# 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

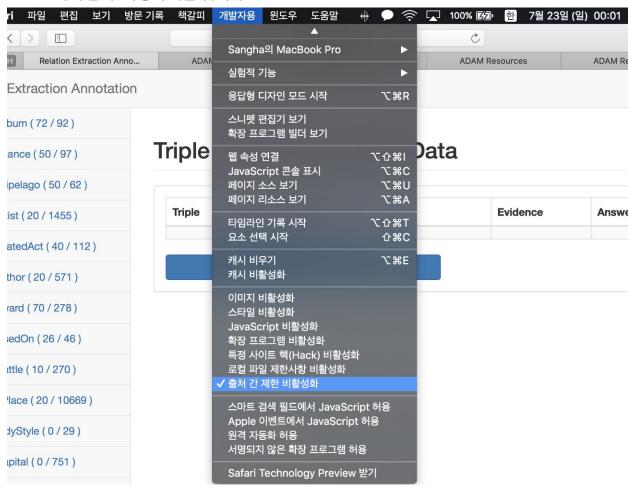
#### 문서 이력

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

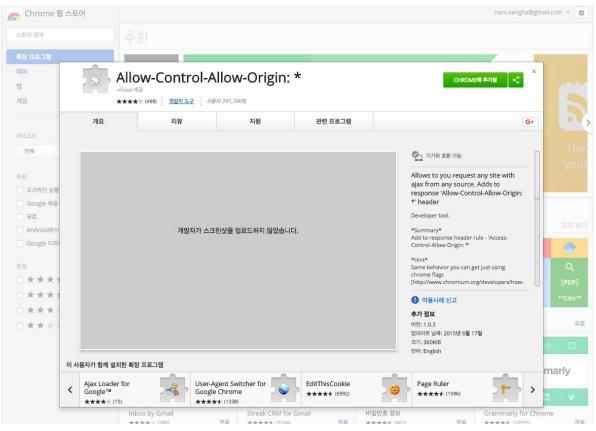
### Prerequisite: 브라우저 세팅

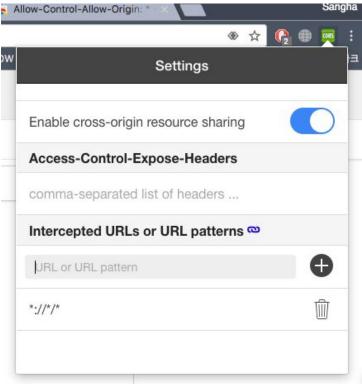
#### 사용 가능 브라우저 : 구글 크롬, 사파리

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



- **크롬 사용자** : 아래 링크의 확장프로그램 설치 후 사용하시면 됩니다.
  - <a href="https://chrome.google.com/webstore/detail/allow-control-allow-origi/nlfbmbojpeacfghk">https://chrome.google.com/webstore/detail/allow-control-allow-origi/nlfbmbojpeacfghk</a> <a href="pbjhddihlkkiljbi?hl=ko">pbjhddihlkkiljbi?hl=ko</a>





