

AI 기반 무역거래 자금세탁방지 솔루션

Anti-TBML 시스템 개발

- 최종 발표 -

하나금융IT

Mentor

김신웅 기술사님

CONTENTS

01

프로젝트 개요

1. Anti-TBML이란?
2. 주요 기능
3. 목표 시스템
4. R & R
5. 협업 방식
6. 개발 일정 & WBS

02

사용 기술

1. AWS RDS
2. MySQL
3. Google Vision API
4. Edit Distance

03

결과물

1. 시스템 구조
2. 상세 기능
3. 시연 영상

04

결론

1. 기대 효과
2. 개선 방안

01

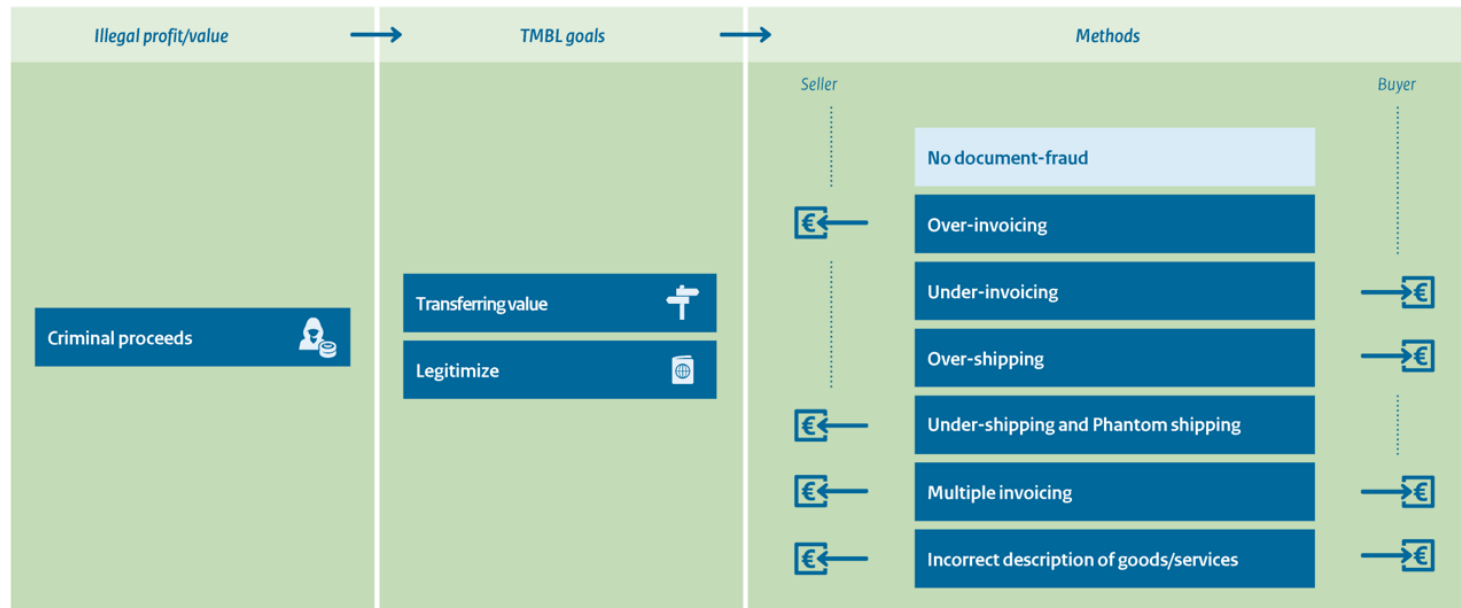
프로젝트 개요

01. Anti-TBML이란?

프로젝트 개요

TBML (Trade-Based Money Laundering) : 무역금융 과정에서 일어날 수 있는 범죄 자금 조달 행위

Trade Based Money Laundering



- 무역 거래를 활용하여 수익을 얻는, 또는 이동시키는 방식으로 이뤄짐
- 해당 수익은 자금 세탁, 밀반입, 테러와 같은 범죄 자금으로 활용
- Anti-TBML은 이러한 범죄 자금 조달 행위를 사전에 탐지하고 차단하기 위한 시스템

