

Trawler

사용자 매뉴얼

v1.0 2021.06

Team. 금은방

Trawler 사용자 매뉴얼

01

시작하기 ... 3p

1. 소개

사이트 주요 기능

2. 서버 사양

02

웹사이트 구성요소 ... 6p

1. 메인페이지 구성요소

2. 게시판 구성요소

탐지 현황

신고 현황

불법 토토

OCR

3. 통계페이지 구성요소

통계 설명

03

프로그램 소개 ... 16p

1. 테이블 정의서

2. 작동 과정

01

시작하기

01-1. 소개

Trawler 웹 사이트를 통해 성매매 알선 광고들에 대한 자료를 획득할 수 있습니다.

다양하게 수집된 광고들의 실태를 파악해보고 신고를 통해 직접 제재해보는 것을 통해 실생활에 얼마나 밀접하게 성매매가 접근해 있는지 알아봅시다.

사이트의 주요 기능



데이터 수집

SNS상에서
성매매 알선광고 수집



광고 유형 분류

수집한
성매매 알선광고들의
유형별 분류



워드 클라우드

분류된 광고와
수집한 광고의
주요 단어 시각화

01-2. 시작하기 : 서버 사양



서버

AWS 서버 T3.small

CPU

Intel(R) Xeon(R) Platinum 8259CL CPU @ 2.50GHz

RAM

2GB 메모리

운영체제

Ubuntu 18.04.5 LTS

02

웹 사이트 구성요소

02-1. 웹 사이트 구성요소 : 메인 페이지 구성요소



01. 실시간 현황

- 탐지 현황 : 하루동안 탐지한 수치
- 신고 현황 : 하루동안 신고된 수치
- 누적 신고 수 : 전체 신고 수치
- 누적 탐지 수 : 누적된 탐지 수치

02. 탐지 현황

매 크롤링시 탐지된 광고 수치를 기록, 그래프화

03. 실시간 신고 현황

실시간으로 최근 신고 현황을 볼 수 있는 현황판

04. 신고 현황

전체 탐지된 광고 중 신고 여부에 따라 비율을 표시

02-2. 웹 사이트 구성요소 : 게시판 구성요소 - 탐지 현황

Menu

Main

공지사항

실시간 현황

탐지 현황

신고 현황

불법토론

OCR

통계

logout

탐지 현황

Simple Datable

10 entries per page

Search...

NUM	ID	NICKNAME	PLATFORM	TIME	DATE
2824	S0VQ3NKYbu1nq7C	파워블예측 파워블게임주소 파워블분석방 파워블분석방법	Twitter	09:38:30	2021-06-11
2819	Lwdf8ZmyAqd7UFB	구미 경주 경산 대구 포항 성주 김천 영천 콜걸샵 출장샵 섹파	Twitter	09:13:15	2021-06-11
2818	ElzrB55WvETVkxx	서초 강남 송파 잠실 강동 천호동 길동 콜걸샵 출장샵 섹파	Twitter	09:13:16	2021-06-11
2817	QQU5lcfux9rdskH	건대 광진 성동 강북 수유 중랑 노원 도봉 콜걸샵 출장샵 섹파	Twitter	09:23:16	2021-06-11
2816	kaCSH8T8l6hpNin	과천 의왕 군포 안산 상록 단원 안양 만안 등안 시흥 콜걸샵 출장샵 섹파	Twitter	09:28:37	2021-06-11
2815	Namjeok1	Namheek	Twitter	09:28:27	2021-06-11
2814	fakkufakku8989	파쿠파쿠	Twitter	09:34:50	2021-06-11
2813	AeuLLwPKPFL5e5V	양산 기장 부산 울산 김해 동래 해운대 서면 콜걸샵 출장샵 섹파	Twitter	09:38:14	2021-06-11
2812	NQFeA9CReIdLQty	김포 부천 중동 광명 시흥 콜걸샵 출장샵 섹파	Twitter	09:48:10	2021-06-11
2811	alstjfkdrkxdl43	마포 용산 서대문 은평 종로 동대문 콜걸샵 출장샵 섹파	Twitter	09:48:53	2021-06-11

Showing 1 to 10 of 2719 entries

1 2 3 4 5 6 7 ... 272

Dashboard / 탐지 현황

03. ID

04. NICKNAME

05. PLATFORM

06. TIME

07. DATE

탐지된 계정의 ID

탐지된 계정의 닉네임

탐지된 플랫폼 ex) 트위터, 인스타그램, 페이스북 등등 기타 SNS명

탐지된 광고의 게시 날짜

탐지된 광고의 게시 시간

02-2. 웹 사이트 구성요소 : 게시판 구성요소 - 탐지 현황

탐지 현황

Dashboard / 신고현황 상세

[Twitter]8591bjm9813gq1

번호	2783	작성일	2021-06-11 18:37:37	플랫폼	Twitter
계정	역삼풀싸롱가격01059326647 유리방조이스 잠실룸싸롱 5692	검색 키워드	노래방 강남	유형	-
URL	https://twitter.com/8591bjm9813gq1				