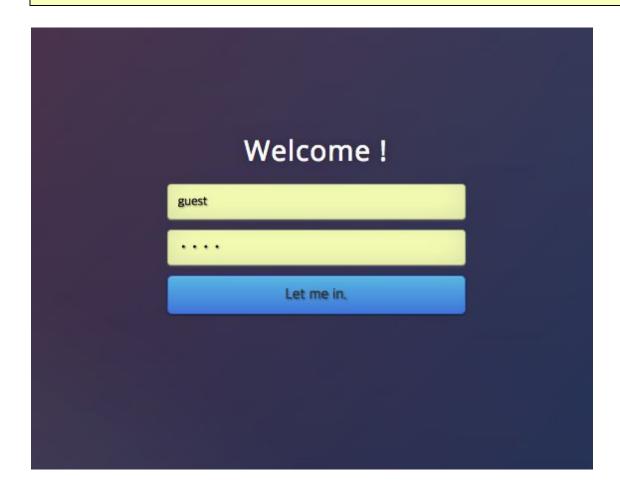
### 어노테이션 페이지 접근 및 로그인

사이트 주소 : <u>http://wisekb.kaist.ac.kr/dsa/</u>

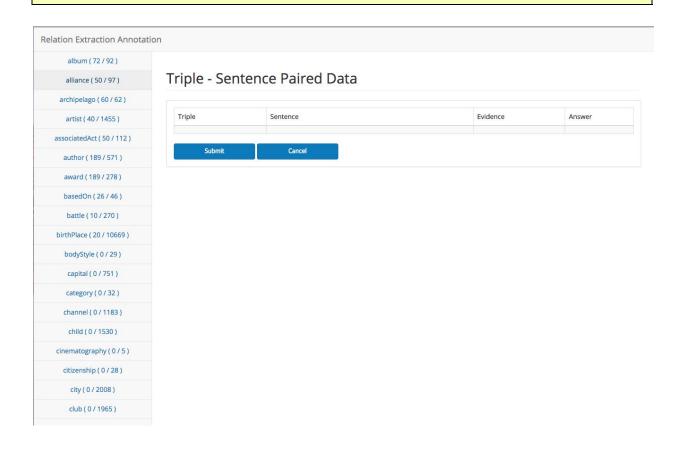
로그인은 본 센터로부터 허가받은 사람에 한해 ID/PW를 발급해드립니다. 본인이 신청한 ID/PW로 로그인 하시기 바랍니다.

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

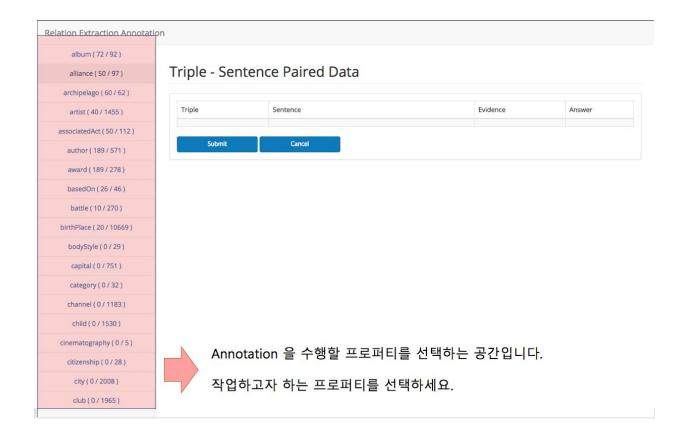
임시 ID/PW : guest/pass



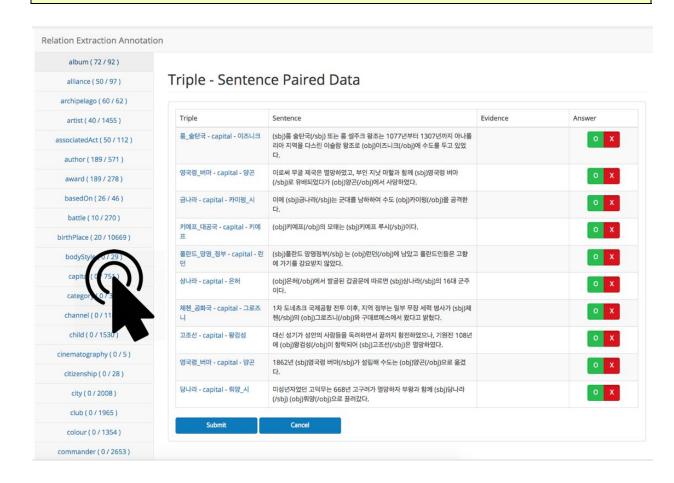
로그인 후, Annotation page 로 잘 들어온 모습입니다. 각각의 기능은 아래 스크린샷과 설명을 참조하세요.



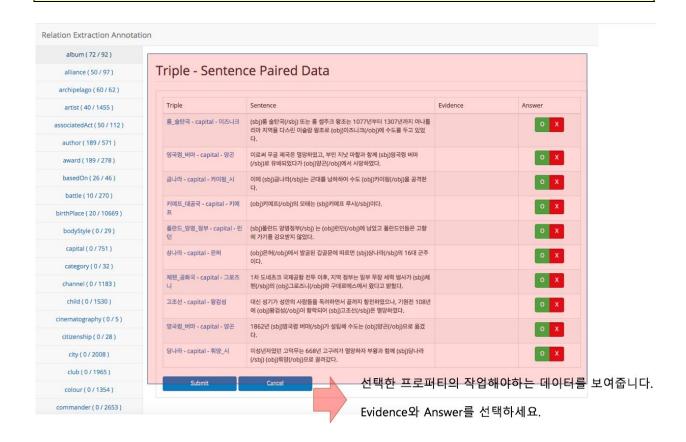
#### 왼쪽에 위치한 서술어 리스트 중에서 어노테이션 작업을 수행할 서술어 하나를 클릭하세요.



