


캡스톤 디자인 I

종합설계 프로젝트

프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼
팀 명	T-SAN(티끌모아 태산)
문서 제목	2차 중간보고서

Version	1.6
Date	2020-05-26

팀원	이정하(팀장)
	박상일
	박지희
	윤여환
	이다은
	장태진
지도교수	최은미 교수님

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26


CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 국민대학교 전자정보통신대학 컴퓨터공학부 및 컴퓨터공학부 개설 교과목 캡스톤 디자인I 수강 학생 중 프로젝트 “집단지성을 이용한 데이터 레이블링 플랫폼”을 수행하는 팀 “T-SAN(티끌모아 태산)”의 팀원들의 자산입니다. 국민대학교 컴퓨터공학부 및 팀 “T-SAN(티끌모아 태산)”의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

문서 정보 / 수정 내역


Filename	2차중간보고서-집단지성을 이용한 데이터 레이블링 플랫폼.doc
원안작성자	이정하, 박상일, 박지희, 윤여환, 이다은, 장태진
수정작업자	이정하, 박상일, 박지희, 윤여환, 이다은, 장태진

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2020-05-18	이정하	1.0	최초 작성	프로젝트 목표 작성
2020-05-19	윤여환	1.1	내용 수정	수정된 연구내용 추가
2020-05-20	이다은	1.2	내용 수정	향후 추진 계획 수정
2020-05-21	박상일	1.3	내용 수정	수행 내용 및 중간 결과 추가
2020-05-23	박지희	1.4	내용 수정	수정사항 추가
2020-05-25	장태진	1.5	내용 수정	향후 계획 추가
2020-05-26	전원	1.6	내용 정리	전체적인 마무리

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

목 차

1	프로젝트 목표	4
2	수행 내용 및 중간 결과	5
2.1	수행 내용	5
2.2	중간 결과	7
3	수정된 연구내용 및 추진 방향	26
3.1	수정사항	26
4	향후 추진계획	28
4.1	향후 계획의 세부 내용	28

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

1 프로젝트 목표

본 프로젝트는 T-SAN이 자체적으로 크롤링하여 모은 데이터를 사용자(레이블링 참여자)들이 레이블링하여 유의미한 정보를 생성하고 플랫폼을 개발하는 것을 목표로 한다.


반응형 웹을 통해 모바일 환경을 함께 지원하는 것을 목표로 한다. React를 통해 프론트엔드를 구현하도록 한다. 보다 빠른 개발 속도를 위해 state 관리 모듈인 mobx와 디자인 프레임 워크인 react-semantic-ui를 적극 사용함으로써 프론트 엔드 개발의 효율을 높일 수 있다.

레이블링 참여자가 레이블링을 통해 포인트를 획득 할 수 있으며, 보상으로 받은 포인트를 추후 제휴를 맺은 다른 기업들의 매장에서 사용이 가능하다. 또한 포인트 관리를 블록체인으로 함으로써 포인트를 투명하게 관리한다.

전반적인 대부분의 기술들은 백엔드에서 각각의 모듈로 작동한다. 회원 정보 관리 및 게시물 업로드 등의 일반적인 기능은 기존 웹사이트 개발 관례에 맞게 개발하고, 특수 기능들은 각각의 모듈이 데이터베이스에 접근하여 처리한다. 웹페이지의 백엔드는 기본 기능 외에 토큰 입금 / 토큰 출금 등의 보상 관련 API들을 처리한다. Django를 통해 서버를 개발하며 GraphQL를 통해 프론트엔드와 통신하는 것을 목표로 한다.

회원DB, 결산 내역 등 데이터가 많지 않거나 중요할 것으로 예상되는 데이터들은 Maria DB에, 크롤링과 라벨링 등으로 방대한 데이터가 저장될 것으로 예상되는 부분은 MongoDB를 통해 데이터를 저장하여 효율적인 DB관리를 하도록 한다.

머신러닝을 이용하여 레이블을 검증하고 레이블링 참여자의 신뢰도를 조정하도록 한다. 정확한 레이블링을 많이 진행한 참여자일 수록 “신뢰도”라는 가중치를 높게 부여한다. 이 신뢰도를 정확한 레이블링 여부를 활용 하는데 사용하여 레이블링 결과의 정확도를 높이는 것을 목표로 한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

2 수행 내용 및 중간 결과

2.1 수행 내용

1) Machine Learning

- 학습을 위한 임시 레이블링
 - 각 데이터의 레이블링 참여자 중 신뢰도가 상위 50%인 참여자들(단, 최소 3인)의 레이블 중 가장 다수가 지정한 레이블을 이미지 데이터의 임시 레이블로 지정하는 과정을 구현한다.
- 이미지 레이블 검증 모델 개발
 - Imagenet의 데이터로 사전학습된 DenseNet201모델의 Convolution 부분과 데이터셋 검증에 사용할 Classifier를 연결하여 임시 레이블링 된 데이터로 Classifier를 K-Fold를 적용하여 적은 데이터셋으로 전이 학습시켜서 데이터셋의 레이블링 검증에 사용하도록 하는 과정 구현한다.
- 텍스트 레이블 검증 모델 개발
 - 한국어 위키의 문장 5M개, 단어 54M개, 한국어 뉴스의 문장 20M개, 단어 270M개로 사전학습 시킨 KOBERT 모델을 가져와서 네이버 영화 리뷰 20만개로 문장의 감정을 분석할 수 있게 꿈 전이학습시킨 모델을 이용하여 텍스트 감정 분석 레이블링 검증에 사용하도록 하는 과정을 구현한다.
- 참여자의 신뢰도 갱신
 - 임시 레이블과 모델 예측 레이블의 일치 여부로 검수 작업을 완료하고 최종 레이블을 확정하고, 참여자들이 레이블링한 값들과 대조하여 참여자들의 레이블 정답 개수로 신뢰도를 갱신하는 과정을 구현한다.

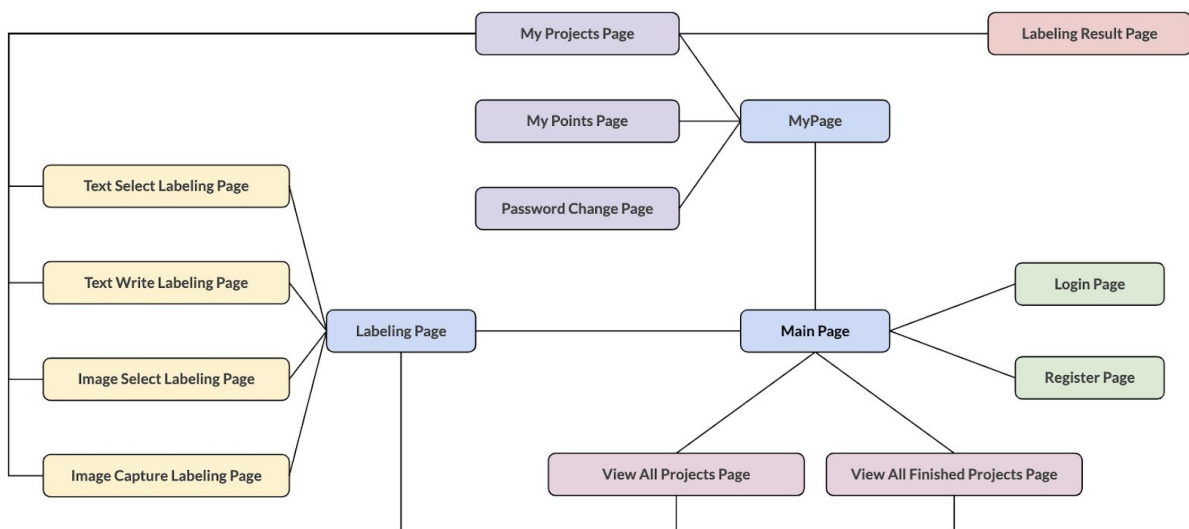
2) Blockchain

- Solidity를 이용한 smart contract 작성
 - 사용자에게 대한 보상 관리를 위해 포인트를 생성, 분배, 제거 등의 기능이 동작할 수 있게 하였다. Solidity 언어를 이용하고 Ethereum ERC20토큰 표준의 인터페이스를 참고하여 Smart Contract를 작성하였고, 이더리움 테스트넷(Ropsten)에 계약 배포를 완료하였다.
- Web3를 이용한 웹 서버와의 연동
 - Web-Backend와 연동을 위해 Web3 라이브러리를 사용했다. Django 서버를 이용하는 이유로, Web3 라이브러리는 Python에서 사용한다. 현재 geth를 이용하여 private network를 구성하여 web3를 이용한 계약 배포 및 transaction 테스트를 완료했으며,

이더리움 테스트넷(Ropsten)에서 테스트 중이다.