02. 주요 기능

프로젝트 개요

데이터베이스

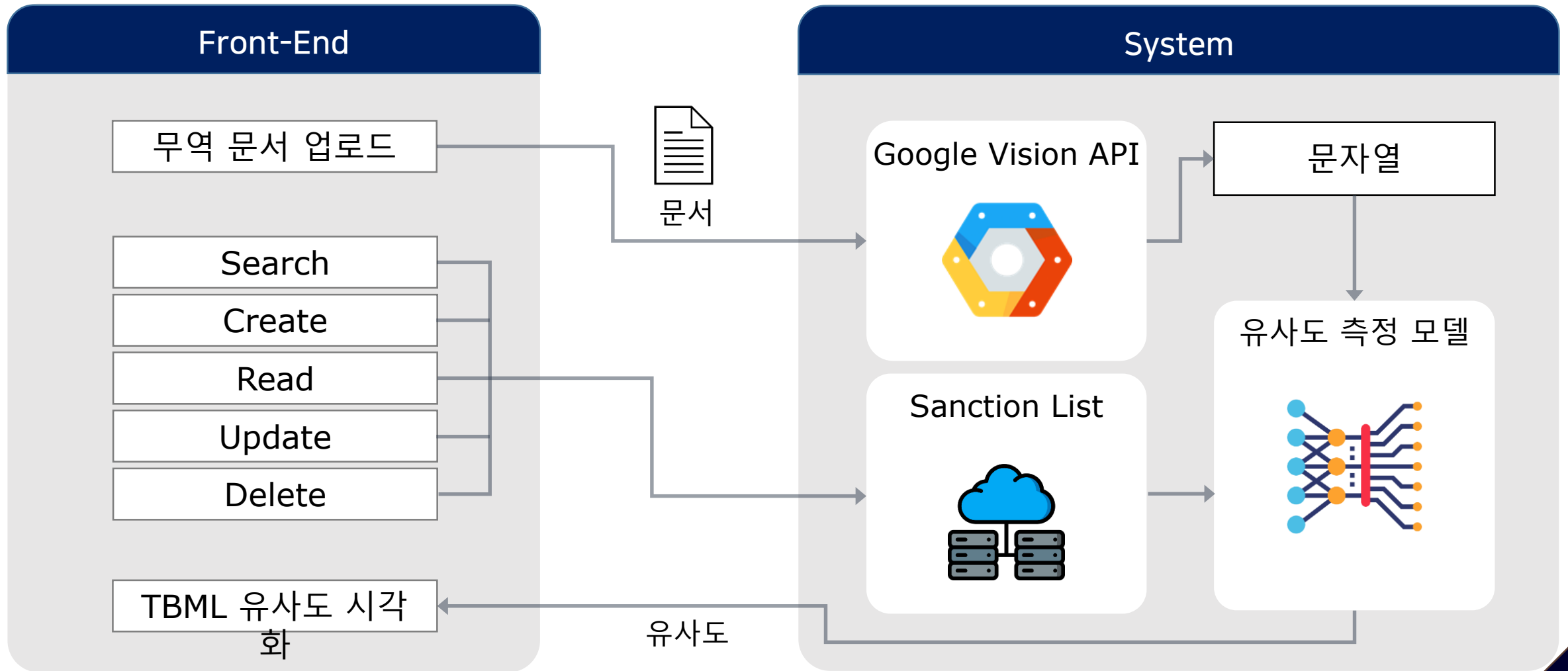
- 미국 재무부 홈페이지에 있는 sanction list (제재 리스트) 를 크롤링하여 데이터베이스에 저장
- 개인·국가·기업·선박 등 다양한 데이터를 저장할 수 있는 테이블을 구성

웹 (Front-End & Back-End)

- sanction list에 대한 CRUD 및 검색 작업 수행
- 사용자가 무역 문서를 업로드
- OCR을 이용하여 해당 문서의 내용을 문자열 단위로 추출
- 추출된 문자열과 text matching 알고리즘 및 모델을 이용하여 TBML과의 유사도 측정

03. 목표 시스템

프로젝트 개요



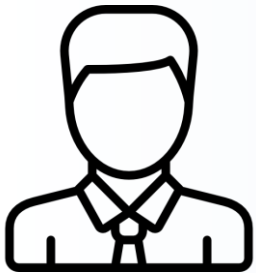
04. R & R

프로젝트 개요



서준교 (PM)