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.

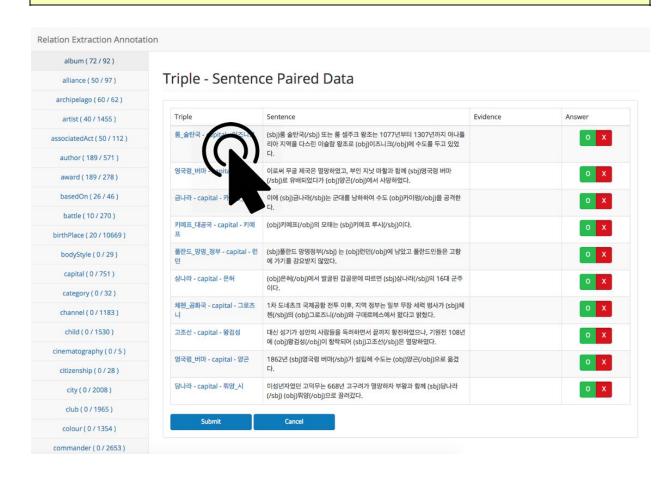
- 1. 문장과 트리플을 읽고, 서로 동일한 의미인지 아닌지에 대해 판단한다.
- 2. 만약 동일하다고 판단된다면,
  - a. 문장에서 나타난 확실한 단서에 대해 드래그를 한다.
  - b. 그 다음 Answer 의 O 버튼을 누르면 자동으로 Evidence 칸이 채워진다.
- 3. 만약 동일하지 않다고 판단되면,
  - a. Answer 의 X 버튼을 누른다.

Triple	Sentence	Evidence	Answer
룸_술탄국 - capital - 이즈니크	(sbj)룸 술탄국(/sbj) 또는 룸 셀주크 왕조는 1077년부터 1307년까지 아나톨리아 지역을 다스린 이슬람 왕조로 (obj)이즈니크(/obj)에 수도를 두고 있었다.		0 X
	DRAG		
Triple	Sentence	Evidence	Answer
룸_술탄국 - capital - 이즈니크	(sbj)룸 술탄국(/sbj) 또는 룸 셀주크 왕조는 1077년부터 1307년까지 아나톨 리아 지역을 다스린 이슬람 왕조로 (obj)이즈니크(/obj) <mark>에 수도를 두고 있었다.</mark>	에 수도를 두고 있었다.	
			Click
Triple	Sentence	Evidence	Answer
룸_술탄국 - capital - 이즈니크	(sbj)룸 술탄국(/sbj) 또는 룸 셀주크 왕조는 1077년부터 1307년까지 아나톨리아 지역을 다스린 이슬람 왕조로 (obj)이즈니크(/obj) <mark>에 수도를 두고 있었다.</mark>	에 수도를 두고 있었다.	o x

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

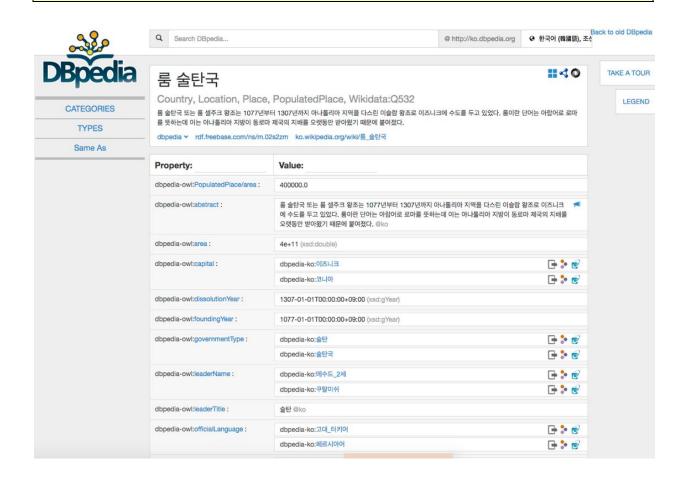
영국령_버마 - capital - 양곤	이로써 무굴 제국은 멸망하였고, 부인 지낫 마할과 함께 (sbj)영국령 버마 (/sbj)로 유배되었다가 (obj)양곤(/obj) <mark>에서 사망하였다.</mark>	에서 사망하였다.	O X
영국령_버마 - <u>capital</u> - 양곤	이로써 무굴 제국은 멸망하였고, 부인 지낫 마할과 함께 (sbj)영국령 버마 (/sbj)로 유배되었다가 (obj)양곤(/obj)에서 사망하였다.		o x
			Click