3) Front-End

- 기존 설계에 맞추어 1차 중간 평가에 맞추어 전체 웹 페이지의 디자인을 구현했다. 디자인 구현을 맞추고 GraphQL 의 query 와 mutation 을 이용하여 Backend 서버와 API 연동을 하여 기능 구현을 완료하는 단계이다.



[그림 1] 웹 페이지 Routing 구성도

4) Web Back-End

- Django를 프레임워크를 사용하여 웹서버와 DB를 연동하고있다. 레이블링 주제 의뢰, 참여자와 의뢰자 회원 관리, 보상 체계와 관련된 데이터는 MariaDB를 사용한다. Django와 MySQL을 연동하고 mysqlclient 모듈을 통해 MySQL과 MariaDB를 연결하였다. 또한 레이블링 주제에 대한 데이터 셋은 MongoDB로 관리한다. Django와 MongoDB는 pymongo 모듈을 사용하여 연결하였다.
- GraphQL API을 사용하여 프론트엔드에서의 HTTP 요청 횟수와 백엔드에서의 응답 크기를 줄이도록 하였다. 데이터 라벨링 플랫폼의 회원, 의뢰, 보상 관련, 데이터 셋 관련 API에 대해 요청 인자를 이용하여 다양한 요청들에 대해 응답할 수 있게 하였다.
- 보다 안정된 웹서비스를 운영하기 위해 uWSGI와 NGINX Reverse Proxy를 함께 이용하여 배포하였다.

2.2 중간 결과

1) Machine Learning

더미 데이터 프레임을 생성하고 레이블링 한 유저 중 신뢰도 상위 50%들의 다수 레이블을 임시 레이블로 지정하는 과정 구현을 완료하였다..

project_id	data_index		data	label_true	label_user	user_id	user_credibility
0	2.0	1.0	../food-images/pizza/1001116.jpg	pizza	pizza	mfs	0.7593
1	2.0	1.0	../food-images/pizza/1001116.jpg	pizza	pizza	ukw	0.8618
2	2.0	1.0	../food-images/pizza/1001116.jpg	pizza	pizza	nyi	0.9949
3	2.0	1.0	../food-images/pizza/1001116.jpg	pizza	pizza	sbs	0.7007
4	2.0	1.0	../food-images/pizza/1001116.jpg	pizza	pizza	vbq	0.9579
...
34995	2.0	5000.0	../food-images/hot_dog/996310.jpg	hot_dog	hot_dog	nyi	0.9949
34996	2.0	5000.0	../food-images/hot_dog/996310.jpg	hot_dog	hot_dog	sbs	0.7007
34997	2.0	5000.0	../food-images/hot_dog/996310.jpg	hot_dog	hot_dog	vbq	0.9579
34998	2.0	5000.0	../food-images/hot_dog/996310.jpg	hot_dog	hot_dog	lzt	0.8922
34999	2.0	5000.0	../food-images/hot_dog/996310.jpg	hot_dog	hot_dog	zgo	0.6822

[그림 2] 더미 데이터셋 생성

	data_index	data	label_temp	credibility_label_temp
0	1.0	../food-images/pizza/1001116.jpg	pizza	0.835571
1	2.0	../food-images/pizza/1008104.jpg	pizza	0.861000
2	3.0	../food-images/pizza/1008144.jpg	pizza	0.845580
3	4.0	../food-images/pizza/1008844.jpg	pizza	0.835571
4	5.0	../food-images/pizza/1008941.jpg	pizza	0.845580
...
4995	4996.0	../food-images/hot_dog/987144.jpg	hot_dog	0.835571
4996	4997.0	../food-images/hot_dog/987193.jpg	hot_dog	0.835571
4997	4998.0	../food-images/hot_dog/98754.jpg	hot_dog	0.835571
4998	4999.0	../food-images/hot_dog/99314.jpg	hot_dog	0.835571
4999	5000.0	../food-images/hot_dog/996310.jpg	hot_dog	0.835571

5000 rows × 4 columns

[그림 3] 임시 레이블 지정

적은 데이터셋으로 훈련-검증 데이터 셋을 생성을 반복하는 K-fold 기법을 적용하여 임시

레이블로 모델의 Classifier를 학습시키는 과정 구현을 완료하였다.

```

Train on 4000 samples, validate on 1000 samples
Epoch 1/300
4000/4000 [=====] - 7s 2ms/step - loss: 0.0607 - accuracy: 0.9880 - val_loss: 0.0483 - val_accuracy: 0.9950
Epoch 2/300
4000/4000 [=====] - 8s 2ms/step - loss: 0.0546 - accuracy: 0.9923 - val_loss: 0.0508 - val_accuracy: 0.9950
Epoch 3/300
4000/4000 [=====] - 8s 2ms/step - loss: 0.0517 - accuracy: 0.9923 - val_loss: 0.0582 - val_accuracy: 0.9930
Epoch 4/300
4000/4000 [=====] - 7s 2ms/step - loss: 0.0483 - accuracy: 0.9940 - val_loss: 0.0678 - val_accuracy: 0.9840
Epoch 5/300
4000/4000 [=====] - 7s 2ms/step - loss: 0.0464 - accuracy: 0.9950 - val_loss: 0.0681 - val_accuracy: 0.9830
Epoch 6/300
4000/4000 [=====] - 6s 2ms/step - loss: 0.0455 - accuracy: 0.9958 - val_loss: 0.0691 - val_accuracy: 0.9850
Epoch 7/300
4000/4000 [=====] - 7s 2ms/step - loss: 0.0424 - accuracy: 0.9962 - val_loss: 0.0794 - val_accuracy: 0.9790
Epoch 8/300
4000/4000 [=====] - 7s 2ms/step - loss: 0.0412 - accuracy: 0.9958 - val_loss: 0.0807 - val_accuracy: 0.9800
Epoch 9/300
4000/4000 [=====] - 7s 2ms/step - loss: 0.0395 - accuracy: 0.9965 - val_loss: 0.0840 - val_accuracy: 0.9760
Epoch 10/300
4000/4000 [=====] - 7s 2ms/step - loss: 0.0374 - accuracy: 0.9970 - val_loss: 0.0902 - val_accuracy: 0.9730
Epoch 11/300
4000/4000 [=====] - 7s 2ms/step - loss: 0.0369 - accuracy: 0.9970 - val_loss: 0.0907 - val_accuracy: 0.9720

```

[그림 4] 이미지 모델의 학습 과정

임시 레이블과 모델의 예측 레이블이 일치한 데이터에 대하여 레이블을 확정하고 임시 레이블과 모델의 예측이 일치하는 데이터들은 레이블을 확정시키고 해당 데이터의 정답률과 올바르게 레이블한 유저의 아이디를 구해서 각 데이터의 레이블링 정답률과 유저의 현재 신뢰도를 바탕으로 이번 프로젝트에서 유저의 신뢰도를 임시로 계산하는 과정을 구현하였다.

project_id	data_index		data	label_true	label_user	user_id	user_credibility	user_credibility_temp
0	2.0	1.0	../food-images/pizza/1001116.jpg	pizza	pizza	mfs	0.7593	0.7600
1	2.0	1.0	../food-images/pizza/1001116.jpg	pizza	pizza	ukw	0.8618	0.8622
2	2.0	1.0	../food-images/pizza/1001116.jpg	pizza	pizza	nyi	0.9949	0.9949
3	2.0	1.0	../food-images/pizza/1001116.jpg	pizza	pizza	sbs	0.7007	0.7015
4	2.0	1.0	../food-images/pizza/1001116.jpg	pizza	pizza	vbq	0.9579	0.9580
...
34995	2.0	5000.0	../food-images/hot_dog/996310.jpg	hot_dog	hot_dog	nyi	0.9949	0.9949
34996	2.0	5000.0	../food-images/hot_dog/996310.jpg	hot_dog	hot_dog	sbs	0.7007	0.7015
34997	2.0	5000.0	../food-images/hot_dog/996310.jpg	hot_dog	hot_dog	vbq	0.9579	0.9580
34998	2.0	5000.0	../food-images/hot_dog/996310.jpg	hot_dog	hot_dog	lzt	0.8922	0.8925
34999	2.0	5000.0	../food-images/hot_dog/996310.jpg	hot_dog	hot_dog	zgo	0.6822	0.6831

