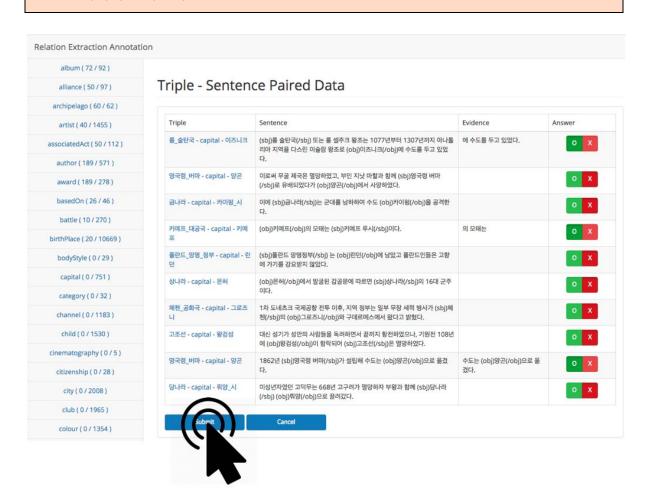
'capital' 서술어를 클릭했을 시 화면

* 서술어의 의미는 작업 시작 전 교육에서 알려드릴 예정입니다. 그래도 작업 중 의미를 잘 모르겠다면 1페이지에 기술된 연락처로 연락주세요.



This content was extracted from Wikipedia and is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported License

어노테이션을 모두 완료하였다면 최종적으로 Submit 버튼을 클릭하세요. 여러분이 작업한 소중한 데이터가 서버에 업로드됩니다.



다른 서술어도 마찬가지 동일한 방법으로 작업하시면 됩니다. 아래 예시는 'child' 서술어에 대한 예시

alliance (50 / 97) archipelago (60 / 62) artist (40 / 1455) associatedAct (50 / 112) author (189 / 571) award (189 / 278) basedOn (26 / 46) battle (10 / 270) birthPlace (20 / 10669) bodyStyle (0/29) capital (0/751) category (0/32) channel (0/1183) cinematography (0/5) citizenship (0/28) city (0/2008) club (0 / 1965) colour (0 / 1354) commander (0/2653)

Triple - Sentence Paired Data

Triple	Sentence	Evidence	Answer
손권 - child - 손화	(obj)순화(/obj)와 그의 동생 손패가 (sbj)순권(/sbj)의 후계를 둘러싸고 다툼 을 벌이자, 적자와 서자의 구분을 분명히 할 것을 주장하며 손화를 지지하였다.		o x
손권 - child - 손량	손침은 (obj)순량(/obj)을 내쫓고 폐제의 형이며 (sbj)손권(/sbj)의 여섯째 아들인 손휴를 황제로 용립하니 곧 오 경제다.		o x
신사임당 - child - 이이	(obj)율곡 이이(/obj)는 (sbj)신사임당(/sbj)의 셋째 아들이다.	의 셋째 아들이다.	o x
김용건 - child - 차현우	본명은 김성훈으로 배우 (sbj)김용건(/sbj)의 아들이며, 동생 (obj)차현우 (/obj)도 배우로 활동하고 있다.	의 아들이며, 동생 (obj)차현우 (/obj)도 배우로 활동하고 있다.	o x
고려_혜종 - child - 경화궁부인	왕규에 의해 왕위를 노린다는 고변을 들게 된 (sbj)혜종(/sbj)은 그를 시험하였고, 이어 944년 의화왕후 소생인 자신의 적녀 (obj)경화궁부인 임씨(/obj)를 후궁으로 주었다.	자신의 적녀	o x
신사임당 - child - 이이	(obj)율곡 이이(/obj)는 1536년 강원도 강룡부 족현동에 있는 의가인 오족현 에서, 덕수 이씨 통덕량 사헌부감찰 이원수와 평산 신씨 (sbj)신사임당(/sbj)의 셋째 아들로 태어났다.	의 셋째 아들로 태어났다.	o x
윤웅렬 - child - 윤치호	종조부는 (sbj)윤웅렬(/sbj)이고, 그의 5촌 당숙은 구한 말의 개화파이자 일제 하 대표적 친일파였던 (obj)윤치호(/obj), 군의관 윤치왕이었다.		o x
조선_세종 - child - 조선_문종	그가 죽자 (sbj)세종(/sbj)은 조회와 저자를 일주일간 정지하고, 조문하고 부의 를 내렸으며 세자인 (obj)조선 문종(/obj)도 치제하였다.	세자인 (obj)조선 문종(/obj)도	o x
조선_세종 - child - 임영대군	신승선의 아들로 어머니 중모현주는 (obj)임영대군(/obj)의 딸이자 (sbj)세종 대왕(/sbj)의 손녀딸이었다.		o x
조선_세종 - child - 임영대군	(obj)임영대군(/obj) 이구는 1420년 1월 6일 (sbj)세종대왕(/sbj)과 심온의 딸 소헌왕후 심씨의 넷째 아들이며 문종, 세조의 친동생이다.	과 심온의 딸 소헌왕후 심씨의 넷 째 아들이며	o x

- 1. 처음 시작하는 작업은 난이도가 가장 낮은 것부터 시작할 예정입니다. 위 예시들은 난이도 중 정도에 해당합니다.
- 2. 작업 시작 전 트리플에 대한 개념과 해석하는 방법 및 어노테이션 작업에 대한 핵심적인 교육이 있을 예정입니다.

그러니 너무 걱정하지 마세요!