데이터베이스 구축 및
Front-End & Back-End 개발



최기훈

text matching 및 유사도 측정
알고리즘 설계 및 평가

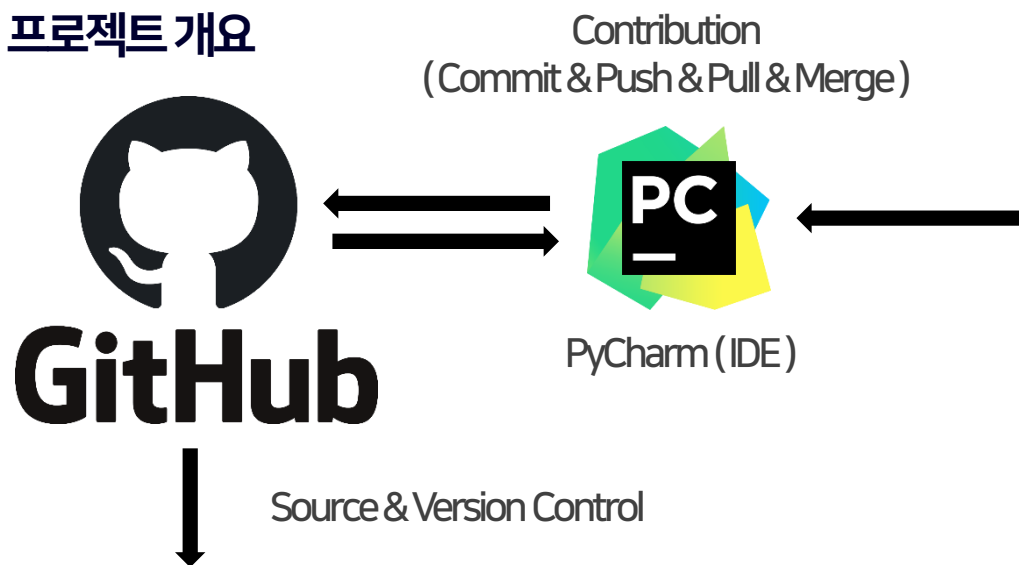


조현아

Front-End & Back-End 개발

05. 협업 방식

프로젝트 개요

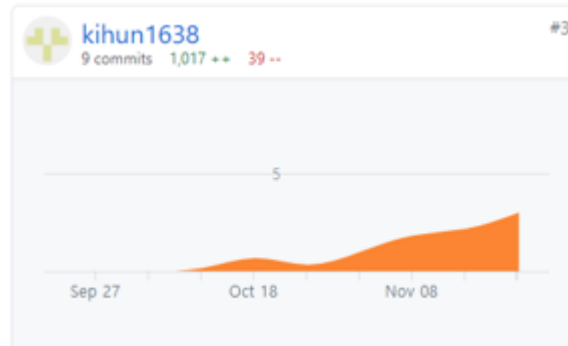
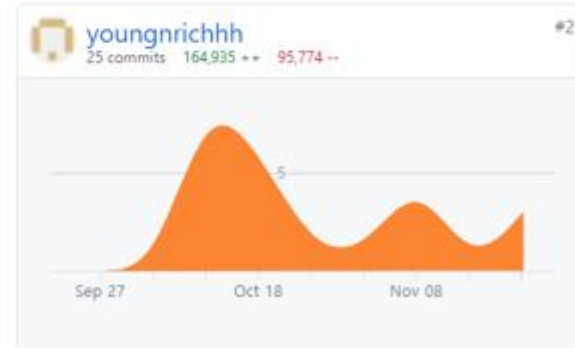
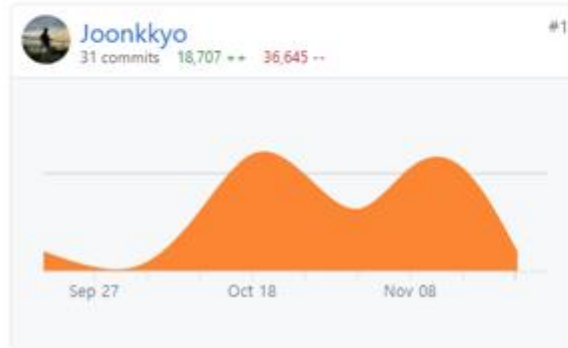


	Joonkkyo threshold 이상의 유사도를 가진 word의 boundary만 출...	c50d1e9 3 days ago	🕒 62 commits
📁 .idea	코드 정리	10 days ago	
📁 assets	navbar 문제 해결	6 days ago	
📁 config	주말 전 수정	10 days ago	
📁 document_inspection	threshold 이상의 유사도를 가진 word의 boundary만 출...	3 days ago	
📁 media/documents	navbar 문제 해결	6 days ago	
📁 model	GOOGLE_VER.py 수정	6 days ago	
📁 sanction	navbar 문제 해결	5 days ago	
📁 templates	navbar 문제 해결	6 days ago	
📄 manage.py	Initial commit	2 months ago	
📄 requirements.txt	requirements.txt 생성	12 days ago	

Sep 20, 2020 – Nov 26, 2020

Contributions: Commits ▾

Contributions to master, excluding merge commits



프로젝트 개요

데이터베이스 구축 및 연동

Text matching 및 유사도 측정 알고리즘 개발

발표준비

[illegible]

02

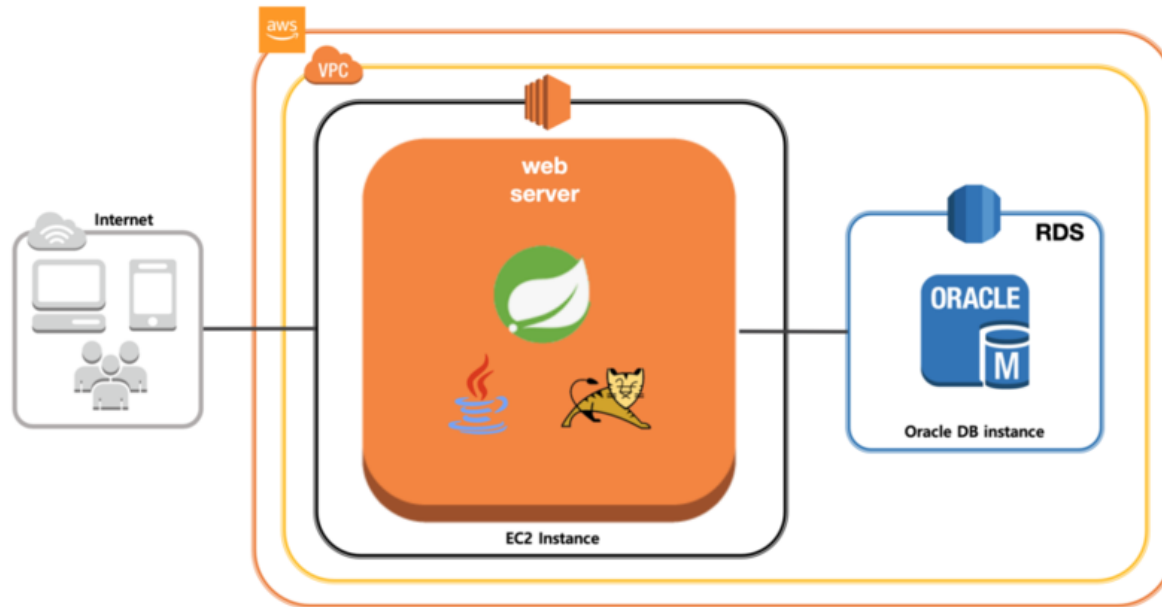
사용 기술

01. AWS RDS

사용 기술

Amazon Relational Database Service (RDS)

- AWS에서 제공하는 클라우드에서 관계형 데이터베이스를 쉽게 설치, 운영 및 확장할 수 있는 웹 서비스
- 애플리케이션 내에서 관계형 데이터베이스의 설정, 운영, 스케일링을 단순케 하도록 설계됨