35000 rows × 8 columns

[그림 5] 임시 신뢰도 계산



2차 중간보고서		
프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
팀 명	T-SAN(티폴모아 태산)	
Confidential Restricted	Version 1.6	2020-05-26

	data_index	data	label_temp	credibility_label_temp	label_predicted	label_second
13	14.0	../food-images/pizza/1038357.jpg	pizza	0.845580	steak	pizza
143	144.0	../food-images/pizza/1512514.jpg	pizza	0.835571	steak	pizza
181	182.0	../food-images/pizza/1660415.jpg	pizza	0.861133	sushi	pizza
244	245.0	../food-images/pizza/1934355.jpg	pizza	0.835571	steak	pizza
257	258.0	../food-images/pizza/1973447.jpg	pizza	0.861133	hot_dog	pizza
...
4800	4801.0	../food-images/hot_dog/505584.jpg	hot_dog	0.835571	steak	hot_dog
4812	4813.0	../food-images/hot_dog/527074.jpg	hot_dog	0.835571	sushi	hot_dog
4852	4853.0	../food-images/hot_dog/607287.jpg	hot_dog	0.835571	pizza	hot_dog
4939	4940.0	../food-images/hot_dog/853994.jpg	hot_dog	0.835571	steak	hot_dog
4997	4998.0	../food-images/hot_dog/98754.jpg	hot_dog	0.835571	sushi	hot_dog

248 rows × 6 columns

[그림 6] 2차 레이블링

유저들의 임시 신뢰도를 이용하여 임시 레이블과 모델의 예측 레이블이 일치하지 않았던 데이터들의 레이블링을 진행한다. 현재는 이를 바탕으로 이미지 데이터의 최종 레이블을 지정하고 유저의 최종 신뢰도를 계산하도록 구현하였다.

	project_id	data_index	data	label_final
0	2.0	1.0	../food-images/pizza/1001116.jpg	pizza
1	2.0	2.0	../food-images/pizza/1008104.jpg	pizza
2	2.0	3.0	../food-images/pizza/1008144.jpg	pizza
3	2.0	4.0	../food-images/pizza/1008844.jpg	pizza
4	2.0	5.0	../food-images/pizza/1008941.jpg	pizza
...
4995	2.0	4996.0	../food-images/hot_dog/987144.jpg	hot_dog
4996	2.0	4997.0	../food-images/hot_dog/987193.jpg	hot_dog
4997	2.0	4998.0	../food-images/hot_dog/98754.jpg	hot_dog
4998	2.0	4999.0	../food-images/hot_dog/99314.jpg	hot_dog
4999	2.0	5000.0	../food-images/hot_dog/996310.jpg	hot_dog

5000 rows × 4 columns

[그림 7] 최종 레이블 확정

	user_id	user_credibility
0	mfs	0.7600
1	ukw	0.8622
2	nyi	0.9949
3	sbs	0.7015
4	vbq	0.9580
5	lzt	0.8925
6	zgo	0.6831

[그림 8] 유저의 신뢰도 최종 계산

텍스트 감정 분석에 사용할 수 있도록 KoBERT 모델을 네이버 영화 리뷰 20만개로 문장의 감정을 분석할 수 있게 끈 전이학습을 진행하였다.


```
Epoch 1/4
1500/1500 [=====] - 944s 630ms/step - loss: 0.2007 - accur
acy: 0.9206 - val_loss: 0.3015 - val_accuracy: 0.8827
Epoch 2/4
1500/1500 [=====] - 946s 631ms/step - loss: 0.1718 - accur
acy: 0.9332 - val_loss: 0.3114 - val_accuracy: 0.8796
Epoch 3/4
1500/1500 [=====] - 945s 630ms/step - loss: 0.1480 - accur
acy: 0.9438 - val_loss: 0.3257 - val_accuracy: 0.8809
Epoch 4/4
1500/1500 [=====] - 945s 630ms/step - loss: 0.1243 - accur
acy: 0.9535 - val_loss: 0.3569 - val_accuracy: 0.8786
```

[그림 9] 텍스트 모델의 학습 과정

전이학습 시킨 모델의 성능을 검사하였다.

	precision	recall	f1-score	support
0	0.89	0.86	0.88	24827
1	0.87	0.89	0.88	25173
accuracy			0.88	50000
macro avg	0.88	0.88	0.88	50000
weighted avg	0.88	0.88	0.88	50000

[그림 10] 텍스트 모델 성능 검사

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티폴모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

모델에 문장을 넣고 감정 분석을 하여 예측 값을 출력하였다.

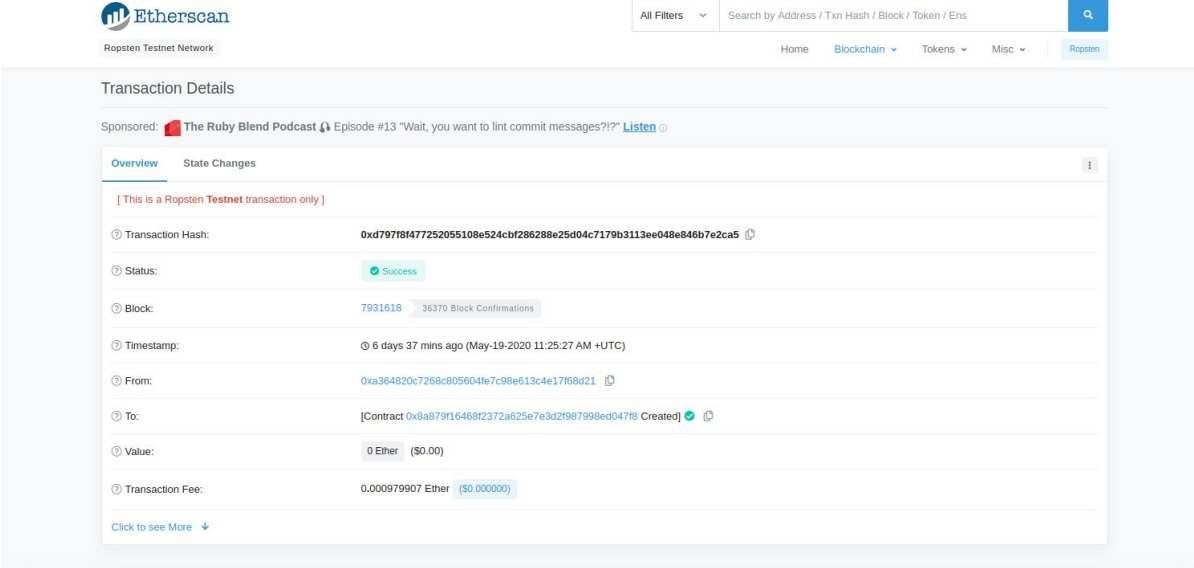
보던거라 계속보고있는데 전개도 느리고 주인공인 은희는 한두컷 나오면서 소극적인 모습에 : (부정 확률 : 0.998) 부정적인 평가입니다.

[그림 11] 모델의 테스트 문장 감정 분석

2) Blockchain


- smart contract 배포 및 테스트


포인트 생성, 전송, 소멸 등의 기능이 작성된 smart contract를 작성하였고, ropsten testnet에 배포 완료하였다. 웹서비스에서 사용할 수 있도록 web3 라이브러리를 이용하여 모듈화하였다.



The screenshot shows the Etherscan interface for a transaction on the Ropsten Testnet. The transaction is successful, with a value of 0 Ether and a transaction fee of 0.000979907 Ether. The transaction hash is 0xd797f8f47725055108e524cbf286288e25d04c7179b3113ee048e846b7e2ca5. The transaction is confirmed by 36370 blocks.