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



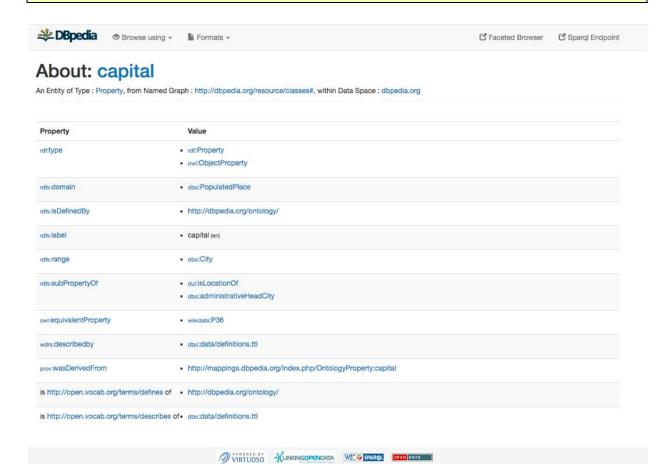
#### '룸\_술탄국' subject 를 클릭했을 시 화면

\* 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.



fia 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.

#### Relation Extraction Annotation album (72 / 92) 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) child (0 / 1530) cinematography (0/5) citizenship (0/28) city (0/2008) club (0/1965) colour (0/1354)

#### Triple - Sentence Paired Data

Triple	Sentence	Evidence	Answer
룸_술탄국 - capital - 이즈니크	(sbj)롬 술탄국(/sbj) 또는 룸 셀주크 왕조는 1077년부터 1307년까지 아나톨 리아 지역을 다스린 이슬람 왕조로 (obj)이즈니크(/obj)에 수도를 두고 있었 다.	에 수도를 두고 있었다.	O X
영국령_버마 - capital - 양곤	이로써 무굴 제국은 멸망하였고, 부인 지낫 마할과 함께 (sbj)영국령 버마 (/sbj)로 유배되었다가 (obj)양곤(/obj)에서 사망하였다.		o x
금나라 - capital - 카이펑_시	이에 (sbj)금나라(/sbj)는 군대를 남하하여 수도 (obj)카이펑(/obj)을 공격한 다.		o x
키예프_대공국 - capital - 키예 프	(obj)키예프(/obj)의 모태는 (sbj)키예프 루시(/sbj)이다.	의 모태는	o x
폴란드_망명_정부 - capital - 런 던	(sbj)폴란드 망명정부(/sbj) 는 (obj)런던(/obj)에 남았고 폴란드인들은 고향 에 가기를 강요받지 않았다.		o x
상나라 - capital - 은허	(obj)은허(/obj)에서 발굴된 갑골문에 따르면 (sbj)상나라(/sbj)의 16대 군주이다.		o x
체첸_공화국 - capital - 그로즈 니	1차 도네츠크 국제공항 전투 이후, 지역 정부는 일부 무장 세력 병사가 (sbj)체 첸(/sbj)의 (obj)그로즈니(/obj)와 구데르메스에서 왔다고 밝혔다.		o x
고조선 - capital - 왕검성	대신 성기가 성안의 사람들을 독려하면서 끝까지 항전하였으나, 기원전 108년 에 (obj)왕검성(/obj)이 함락되어 (sbj)고조선(/sbj)은 멸망하였다.		o x
영국령_버마 - capital - 양곤	1862년 (sbj)영국령 버마(/sbj)가 설립해 수도는 (obj)양곤(/obj)으로 옮겼다.	수도는 (obj)양곤(/obj)으로 옮 겼다.	o x
당나라 - capital - 뤄양_시	미성년자였던 고덕무는 668년 고구려가 멸망하자 부왕과 함께 (sbj)당나라 (/sbj) (obj)뤄양(/obj)으로 끌려갔다.		o x

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



#### 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, 작자와 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.	0 X
고려_혜종 - child - 경화궁부인	왕규에 의해 왕위를 노린다는 고변을 듣게 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 강릉부 죽헌둥에 있는 외가인 으죽헌 에서, 덕수 이씨 통덕랑 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 심은의 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.