- AWS EC2가 서비스되고 있고, RDS 서비스가 EC2와 연동되어 클라이언트가 **EC2에 접속하여 데이터베이스 정보를 조회하는 구조**
- **사용 목적:** 원격 개발 환경에서 데이터베이스에 상시 접근할 수 있어야 해야 하기 때문에, 클라우드 환경에서 안정적으로 데이터베이스 서비스를 제공하는 AWS RDS 사용

02. MySQL

사용 기술

데이터베이스 설계 (MySQL)

- Id, name, type, program으로 총 네 가지의 field로 구성
- type는 제재 대상의 종류 (선박, 기업, 인물 등), program은 제재 프로그램의 명칭을 의미
- Id에 auto_increment 옵션을 적용하여 sanction list에 대한 index 관리에 용이하도록 함

```
mysql> desc sanction;
```

Field	Type	Null	Key	Default	Extra
id	int	NO	PRI	NULL	auto_increment
name	varchar(130)	YES		NULL	
type	varchar(20)	YES		NULL	
program	varchar(60)	YES		NULL	

```
4 rows in set (0.01 sec)
```

< sanction 테이블 구조 >

02. MySQL

사용 기술



< 미국 재무부 sanction list >

django

< Django back-end >



```
FIRST DIVISION OF THE SYRIAN ARAB ARMY  
XINJIANG PRODUCTION AND CONSTRUCTION CORPS  
SUN, Jinlong  
PENG, Jiarui  
HAMUD, Faruq  
TAGWIREI, Kudakwashe Regimond  
SAKUNDA HOLDINGS  
PARTHIA CARGO LLC  
MAHDAVI, Amin  
DELTA PARTS SUPPLY FZC  
AL SHIBL, Luna  
SAATI BIN MOHAMAD NAWZAD, Mohamad Ammar  
ISMAIL, Samer  
DALAH, Ghaith  
SAQR, Fadi  
IBRAHIM, Yasser Hussein
```

3291 rows in set (0.00 sec)

< MySQL Database >

02. MySQL

사용 기술

< AWS RDS >

anti-tbml 수정 작업 ▼

요약

DB 식별자 anti-tbml	CPU 1.67%	정보 ✓ 사용 가능	클래스 db.t2.micro
역할 인스턴스	현재 활동 2 연결	엔진 MySQL Community	리전 및 AZ ap-northeast-2c

```
DATABASES = {  
    'default': {  
        'ENGINE': 'django.db.backends.mysql',  
        'NAME': 'ANTI_TBML',  
        'USER': 'admin',  
        'PASSWORD': '  
        'HOST': 'anti-tbml.cm13xr80y2ts.ap-northeast-2.rds.amazonaws.com',  
        'PORT': '3306',  
        'OPTIONS': {  
            'init_command': 'SET sql_mode="STRICT_TRANS_TABLES"'  
        }  
    }  
}
```

연결 & 보안

엔드포인트 및 포트

엔드포인트

anti-tbml.cm13xr80y2ts.ap-northeast-2.rds.amazonaws.com

포트

3306

< Django의 settings.py >

SANCTION

Sanction lists + 추가

인증 및 권한

그룹 + 추가

사용자(들) + 추가

변경할 sanction list 선택

SANCTION LIST 추가 +

액션:

 실행 100 중 아무것도 선택되지 않았습니다.

<input type="checkbox"/>	ID	NAME	TYPE	PROGRAM
<input type="checkbox"/>	8292	IBRAHIM, Yasser Hussein	individual	SYRIA-E013894
<input type="checkbox"/>	8291	SAQR, Fadi	individual	SYRIA-E013894
<input type="checkbox"/>	8290	DALAH, Ghaith	individual	SYRIA-E013894
<input type="checkbox"/>	8289	ISMAIL, Samer	individual	SYRIA-E013894
<input type="checkbox"/>	8288	SAATI BIN MOHAMAD NAWZAD, Mohamad Ammar	individual	SYRIA
<input type="checkbox"/>	8287	AL SHIBL, Luna	individual	SYRIA

•
•
•

<input type="checkbox"/>	8199	MORON HERNANDEZ, Santiago Jose	individual	VENEZUELA
<input type="checkbox"/>	8198	KALAI, Nader	individual	SYRIA] [SYRIA-CAESAR
<input type="checkbox"/>	8197	AL-ZUBAIDI, Khaled	individual	SYRIA] [SYRIA-CAESAR
<input type="checkbox"/>	8196	EBLA HOTEL	-0-	SYRIA
<input type="checkbox"/>	8195	CASTLE INVESTMENT HOLDING	-0-	SYRIA
<input type="checkbox"/>	8194	ZUBAIDI AND QALEI LLC	-0-	SYRIA
<input type="checkbox"/>	8193	TELEFOCUS CONSULTANTS INC	-0-	SYRIA

03. Google Vision API

사용 기술

Google Cloud의 Vision API

- REST 및 RPC API를 통해 선행 학습된 강력한 머신러닝 모델 기반 API
- 이미지에서 텍스트를 감지하고 추출하기 위한 목적으로 사용
- 해당 문서에 대한 페이지, 블록, 단락, 단어, 줄바꿈 정보를 JSON 형태로 반환