[그림 12] 배포된 계약 Overview(etherscan)

 <div> 국민대학교 컴퓨터공학부 캡스톤 디자인 I </div>	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티폴모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26



All Filters


Search by Address / Txn Hash / Block / Token / Ens



Ropsten Testnet Network

Home
Blockchain
Tokens
Misc
Ropsten

Address
0xA364820C7268C805604Fe7C98E613c4E17F68d21

Sponsored:

[Bitcoin Virtual Hackathons- Earn crypto building OSS. New prizes/webinars every week](#)
[Sign up + learn more](#)

Overview







Balance:
4.98772460047968 Ether

More Info

My Name Tag:
Not Available

Transactions

Latest 25 from a total of 26 transactions

Txn Hash	Block	Age	From	To	Value	[Txn Fee]
0xbd10adeee99fd9...	7973077	1 min ago	0xa364820c7268c8...	OUT  0xdc6aca4e6f5c59f...	0 Ether	0.000755265
0x7f9f49c10f05e486...	7973075	1 min ago	0xa364820c7268c8...	OUT  0xdc6aca4e6f5c59f...	0 Ether	0.000755265
0x51bb91e5ac2f1c1...	7973068	4 mins ago	0xa364820c7268c8...	OUT  0xdc6aca4e6f5c59f...	0 Ether	0.001147188
0x1dec40ab93a8be...	7972997	21 mins ago	0xa364820c7268c8...	OUT  Contract Creation	0 Ether	0.001240626
0x004e3d2533a4ab...	7972957	30 mins ago	0xa364820c7268c8...	OUT  0x8a879f16468f237...	0 Ether	0.000540624
0xbcd519ee184cad...	7972517	2 hrs 37 mins ago	0xa364820c7268c8...	OUT  0x8a879f16468f237...	0 Ether	0.000614002

This website uses cookies to improve your experience and has an updated Privacy Policy.
Get it


[그림 13] 배포된 contract의 State Changes log(etherscan)

캡스톤 디자인 I

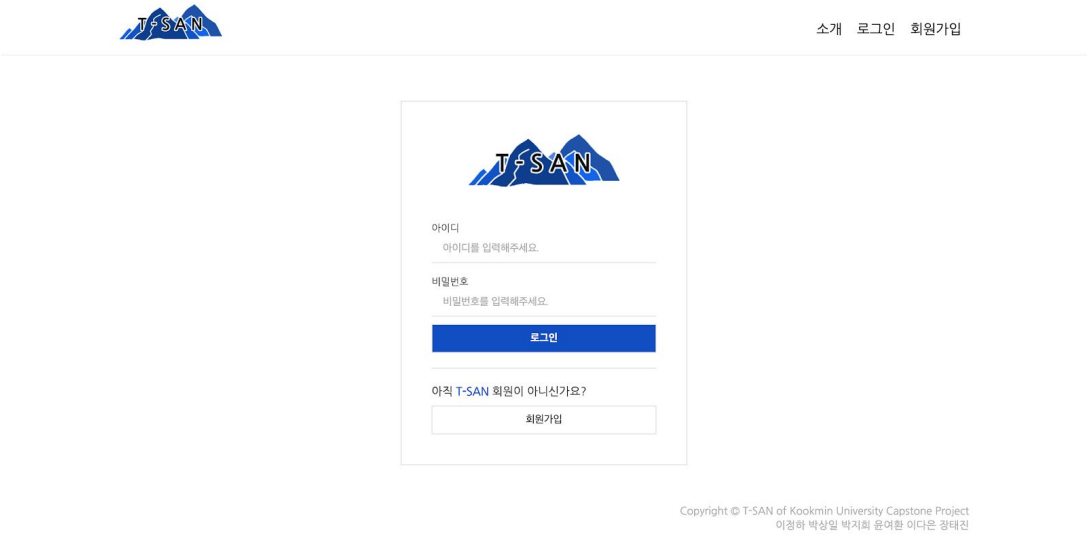
Page 12 of 29

중간보고서

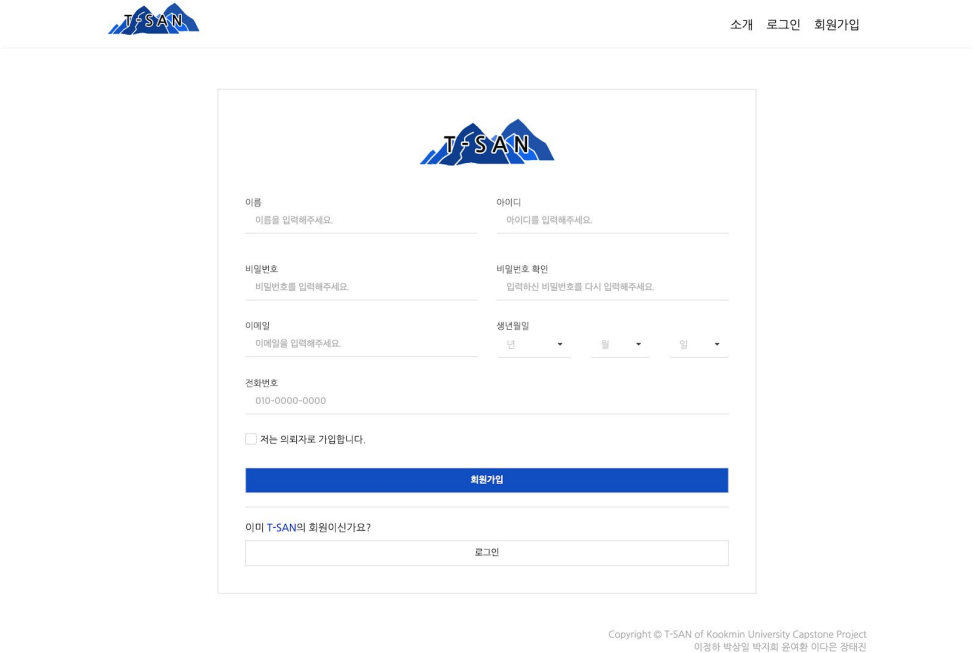
All rights are reserved. Reproduction in whole or in parts is prohibited without the written consent of the copyright owner.

 <div> <p>국민대학교 컴퓨터공학부 캡스톤 디자인 I</p> </div>	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26


3) Front-End



[그림 14] 로그인 페이지



[그림 15] 회원가입 페이지

 <div> 국민대학교 컴퓨터공학부 캡스톤 디자인 I </div>	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

- T-SAN의 첫 페이지로, 진행중인 프로젝트와 완료된 프로젝트를 카드형식으로 보여준다.



[소개](#)
[프로젝트 등록](#)
[마이페이지](#)
[로그아웃](#)

T-SAN, 티끌모아 태산

티끌 모아 태산 T-SAN에서 데이터 레이블링을 통한 보상을 얻어보세요.
 저희 T-SAN 플랫폼은 일반적인 데이터 레이블링 의뢰뿐만 아니라 자체적으로 데이터셋을 확보하여 레이블링을 진행해서 레이블링된 데이터를 공공 데이터로 공개하여 인공지능 사업 발전에 기여할 수 있습니다.

진행중인 라벨링 프로젝트



[IMAGE-영역지정] 사진에서 ...
 daeunLee | 3770P | 122일 뒤 종료

이미지캡처를 이용하여 사진에서 침대 부분을 캡처해주세요.

6/20



[IMAGE-객관식] 식품 상세정...
 sangil | 276P | 1년 1개월 12일 뒤 종료

식품 상세정보를 분석하기 위한 데이터를 수집하고 있습니다.

239/250



[TEXT-객관식] 화장품 성분 분석
 jen | 115P | 1년 1개월 5일 뒤 종료

화장품 성분 분석을 위한 데이터가 필요합니다.

286/600



[TEXT-객관식] 육설 및 비방글 ...
 jen | 100P | 23일 뒤 종료

육설 및 비방글 구분을 하기 위해 텍스트 데이터 레이블링을 진행하고 ...

78/100

완료된 라벨링 프로젝트



[IMAGE-영역지정] 화재 판단
 jiheepark | 1140P | 62일 뒤 종료

사진 속에서 화재 부분을 캡처해주세요.