역삼풀싸롱가격01059326647 유리방조이스 잠실룸싸롱 5692
@8591bjm9813gq1

#밤의전쟁트위터010X5932X6647 #북창동식노래방 010X5932X6647 #강남폭스010X5932X6647 성공의 겉모습 만큼 성공하는 것은 없다. Nothing succeeds like the appearance of success. June 11, 2021 at 06:31PM 밤의전쟁트위터 북창동식노래방 강남폭스

6:31 PM · Jun 11, 2021

Copy link to Tweet

이전글

다음글

jsouxus

greenenblue

광고 유형

신고 완료

01. 번호

02. 작성일

03. 플랫폼

04. 계정

05. 검색키워드

06. 유형

07. URL

08. 이전글과 다음글로 갈수있는 링크 보이는 항목은 계정ID

09. 유형선정 과정

탐지 현황의 게시물 번호

해당 광고가 탐지된 시간

해당 광고가 게시된 플랫폼

해당 광고를 게시한 계정 이름

해당 게시글을 탐지한 키워드

09.에서 선택한 광고 유형이 기입됨

해당 광고의 URL

2021 © TheGoldStore

Crafted with  by TheGoldStore

representative slammed the group's "tropes".

얼음
빙두
아이스

파워볼사이트추천

해피천사파워볼
Ro-300.com
추천인 8090
승인전화 없는
안전하고 편안한
미니게임 놀이터
TALK
hpy1004

6:38 PM · Jun 11, 2021

30 32 Copy link to Tweet

광고 유형
주점형
출장형
조건만약
마사지형
불법포로
오탁
광고 유형 ▾

Lwdf8ZmyAqd7UFB

다음글이 없습니다.

신고 완료

02. 광고 유형을 클릭 후 해당 광고 유형으로 판단되는 불법도토를 선택, 신고완료 버튼 클릭

02-2. 웹 사이트 구성요소 : 게시판 구성요소 - 유형선택

불법토토 현황

Dashboard / 불법토토 현황

Simple Datatable

10▼ entries per page

Search...

NUM	ID	NICKNAME	PLATFORM	PARTTEN	TIME	DAT
1597	S0VQ3NKYbu1nq7C	파워볼예측 파워볼게임주소 파워볼분석방 파워볼분석방법	Twitter	불법토토	09:38:30	2021-0
1596	bayronjgfyh1	은게임랜드drmc-999 .com	Twitter	불법토토	09:40:07	2021-0
1595	CZsOF6fVKys7Sp4	카톡판매	Twitter	불법토토	20:18:20	2021-0
1594	iO2QOmjuNG69s8	폰팅 060-900-0009 폰팅 좋아 폰섹 #폰팅 #350원060폰팅	Twitter	불법토토	10:16:41	2021-0
1593	Arr_2120	남서우	Twitter	불법토토	09:57:43	2021-0
1532	uLZsZoa2mEmZBbL	심바먹튀검증 심바추천인 심바가입코드	Twitter	불법토토	03:41:04	2021-0
1232	uLZsZoa2mEmZBbL	심바먹튀검증 심바추천인 심바가입코드	Twitter	불법토토	03:41:04	2021-0
947	uLZsZoa2mEmZBbL	심바먹튀검증 심바추천인 심바가입코드	Twitter	불법토토	03:41:04	2021-0
537	uLZsZoa2mEmZBbL	심바먹튀검증 심바추천인 심바가입코드	Twitter	불법토토	03:41:04	2021-0
288	upline80	그래프게임 사이트	Twitter	불법토토	07:48:41	2021-0

Showing 1 to 10 of 11 entries

유형선택한 불법 토토가 해당 항목에 입력되는 것을 확인

[Twitter]S0VQ3NKYbu1nq7C					
번호	1597	작성일	2021-06-11 18:37:37	플랫폼	Twitter
계정	파워볼예측 파워볼게임주소 파워볼분석방 파워볼분석방법	검색 키워드	추천 사이트	유형	불법토토
URL	https://twitter.com/S0VQ3NKYbu1nq7C				

02-2. 웹 사이트 구성요소 : 게시판 구성요소 - 신고 현황

01. 한 페이지에서 볼 수 있는 게시물 개수를 지정하는 항목

02. 검색을 할 수 있는 항목

신고 현황

Simple Datatable

10 entries per page

Search...

NUM	ID	NICKNAME	PLATFORM	PARTTEN	TIME	DATE
1	gxxjbtueo25	오피가자 opgaja	Twitter	조건만남형	09:29:09	2021-06-11
2	7625oadm9486jh1	강남식스 01031441175	Twitter	주점형	09:31:03	2021-06-11
3	misokim00065	ㅎㅏㅓㅓ 팔당 구르ㅣ 양평 ㅏㅓㅓ 평 남양ㅓㅓ 콜걸샵 출장샵 섹ㅓㅓ	Twitter	출장형	09:09:11	2021-06-11