JSON

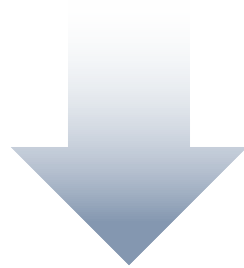
```
{
  "requests": [
    {
      "image": {
        "content": "base64-encoded-image"
      },
      "features": [
        {
          "type": "TEXT_DETECTION"
        }
      ]
    }
  ]
}
```

03. Google Vision API

사용 기술

< JSON data from Google Vision API >

```
{'ORIGINAL TEL : SWIFT : KOEX': {'place': [(1269, 66), (1506, 66), (1506, 171), (1269, 171)], 'danger': [], 'similar_word': []}, '*** This is a constituent and integral part of advice No : A12WH191': {'place': [(133, 274), (1234, 268), (1234, 299), (133, 305)], 'danger': [], 'similar_word': []}, '( OP )': {'place': [(1347, 270), (1399, 270), (1399, 297), (1347, 297)], 'danger': [], 'similar_word': []}, '* KRYEE ==': {'place': [(133, 331), (273, 331), (273, 341), (133, 341)], 'danger': [], 'similar_word': []}, 'PREPAID INDICATING LC NUMBER , NUMBER NIE OF APPLICANT': {'place': [(233, 362), (1107, 357), (1107, 380), (233, 385)], 'danger': [], 'similar_word': []}, '09981': {'place': [(234, 402), (310, 401), (310, 422), (234, 423)], 'danger': [], 'similar_word': []}, 'ETABLISHED IN SOUTH KOREA .': {'place': [(496, 402), (919, 400), (919, 422), (496, 424)], 'danger': [], 'similar_word': []}, 'COMMERCIAL': {'place': [(234, 482), (396, 482), (396, 511), (234, 511)], 'danger': [], 'similar_word': []}, 'INVOICE': {'place': [(416, 482), (528, 482), (528, 511), (416, 511)], 'danger': [], 'similar_word': []}, 'IN': {'place': [(238, 1479), (267, 1479), (267, 1503), (238, 1503)], 'danger': [], 'similar_word': []}, '03': {'place': [(598, 482), (626, 482), (626, 511), (598, 511)], 'danger': [], 'similar_word': []}, 'ORIGINALS': {'place': [(646, 482), (792, 482), (792, 511), (646, 511)], 'danger': [], 'similar_word': []}, 'INDICATED': {'place': [(814, 482), (960, 482), (960, 511), (814, 511)], 'danger': [], 'similar_word': []}, 'AMOUNT': {'place': [(764, 1561), (861, 1561), (861, 1589), (764, 1589)], 'danger': [], 'similar_word': []}, ',': {'place': [(784, 1521), (789, 1521), (789, 1549), (784, 1549)], 'danger': [], 'similar_word': []}, 'ORIGIN': {'place': [(696, 649), (793, 649), (793, 675), (696, 675)], 'danger': [], 'similar_word': []}, 'OF': {'place': [(881, 1561), (912, 1561), (912, 1589), (881, 1589)], 'danger': [], 'similar_word': []}, 'GOODS': {'place': [(283, 526), (361, 526), (361, 552), (283, 552)], 'danger': [], 'similar_word': []}, 'AND': {'place': [(1162, 1479), (1211, 1479), (1211, 1503), (1162, 1503)], 'danger': [], 'similar_word': []}, 'PAYMENT': {'place': [(448, 525), (560, 525), (560, 552), (448, 552)], 'danger': [], 'similar_word': []}, 'TERMS': {'place': [(580, 525), (658, 525), (658, 551), (580, 551)], 'danger': [], 'similar_word': []}, 'BY': {'place': [(516, 1521), (546, 1521), (546, 1549), (516, 1549)], 'danger': [], 'similar_word': []}, 'LETTER': {'place': [(729, 524), (826, 524), (826, 550), (729, 550)], 'danger': [], 'similar_word': []}, 'CREDIT': {'place': [(1163, 1521), (1260, 1521), (1260, 1549), (1163, 1549)], 'danger': [], 'similar_word': []}, 'SIGNED': {'place': [(235, 694), (330, 694), (330, 715), (235, 715)], 'danger': [], 'similar_word': []}, 'STAMPED': {'place': [(236, 567), (347, 567), (347, 590), (236, 590)], 'danger': [], 'similar_word': []}, 'THE': {'place': [(1229, 1687), (1275, 1687), (1275, 1707), (1229, 1707)], 'danger': [], 'similar_word': []}, 'BENEFICIARY': {'place': [(779, 692), (959, 691), (959, 713), (779, 714)], 'danger': [], 'similar_word': []}, 'COPY': {'place': [(234, 651), (297, 651)
```



```
'ACCEPTANCE': {'place': [(334, 1521), (496, 1521), (496, 1549), (334, 1549)], 'danger': [1.0], 'similar_word': ['ACCEPTANCE']},
```


ORIGINAL

TEL:
SWIFT:KOEX

*** This is a constituent and integral part of advice No : A12WH191000040 (OP)

:44E Port of Loading/Airport of Departure :PORT SOUTH KOREA
:44F Port of Discharge/Airport of Destination:PORT ORAN
:44C Latest Date of Shipment :191210
:45A Description of Goods and/or Services :
PORT ORAN (INCOTERMS 2010)

FULL AUTOMATIC DISPOSAL SYRINGE PRODUCTION IN LINE FOR
5 ML

FOLLOWING TO THE PROFORMA INVOICE NO:DMT0916
DATED:16/09/2019

:46A Documents Required :
95 PCT OF THE AMOUNT OF THE L/C OF USD 190.000,00 PAYABLE
AGAINST PRESENTATION OF THE FOLLOWING DOCUMENTS:

COMMERCIAL INVOICE IN 03 ORIGINALS INDICATED AMOUNT,ORIGIN
OF GOODS AND PAYMENT TERMS BY LETTER OF CREDIT, SIGNED AND
STAMPED BY THE BENEFICIARY

2/3 BILL OF LADING ORIGINALS ON BOARD TO THE ORDER OF BEA
BRANCH TLEMEN EL KIFFANE NOTIFY APPLICANT MARKED FREIGHT
PREPAID INDICATING LC NUMBER, NUMBER NIF OF APPLICANT
ESTABLISHED IN SOUTH KOREA.