48/50



[TEXT-객관식] 육설 및 비방글 ...
 jen | 100P | 11일 뒤 종료

육설 및 비방글 구분을 하기 위해 텍스트 데이터 레이블링을 진행하고 ...

100/100



[TEXT-객관식] 영어 문장 내 주...
 jen | 426P | 종료

영어 문장에서 주어를 찾아주세요.

230/230




[IMAGE-객관식] 혼잡도에 따...
 jen | 1000P | 종료

혼잡도에 따른 버스 사진을 선택해주세요.


45/50

Copyright © T-SAN of Kookmin University Capstone Project
 이경하 박상일 박지화 윤여환 이다은 강태진

[그림 16] 메인 페이지

 <div> 국민대학교 컴퓨터공학부 캡스톤 디자인 I </div>	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티폴모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

- 진행중인 프로젝트를 모두 보여준다.



[소개](#)
[프로젝트 등록](#)
[마이페이지](#)
[로그아웃](#)

현재 진행중인 프로젝트


프로젝트 이름

검색

#	주제명	레이블링 유형	시작일	마감일	진행 상황
1	대학 주제 파악	TEXT-객관식	2020.05.27	2020.07.27	0/50
2	대학 주제 파악하기	TEXT-객관식	2020.05.27	2020.07.27	0/50
3	사진에서 칭대찾기	IMAGE-영역지정	2020.02.12	2020.06.12	6/20
4	식품 상세정보 분석	IMAGE-객관식	2020.05.25	2021.06.07	239/250
5	화장품 성분 분석	TEXT-객관식	2020.05.25	2021.05.30	286/600
6	육설 및 비방글 구분을 위한 데이터 레이블링	TEXT-객관식	2020.05.25	2020.06.18	78/100
7	startTest	TEXT-객관식	2020.05.23	2020.09.30	0/10
8	state 변경 테스트	TEXT-객관식	2020.05.23	2020.09.30	0/10
9	timeout 테스트9	TEXT-객관식	2020.05.21	2020.09.30	0/10
10	timeout 테스트8	TEXT-객관식	2020.05.21	2020.09.30	0/10

[그림 17] 진행 중인 프로젝트 더보기 페이지

- 완료된 프로젝트를 모두 보여준다.



[소개](#)
[프로젝트 등록](#)
[마이페이지](#)
[로그아웃](#)


완료된 프로젝트

프로젝트 이름

검색

#	주제명	레이블링 유형	시작일	마감일	진행 상황
1	대학 주제 파악	TEXT-객관식	2020.04.02	2020.05.27	8/15
2	보도블럭에서 장애물 구분	IMAGE-영역지정	2020.03.05	2020.05.25	151/160
3	화재 판단	IMAGE-영역지정	2020.03.25	2020.05.25	48/50
4	육설 및 비방글 구분을 위한 데이터 레이블링	TEXT-객관식	2020.05.25	2020.05.25	100/100
5	영어 문장 내 주어 구분	TEXT-객관식	2020.05.25	2020.05.25	230/230
6	혼잡도에 따른 버스 사진 선택	IMAGE-객관식	2020.05.24	2020.05.25	45/50
7	모로코 요리	IMAGE-객관식	2020.05.24	2020.05.25	0/30
8	완료된 프로젝트 8	TEXT-객관식	2020.02.12	2020.04.30	0/10
9	완료된 프로젝트 7	TEXT-객관식	2020.02.12	2020.04.30	0/10
10	완료된 프로젝트 6	TEXT-객관식	2020.02.12	2020.04.30	0/10
11	완료된 프로젝트 5	TEXT-객관식	2020.02.12	2020.04.30	45791/10
12	완료된 프로젝트 4	TEXT-객관식	2020.02.12	2020.04.30	0/10
13	완료된 프로젝트 3	TEXT-객관식	2020.02.12	2020.04.30	0/10

[그림 18] 완료된 프로젝트 더보기 페이지

 <div> <p>국민대학교 컴퓨터공학부 캡스톤 디자인 I</p> </div>	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

- T-SAN 프로젝트의 개요 및 멤버를 소개한다.



소개 로그인 회원가입



소개영상



보상을 통해 데이터 라벨링을 활성화하여 AI 발전에 기여하는



프로젝트 개요

AI 시장이 확대되면서 AI 기술 개발에 사용할 수 있는 데이터셋의 필요성이 증가하고 있다. 그로 인해 데이터셋 생성의 가장 기본적인 단계인 데이터 레이블링 또한 중요해지고 있다. 데이터 레이블링이란 데이터에 대한 결과값을 붙여서 AI의 학습 단계에 사용할 수 있도록 하는 작업이다. 오늘날 데이터 레이블링 작업은 대부분 사람들에게 의해 수작업으로 진행된다. 그래서 데이터셋을 생성하는 작업은 많은 비용을 필요로 한다. 따라서 많은 기업과 연구 기관들은 AI 기술 개발에 사용할 데이터셋을 생성하는데 많은 비용을 지불하고 있으며 비용 문제로 인해 AI 기술 개발을 진행하는 것에 어려움을 겪기도 한다. 그러므로 이와 같이 데이터셋을 생성하는데 많은 비용이 드는 문제를 해결하기 위해 비용을 크게 절감 시킬 수 있는 효율적으로 데이터셋을 생성할 수 있도록 해주는 데이터 마이닝 플랫폼이 필요하다.

T-SAN 프로젝트는 클라우드소싱에 포함된 데이터 마이닝 플랫폼을 개발하는 것을 목표로 한다. 플랫폼 사용자는 의뢰자와 레이블링 참여자로 나뉜다. 의뢰자는 데이터셋과 레이블링 방법, 두 가지 정보를 제공해야 한다. 레이블링 참여자는 레이블링 작업을 할 때 보상으로 포인트를 받을 수 있다. 또한 레이블링 된 데이터를 검수하는 과정에서 머신러닝 알고리즘을 이용하여 검수 작업에 필요한 인건비를 줄이면서 신뢰도 있는 데이터셋을 수집할 수 있다. 결과적으로 T-SAN은 다수의 집단지성을 이용한 라벨링 작업과 머신 러닝 알고리즘을 이용한 검수 과정을 통해 기존의 방법 대비 적은 비용으로 데이터셋을 생성해주는 플랫폼 역할을 수행할 수 있다.