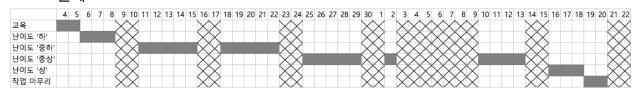
#### 그러니 너무 걱정하지 마세요!

## Program

기간	프로그램 (7명 연구원)	내용
9월 2주차	1. 시맨틱웹 기초 교육 2. 어노테이션 페이지 기능 교육 3. 난이도 '하' 작업 시작 및 완료 (0.6 M/M)	1. 어노테이션 작업에 필요한 Triple의 개념 및 Class, Property (Definition, Domain, Range) 등에 대한 기본 교육2. 연구원별 어노테이션 페이지 계정 부여 및 샘플 데이터로 페이지의 기능에 대한 교육3. 난이도 '하' 데이터에 대한 어노테이션 작업시작 및 완료 (약 3일)
9월 3주차	1. 난이도 ' <b>하</b> ' 작업에 대한 Summary 및 Feedback 2. 난이도 ' <b>중하</b> ' 작업 시작 (3.3 M/M)	1.1 난이도 ' <b>하</b> ' 작업에 대한 Summary : 연구원 각자 어노테이션 작업 시 '작업노트' 작성 및 공유 (작업노트: 시간별 작업 양, 작업시 발생한 의문점, 어려운 문제에 대한 각자의 해결 방법 등) 1.2 난이도 ' <b>하</b> ' 작업에 대한 Feedback : 센터 내 어노테이션 전문가들이 작업한 것과 비교 (약 10%), 평가 후 품질이 아주 나쁜 경우에는 재작업 실시 2. 난이도 ' <b>중하</b> ' 작업 시작 (약 10일)
9월 4주차	1. 난이도 ' <b>중하</b> ' 작업에 대한 중간 Summary 및 Feedback 2. 난이도 ' <b>중하</b> ' 작업 완료	1. 난이도 ' <b>중하</b> ' 작업에 대한 중간 Summary 및 Feedback 2. 난이도 ' <b>중하</b> ' 작업 완료
9월 5주차	1. 난이도 ' <b>중하</b> ' 작업에 대한 전체 Summary 및 Feedback 2. 난이도 ' <b>중상</b> ' 작업 시작 (3.7 M/M)	1. 난이도 ' <b>중하</b> ' 작업에 대한 전체 Summary 및 Feedback 2. 난이도 ' <b>중상</b> ' 작업 시작 (약 11일)
10월 1주차	-추석연휴- (근무일 1일) 1. 난이도 ' <b>중상</b> ' 작업 계속	1. 난이도 ' <b>중상</b> ' 작업 계속
10월 2주차	1. 난이도 ' <b>중상</b> ' 작업에 대한 중간 Summary 및 Feedback 2. 난이도 ' <b>중상</b> ' 작업 완료	1. 난이도 ' <b>중상</b> ' 작업에 대한 중간 Summary 및 Feedback 2. 난이도 ' <b>중상</b> ' 작업 완료
10월 3주차	1. 난이도 ' <b>중상</b> ' 작업에 대한 전체 Summary 및 Feedback 2. 난이도 ' <b>상</b> ' 작업 시작 및 완료 (0.8 M/M) 3. 어노테이션 작업 마무리	1. 난이도 ' <b>중상</b> ' 작업에 대한 전체 Summary 및 Feedback 2. 난이도 ' <b>상</b> ' 작업 시작 및 완료 (약 3일) 3. 작업 우수자 선정 및 포상 지급 (도서상품권 등)

### 근무일자별 스케쥴

#### - 전체



#### - 월별 (9월)

	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
교육						X	X						$\times$	$\times$						$\times$	$\times$						X
난이도 '하'						X	X						X	X						X	X						X
난이도 '중하'						X	X						X	X						X	X						X
난이도 '중상'						X	X						X	X						X	X						X
난이도 '상'						X	X						X	X						X	X						X
작업 마무리						X	X						X	X						X	X						X

#### - 월별 (10월)