ORIGINAL CERTIFICATE OF ORIGIN (ORIGIN SOUTH KOREA) DELIVERED
AND SIGNED BY CHAMBER OF COMMERCE OF BENEFICIARY COUNTRY

COPY CERTIFICATE OF CONFORMITY SIGNED AND STAMPED BY
BENEFICIARY

COPY CERTIFICATE OF WARRANTY ESTABLISHED BY THE MANUFACTURER
(TECHNO).

PACKING LIST SIGNED AND STAMPED BY THE SELLER (
TECHNO) IN 2 COPIES

BENEFICIARY'S ATTESTATION CERTIFYING THAT THE FOLLOWING
DOCUMENTS HAVE BEEN SENT BY SHIP BAG(EXPRESS MAIL) UPS/DHL
DIRECTLY TO THE APPLICANT:

1/3 BILL OF LADING ORIGINAL ON BOARD TO THE ORDER OF
BRANCH TLEMEN EL KIFFANE NOTIFY APPLICANT MARKED FREIGHT

*** This advice to be continued next page

(A12WH1910

)

ORIGINAL
ORIGINAL
TEL: SWIFT: KOEX
TEL:
SWIFT: KOEX

*** This is a constituent and integral part of advice No : A12WH191000040 (OP)
*** This is a constituent and integral part of advice No : A12WH191000040 (OP)

:44E Port of Loading/Airport of Departure :PORT SOUTH KOREA
:44F Port of Discharge/Airport of Destination:PORT ORAN
:44C Latest Date of Shipment :191210
:45A Description of Goods and/or Services :
PORT ORAN (INCOTERMS 2010)

FULL AUTOMATIC DISPOSAL SYRINGE PRODUCTION IN LINE FOR
5 ML

FOLLOWING TO THE PROFORMA INVOICE NO:DMT0916
DATED:16/09/2019

:46A Documents Required :
95 PCT OF THE AMOUNT OF THE L/C OF USD 190.000,00 PAYABLE
AGAINST PRESENTATION OF THE FOLLOWING DOCUMENTS:

COMMERCIAL INVOICE IN 03 ORIGINALS INDICATED AMOUNT,ORIGIN
OF GOODS AND PAYMENT TERMS BY LETTER OF CREDIT, SIGNED AND
STAMPED BY THE BENEFICIARY

2/3 BILL OF LADING ORIGINALS ON BOARD TO THE ORDER OF BEA
BRANCH TLEMEN EL KIFFANE NOTIFY APPLICANT MARKED FREIGHT
PREPAID INDICATING LC NUMBER, NUMBER NIF OF APPLICANT
ESTABLISHED IN SOUTH KOREA.

ORIGINAL CERTIFICATE OF ORIGIN (ORIGIN SOUTH KOREA) DELIVERED
AND SIGNED BY CHAMBER OF COMMERCE OF BENEFICIARY COUNTRY

COPY CERTIFICATE OF CONFORMITY SIGNED AND STAMPED BY BENEFICIARY

COPY CERTIFICATE OF WARRANTY ESTABLISHED BY THE MANUFACTURER
(TECHNO).

PACKING LIST SIGNED AND STAMPED BY THE SELLER (
TECHNO) IN 2 COPIES

BENEFICIARY'S ATTESTATION CERTIFYING THAT THE FOLLOWING
DOCUMENTS HAVE BEEN SENT BY SHIP BAG(EXPRESS MAIL) UPS/DHL
DIRECTLY TO THE APPLICANT:

1/3 BILL OF LADING ORIGINAL ON BOARD TO THE ORDER OF
BRANCH TLEMEN EL KIFFANE NOTIFY APPLICANT MARKED FREIGHT

*** This advice to be continued next page

(A12WH1910

)