지도교수 및 팀원 소개



최은미
Professor



이정하
Front-end developer



박상일
AI developer



박지희
Blockchain developer



윤여환
Front-end developer




이다은
Back-end developer



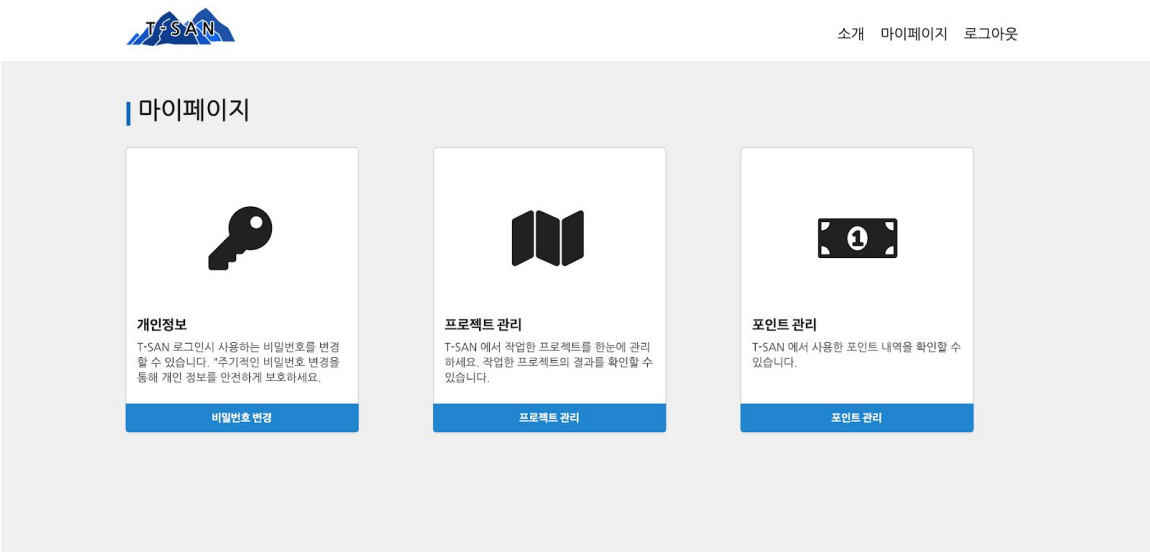
장태진
Back-end developer

Copyright © T-SAN of Kookmin University Capstone Project
이정하 박상일 박지희 윤여환 이다은 장태진

[그림 19] 프로젝트 소개 페이지

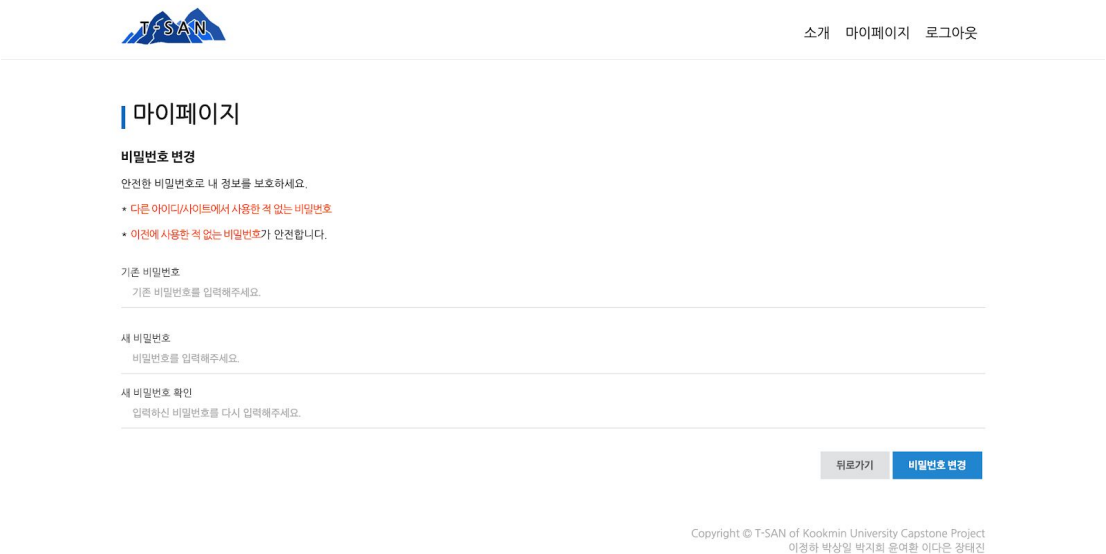
 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

- 회원들의 개인 정보, 프로젝트, 포인트를 관리할 수 있는 private한 페이지이다.




[그림 20] 마이페이지

- 사용자가 비밀번호를 변경할 수 있다.



[그림 21] 비밀번호 변경 페이지

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

- 의뢰자의 경우, 의뢰했던 프로젝트를 관리하고 레이블링 결과를 볼 수 있다. 레이블링 참여자의 경우, 참여했던 프로젝트를 볼 수 있다.



[소개](#) [프로젝트 등록](#) [마이페이지](#) [로그아웃](#)

내 프로젝트 관리

#	주제명	레이블링 유형	시작일	마감일	시작	종료	비고
1	식품 상세정보 분석	[image] 객관식	2020.05.25	2021.06.07	▶	▶	진행중

Copyright © T-SAN of Kookmin University Capstone Project
이경하 박상일 박지희 윤여환 이다은 정태진

[그림 22] 프로젝트 관리 페이지

- 사용자가 획득하거나 사용하였던 포인트 거래현황을 볼 수 있다.



[소개](#) [프로젝트 등록](#) [마이페이지](#) [로그아웃](#)

포인트 관리

[충전하기](#)

현재 회원님이 보유하고 계신 포인트


80,000P

최근 포인트 수입/지출 내역


#	포인트 변동	포인트 누적	내용
1	+100,000P	100,000P	포인트 충전
2	-10,000P	90,000P	프로젝트 의뢰
3	-10,000P	80,000P	프로젝트 의뢰

Copyright © T-SAN of Kookmin University Capstone Project
이경하 박상일 박지희 윤여환 이다은 정태진

[그림 23] 포인트 관리 페이지

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

- 의뢰자가 프로젝트를 등록할 수 있다. (레이블링 참여자는 접근이 불가하다.)



[소개](#)
[프로젝트 등록](#)
[마이페이지](#)
[로그아웃](#)

라벨링 주제 등록

주제

주제명을 적어주세요. (20자)

썸네일 이미지 (jpg, png)

Choose File

No file chosen

라벨링 유형

라벨링 유형을 선택해주세요.

한줄 설명

한줄 설명을 적어주세요. (30자)

상세 설명

내용을 입력해주세요.

총 보상설정

보상 금액을 입력해주세요.

총 참여인원 수

총 횟수를 적어주세요.

키워드

라벨링 키워드를 해시 태그 형태로 입력해주세요. (#클라 #사이다 #맥주 #소주 ...)

데이터셋을 선택해주세요.

데이터셋을 선택해주세요.

기한 설정

시작일

년

월

일

마감일

년

월

일

데이터셋에서 추출할 임의의 데이터 개수

데이터셋에서 몇개의 데이터를 뽑아서 프로젝트를 진행할까요?

제출

Copyright © T-SAN of Kookmin University Capstone Project

이정하 박상일 박지희 윤여환 이다은 정태진


[그림 24] 프로젝트 등록 페이지

캡스톤 디자인 I

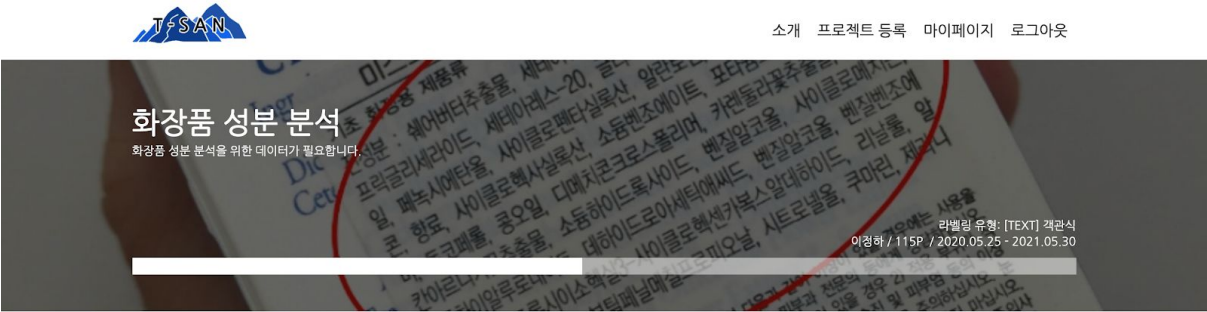
Page 19 of 29

중간보고서

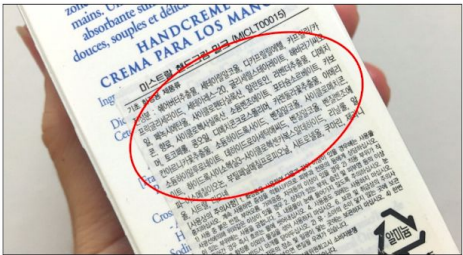
All rights are reserved. Reproduction in whole or in parts is prohibited without the written consent of the copyright owner.

 <div> 국민대학교 컴퓨터공학부 캡스톤 디자인 I </div>	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티골모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

- 레이블링 참여하기 전, 해당 레이블링 프로젝트에 대한 안내 페이지이다.



화장품 성분 분석



의뢰자
이정하
라벨링 유형
[TEXT] 객관식
기간
2020.05.25 - 2021.05.30
한 줄 설명
화장품 성분 분석을 위한 데이터가 필요합니다.
보상 금액
115P


상세 설명

안녕하세요, 저희는 화장품 성분 분석을 AI 학습을 통해 연구하는 화해팀입니다. 여러분의 많은 관심과 정성 부탁드립니다. 더불어 저희 화해도 많이 사랑해주세요. 감사합니다.

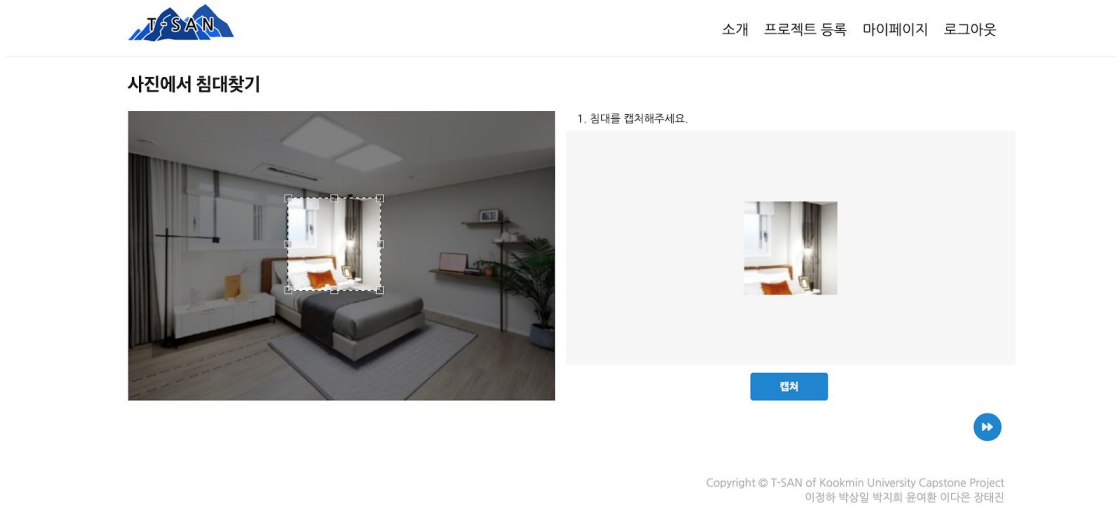
[시작하기](#)

Copyright © T-SAN of Kookmin University Capstone Project
 이정하 박상일 박지희 윤여환 이다은 강태진

[그림 25] 레이블링 안내 페이지

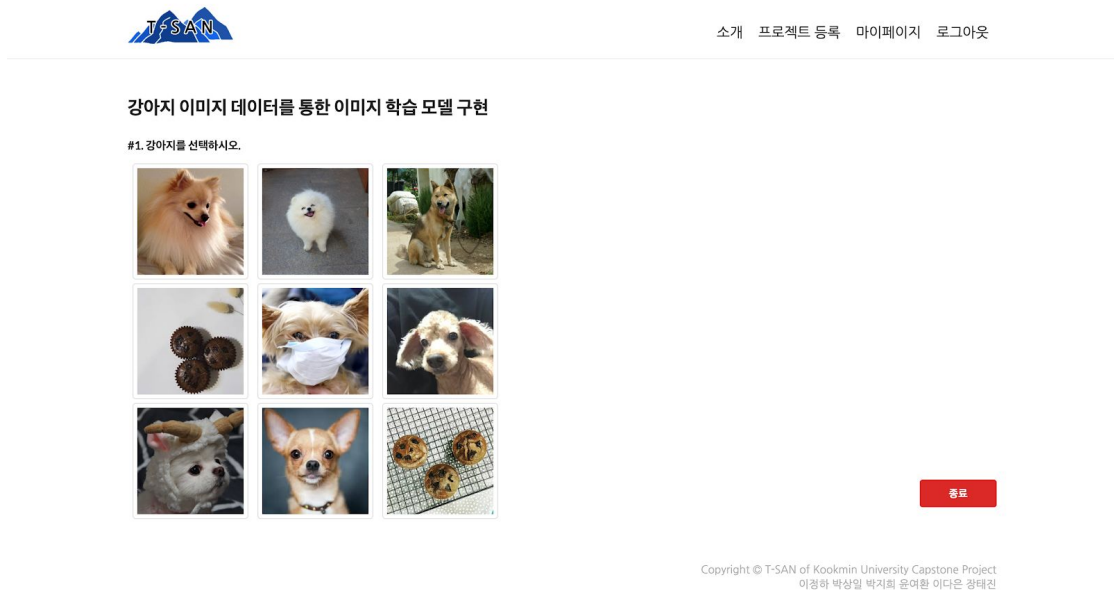
 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

- 레이블링 페이지
 - 이미지 캡처 유형의 레이블링 페이지이다.




[그림 26] [이미지 캡처] 레이블링 페이지


- 이미지 선택 유형의 레이블링 페이지이다.



[그림 27] [이미지 선택] 레이블링 페이지

 <div> 국민대학교 컴퓨터공학부 캡스톤 디자인 I </div>	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

- 텍스트 선택 유형의 레이블링 페이지이다.



[소개](#)
[프로젝트 등록](#)
[마이페이지](#)
[로그아웃](#)

뉴스 데이터 분석

웹베이스 40mm 늘어난 현대 더 뉴 그랜저 (고양=뉴스1) 유승관 기자 = 19일 경기 고양시 일산 빛마루 방송지원센터 열린 현대자동차 '더 뉴 그랜저' 신차발표회에서 새로운 그랜저가 공개되고 있다. 더 뉴 그랜저는 2016년 11월 출시 이후 3년 만에 선보이는 6세대 그랜저의 페이스리프트 모델이다. 더 뉴 그랜저는 지난 4일부터 18일까지 영업일 기준 11일 동안 사전계약 3만2179대를 기록했다. 2019.11.19/뉴스1fotogyoo@news1.kr), 무단 전재 및 재배포 금지]

다음 기사의 주제에 맞는 것을 선택해주세요.

정치/사회

연예

경제

기타

▶

남은 라벨링: 999

Copyright © T-SAN of Kookmin University Capstone Project
이정하 박상일 박지희 윤여환 이다은 정태진

[그림 28] [텍스트 선택] 레이블링 페이지

4) Web Back-End

- GraphQL을 사용하여 다양한 요청에 대한 응답이 가능하게 구현

요청받은 질의에 대한 데이터를 DB에 조회 / 저장한다. 아래 예시는 구현한 API 요청/응답 예시 중 하나로, 특정 의뢰자가 등록한 데이터 레이블링 주제에 대한 요청/응답이다. 오른쪽 응답에 대한 결과를 보면 해당 프로젝트에 대한 username의 값이 모두 각각 'jen'인 것을 확인할 수 있다. JWT token에 내재한 사용자 정보를 해독한 후, 해당 사용자의 프로젝트 인스턴스들을 추출한다. 또한 해당 API는 offset, limit라는 인자를 옵션으로 사용하여 특정 순서 이후의 인스턴스와 인스턴스의 개수를 임의로 지정하여 요청 받을 수 있도록 구현하였다.

GraphQL
Prettify
Merge
Copy
History
Docs

```

1 query {
2   requesterRequest(token: "eyJ0eXAiOiJKV1QiLCJhbGciOiJI
3     message {
4       status
5       message
6     }
7   requests {
8     idx
9     user {
10      username
11    }
12    subject
13    description
14  }
15 }
16


```

```

{
  "data": {
    "getRequesterRequest": {
      "message": {
        "status": true,
        "message": ""
      },
      "requests": [
        {
          "idx": "55",
          "user": {
            "username": "jen"
          },
          "subject": "화장품 성분 분석",
          "description": "안녕하세요. 저희는 화장품 성분 분석을 AI 학습을 통해 연구하는 화해팀입니다. 여러 분들의 많은 관심과 정성 부탁드립니다. 더불어 저희 화해도 많이 사랑해주세요. 감사합니다. "
        },
        {
          "idx": "54",
          "user": {
            "username": "jen"
          },
          "subject": "육질 및 비방글 구분을 위한 데이터 레이블링",
          "description": "안녕하세요. 사이트 내의 육질 및 비방글을 필터링 하기 위한 인공지능 학습 데이터 셋을 수집하고 있습니다. 잘 부탁드립니다."
        }
      ]
    }
  }
}