04. Edit Distance

사용 기술

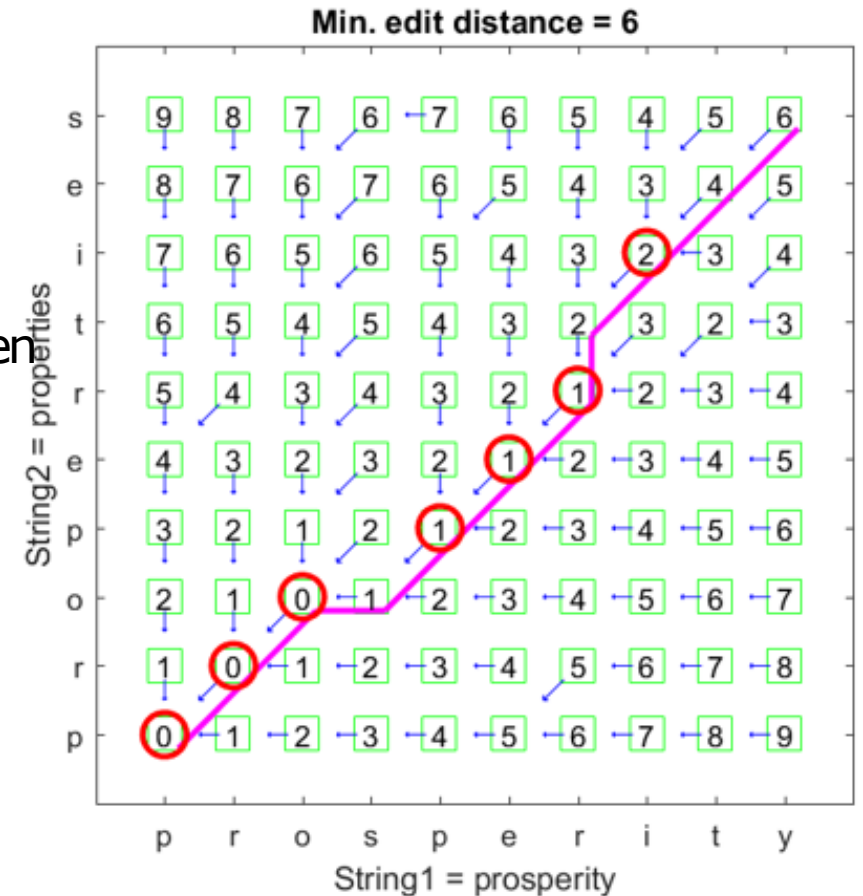
Edit Distance 알고리즘

- 두 문자열의 유사도를 판단하는 알고리즘
- 두 문자열이 같아지기 위해서 필요한 연산의 최소 횟수를 계산
- 시간 복잡도: $O(MN)$, 유사도 : $(\text{len}(\text{str1}) - \text{table}[\text{len2}][\text{len1}]) / \text{len}(\text{str1})$

EDISTANCE(s_1, s_2)

```
1  int m[i, j] = 0
2  for i ← 1 to |s1|
3  do m[i, 0] = i
4  for j ← 1 to |s2|
5  do m[0, j] = j
6  for i ← 1 to |s1|
7  do for j ← 1 to |s2|
8      do m[i, j] = min{m[i - 1, j - 1] + if (s1[i] = s2[j]) then 0 else 1,
9                      m[i - 1, j] + 1,
10                     m[i, j - 1] + 1}
11 return m[|s1|, |s2|]
```

< Pseudo Code of Edit Distance Algorithm >

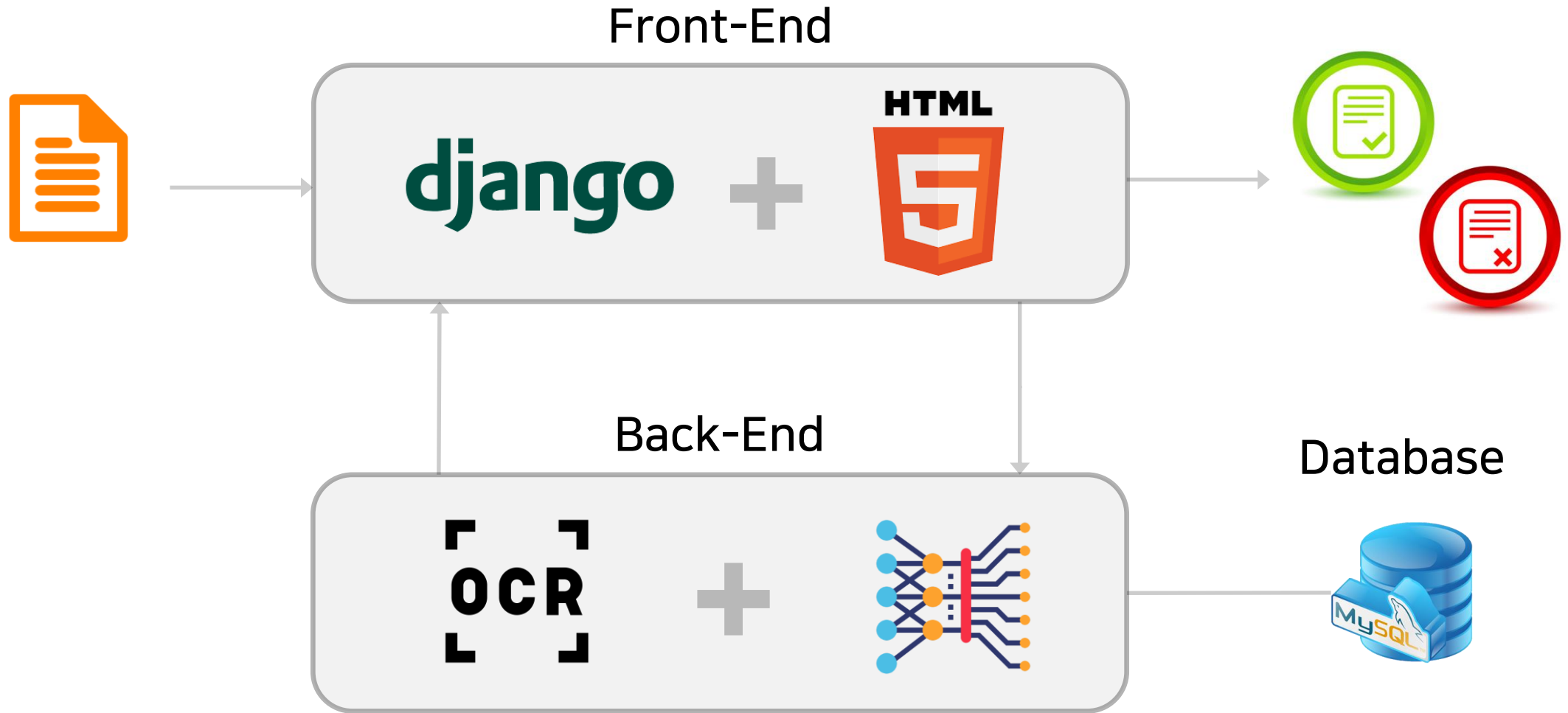


03

결과물

01. 시스템 구조

결과물

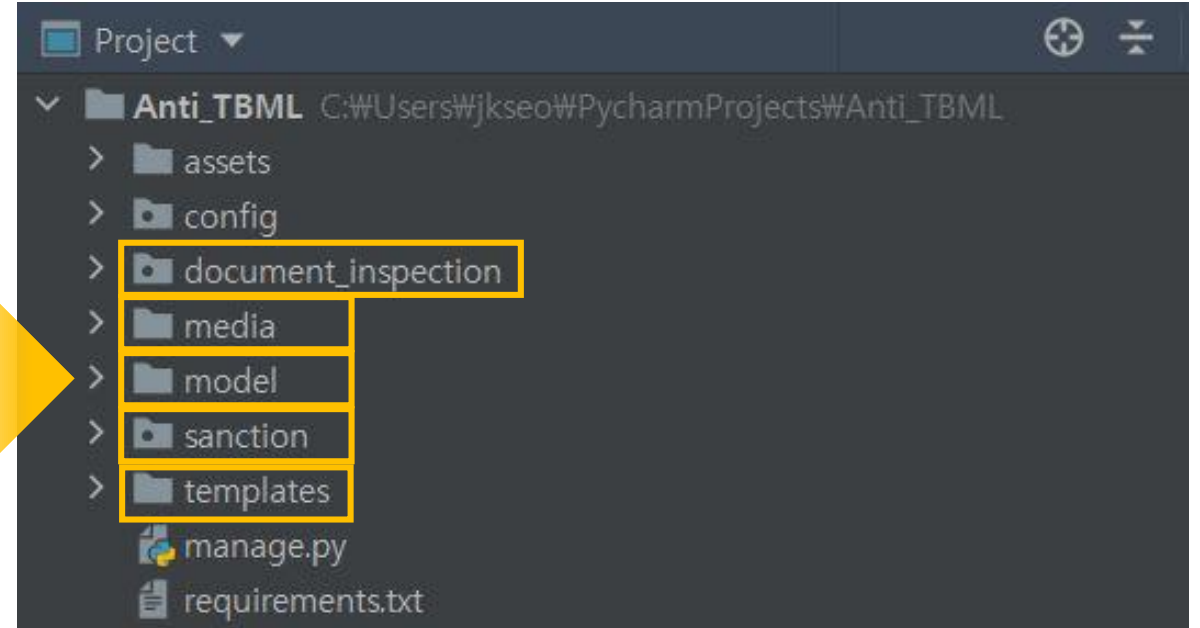


02. 상세 기능

결과물

프로젝트 핵심 기능

1. 로그인 및 회원가입을 통한 인증 기능
2. Sanction List에 대한 CRUD 및 검색 기능
3. 문서 스캔을 통한 유사도 측정 기능



< Django 프로젝트 구조 >

04

결론

01. 기대 효과

결론

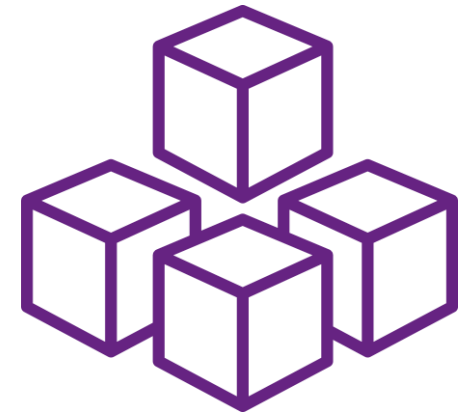
1. 금융 범죄 예방 효과

- 자금 세탁 및 관련 금융 범죄에 연루되어 있을 가능성이 있는 문서를 사전에 탐지함으로써 **경제적 손실 방지 가능**



2. 데이터 사용 편의성

- 웹에서 Sanction List에 대한 CRUD 작업이 클라우드 기반 DB와 실시간 연동되어 데이터 확장이 매우 용이하여 **데이터 관리 편의성 제고**



02. 개선 방안

결론

1. 알고리즘 시간 복잡도 개선

- 시간 복잡도가 $O(MN)$ 인 Edit Distance 알고리즘에 의존하기 때문에 Sanction List의 수가 늘어날 수록 유사도 계산 시간이 급격히 증가
⇒ 관련 알고리즘 개선 필요

```
def find_similar_word(word, sanc_list, thresh):  
    similar_word = {}  
    for sanc in sanc_list:  
        similarity = dist_sen(word, sanc, thresh)  
  
        if similarity > thresh:  
            similar_word[sanc] = similarity  
  
    return list(similar_word.keys()), list(similar_word.values())
```

2. 자연어 처리 개선

- 범용적인 단어를 Sanction List에 존재하는 단어로 인식하는 경우가 존재
⇒ 품사를 기준으로 자연어 처리 개선 필요



Thank you

ANTI-TBML