```


[그림 29] 특정 의뢰자가 등록한 데이터셋 조회 API 요청과 응답

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

● 구현한 API 목록


[표1] 구현한 API목록

구분	함수명	비고
Mutation	CreateAccount	회원가입API
Mutation	CreateDataset	데이터셋 등록API
Mutation	CreateCategory	라벨링 유형 등록API
Mutation	CreateRequest	라벨링 프로젝트 등록API
Mutation	CreatePaymentLog	포인트 지급 기록API
Mutation	LoginAccount	로그인API
Mutation	UpdateCategory	라벨링 유형 수정API
Mutation	UpdateAccount	개인 정보 수정API
Mutation	UpdateRequest	라벨링 프로젝트 수정API
Mutation	UpdatePaymentLog	포인트 지급 기록 수정API
Query	getAllDataset	모든 데이터셋 조회API
Query	getAllRequest	모든 라벨링 프로젝트 조회API
Query	getRequesterRequest	특정 의뢰자 프로젝트 조회API
Query	getAllCategory	모든 라벨링 유형 조회API
Query	getAllUser	모든 사용자 조회API
Query	getAllPaymentLog	모든 포인트 지급 조회API

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

[표2] 중간평가 이후 구현한 API목록

구분	함수명	비고
Mutation	addPoint	포인트 충전API
Mutation	startRequest	라벨링 프로젝트 시작API
Mutation	endRequest	라벨링 프로젝트 종료API
Mutation	deleteUser	회원 탈퇴API
Mutation	deletePaymentlog	금액 지급 내역 삭제API
Mutation	deleteLabelerTakenProject	참여자가 신청한 프로젝트 삭제API
Mutation	updatePassword	비밀번호 변경API
Mutation	takeProject	참여자가 프로젝트 신청API
Query	getLabelerTakenProject	참여자가 신청한 프로젝트 조회API
Query	getIdxRequest	특정 라벨링 프로젝트 조회API
Query	getSubjectRequest	제목으로 라벨링 프로젝트 검색API
Query	getSubjectRunningRequest	진행 중인 프로젝트 중 제목으로 검색API
Query	getSubjectEndRequest	끝난 프로젝트 중 제목으로 검색API
Query	getStateRequest	특정 상태의 프로젝트 조회API

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

3 수정된 연구내용 및 추진 방향

3.1 수정사항

1) Machine Learning


- 텍스트 데이터 검증에 딥러닝 모델 사용
 - 기존에 클러스터링으로 진행하는 과정을 비지도학습 특성으로 인한 검증 과정 조정과 검증 정확도 개선의 어려움이 발생하여 딥러닝 모델을 학습시켜서 사용하도록 수정하였다. 딥러닝 모델은 적은 데이터 셋으로도 전이 학습이 가능한 BERT를 선정하였고, 그 중 한국어 corpus로 사전 학습 시킨 KoBERT를 사용하도록 한다.
- 이미지 데이터 검증 모델 재학습 삭제
 - 처음 학습 단계에서도 overfitting 양상을 보이며, 실제 재학습을 시켜도 별다른 차이가 없음을 확인하였다.
- T-SNE 시각화 모듈 삭제
 - 최종 결과를 레이블링 결과와 신뢰도 수치를 그래프화 하여 보여주는 것으로 수정하였다.
- 2차 레이블링 요청 기능 삭제
 - 2차 레이블링을 위해 검증하던 작업을 중지하고 데이터들을 보존하고 있는 것을 구현하는데 필요한 과정을 구현하여 백엔드와 연결시키는 것이 한계로 인해 삭제하고, 최대한 레이블링을 확정하는 것으로 수정하였다.

2) Blockchain

- testnet(ropsten) 배포
 - 기존에는 geth를 이용하여 private network에 계약을 배포하고 테스트 해보았으나, 로컬 환경에서 개발이 완료되었다고 판단하여 테스트넷 환경에서 실행 및 테스트 중이다.
 - 테스트넷에 직접 fully sync를 하기 위해서 저장공간 및 시간이 필요하기 때문에 Infura를 이용하여 테스트넷 노드에 접근했다.

3) Front-End


- 반응형 웹 페이지 구현 축소

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티폴모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

- 반응형 웹 페이지를 구현하기 보다는 고정 웹 페이지의 완성도를 높이는 것으로 추진 방향을 수정했다.
- 데이터셋 업로드 기능 삭제
 - .zip 파일을 받아서 서버에서 처리하는 것을 구현하지 않고 T-SAN 에서 데이터를 제공해주는 방식으로 구현하는 것으로 수정했다.
- 회원 가입시 번호 인증 삭제
 - 회원 가입 기능 구상 당시에 번호 인증 기능이 필요하다고 판단하여 회원 가입 페이지의 초기 디자인에는 번호 인증 기능이 있었으나, NICE 아이디 - 휴대폰 본인 확인 서비스를 도입하기에는 비용 문제가 있어서 구현하지 않는 것으로 수정했다.

4) Back-End

- 데이터베이스 필드 추가
 - 반복적인 검토로 요구사항이 구체화되면서 프로젝트 테이블에 대한 썸네일, 설명문, 프로젝트 진행 상태 등의 필드를 추가했다. 또한 회원 테이블에서 동명이인이 가입하는 경우를 위해 실명(fullname)과 닉네임(username)을 구분하고, boolean타입의 is_requester 필드를 사용하여 의뢰자와 참여자에 대해 구분짓는 구조로 변경하였다.
- API 수정
 - API 기능적 측면에서 구체화가 필요한 작업을 추가적으로 진행한다. 예를 들어, 데이터 라벨링 프로젝트 검색 및 조회에서 참여자가 여러 조건에 대해 필터링할 수 있는 기능을 추가하였다.
 - UX/UI의 연동과정에서 요구된 인터페이스 입출력에 대해 수정한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

4 향후 추진계획

4.1 향후 계획의 세부 내용

1) Machine Learning


- 레이블 검증 모듈 AWS EC2에 구동
 - EC2의 컴퓨팅 리소스를 이용하여 레이블 검증 모듈을 구동시킨다.
- 백엔드와의 연동
 - 백엔드로부터 검증 작업을 진행한 데이터 셋을 받은 뒤, 데이터 셋 레이블을 확정하고, 회원의 수정된 신뢰도를 백엔드로 보낸다.
- 최종 레이블 확정 과정 개선
 - 머신러닝의 예측 레이블을 반영하여 참여자 다수에 의해 설정된 임시 레이블의 오류를 찾아 개선할 수 있는 방향으로 최종 레이블 확정 과정을 개선한다.

2) Blockchain

- transaction hash 확인 기능
 - 포인트의 투명성을 위해 포인트 거래 마다 Transaction이 생성되면, 이것을 확인할 수 있는 기능을 구현할 예정이다.
- 서버와의 연동
 - 실제 프로젝트에서의 사용을 위해 Back-end와 연동하여 API로 구현한다.

3) Front-End

- Console error 검수 작업
 - Product 를 배포하기 전, 모든 페이지의 오류를 확인하여 오류가 없는 product를 배포할 수 있도록 코드를 검토할 계획이다.
- UX/UI 검수 작업
 - 웹 페이지의 모든 기능을 테스트하면서 사용자에게 가장 친화적인 UX/UI 를 제공할 수 있도록 UX/UI 오류를 찾아서 수정할 계획이다.
- 지속적인 유지 보수 및 테스트
 - 기존의 사용하고 있는 GraphQL query 와 mutation 이 Back-End 개발자에 의해서 수정될 수 있기 때문에 Back-End 개발자와 지속적으로 커뮤니케이션하며 query 와 mutation 에 수정 사항이 있다면 수정하고 테스트를 할 계획이다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	2차 중간보고서		
	프로젝트 명	집단지성을 이용한 데이터 레이블링 플랫폼	
	팀 명	T-SAN(티끌모아 태산)	
	Confidential Restricted	Version 1.6	2020-05-26

4) Back-End

- Machine Learning 모듈과 연동 작업
 - ML 모듈에서 계산한 신뢰도 결과를 해당 회원의 신뢰도에 저장하는 작업을 진행한다.
회원의 신뢰도 데이터는 MariaDB에서 관리된다.
- 블록체인 모듈과 연동 작업
 - 의뢰자 프로젝트 생성, 참여자에 대한 보상에 관련하여 블록체인 API를 연동하는 작업을 진행한다.
- API 함수 유지 보수
 - 프론트엔드에서 API 연동을 진행하면서 생기는 이슈에 대해 검토하며 수정을 진행한다. 또한 검색 기능 등 추가적으로 요구하는 API를 구현 및 수정한다.