Showing 1 to 3 of 3 entries

2021 © TheGoldStore

Crafted with by TheGoldStore

03. ID

04. NICKNAME

05. PLATFORM

06. TIME

07. DATE

08. PARTTEN
- 탐지된 계정의 ID

탐지된 계정의 닉네임

탐지된 플랫폼 ex) 트위터, 인스타그램, 페이스북 등등 기타 SNS명

탐지된 광고의 게시 날짜

탐지된 광고의 게시 시간

탐지현황에서 설정된 유형

02-2. 웹 사이트 구성요소 : 게시판 구성요소 - 신고 현황

신고 현황

Dashboard / 신고현황 상세

[Twitter]gzxjbtueo25

번호	3806	작성일	2021-06-11 18:37:37	플랫폼	Twitter
계정	오피가자 opgaja	검색 키워드	추천 오피	유형	조건만남형
URL	https://twitter.com/gzxjbtueo25				

오피가자 opgaja
@gzxjbtueo25

오피가자 추천정보 네이버뷰 http://naverbam.com 네이버밤 오피가자주소 오피가자트위터 오피가자 추천정보 네이버뷰 http://naverbam.com 네이버밤 오피가자주소 오피가자트위터 932

6:29 PM · Jun 11, 2021

Copy link to Tweet

이전글	7625oadm9486jh1
다음글	다음글이 없습니다.

01. 번호
02. 작성일
03. 플랫폼
04. 계정
05. 검색 키워드
06. 유형
07. URL

탐지 현황의 게시글 번호
해당 광고가 탐지된 시간
해당 광고가 게시된 플랫폼
해당 광고를 게시한 계정 이름
탐지하는 키워드
선택된 유형이 입력되는 곳
해당 광고가 표시된 URL

08. 이전 글과 다음 글로 갈수 있는 링크
보이는 항목은 계정ID

2021 © TheGoldStore

Crafted with by TheGoldStore

02-2. 웹 사이트 구성요소 : 게시판 구성요소 - 불법 토토

불법토토 현황

01. 한 페이지에 볼 수 있는 게시물 개수를 지정하는 항목
02. 검색을 할 수 있는 항목

Simple Datatable

10 ▾ entries per page

Search...

NUM	ID	NICKNAME	PLATFORM	PARTTEN	TIME	DATE
1	bayronjgfyh1	온게임랜드drmc-999 .com	Twitter	불법토토	09:40:07	2021-06-11
2	CZsOF6fVKys7Sp4	카톡판매	Twitter	불법토토	20:18:20	2021-06-07
3	iO2QOrnjuNG69s8	폰팅 060-900-0009 폰팅 좋아 폰섹 #폰팅 #350원060폰팅	Twitter	불법토토	10:16:41	2021-06-09
4	Arr_2120	남서우	Twitter	불법토토	09:57:43	2021-06-10
5	uLZsZoa2mEmZBbL	심바먹튀검증 심바추천인 심바가입코드	Twitter	불법토토	03:41:04	2021-06-08
6	uLZsZoa2mEmZBbL	심바먹튀검증 심바추천인 심바가입코드	Twitter	불법토토	03:41:04	2021-06-08
7	uLZsZoa2mEmZBbL	심바먹튀검증 심바추천인 심바가입코드	Twitter	불법토토	03:41:04	2021-06-08
8	uLZsZoa2mEmZBbL	심바먹튀검증 심바추천인 심바가입코드	Twitter	불법토토	03:41:04	2021-06-08
9	upline80	그래프게임 사이트	Twitter	불법토토	07:48:41	2021-06-09
10	uLZsZoa2mEmZBbL	심바먹튀검증 심바추천인 심바가입코드	Twitter	불법토토	03:41:04	2021-06-08

Showing 1 to 10 of 10 entries

03. ID

04. NICKNAME

05. PLATFORM

06. TIME

07. DATE

08. PARTTEN
- 탐지된 계정의 ID

탐지된 계정의 닉네임

탐지된 플랫폼 ex) 트위터, 인스타그램, 페이스북 등등 기타 SNS명

탐지된 광고의 게시 날짜

탐지된 광고의 게시 시간


탐지현황에서 설정된 유형

02-2. 웹 사이트 구성요소 : 게시판 구성요소 - 불법 토토

불법토토 현황

Dashboard / 불법토토 상세

[Twitter]bayronjgfyh1					
번호	1596	작성일	2021-06-11 18:37:37	플랫폼	Twitter
계정	온게임랜드drmc-999 .com	검색 키워드	추천 사이트	유형	불법토토
URL	https://twitter.com/bayronjgfyh1				



온게임랜드drmc-999 .com

@bayronjgfyh1

#마이다스카지노

#호텔카지노

#슬레이카지노

#폰배팅카지노

drmc-999 .com

추천코드:vvv1

카톡:younme5577

안전하고 믿을수있는 사이트만 제공합니다.

#온게임랜드 에서 여러분께 100% 보증하는곳!!!

6:40 PM · Jun 11, 2021

Copy link to Tweet

이전글

CZsOF6fVKys7Sp4

다음글

다음글이 없습니다.

01. 번호

02. 작성일

03. 플랫폼

04. 계정

05. 검색 키워드

06. 유형

07. URL
- 탐지 현황의 게시물 번호

해당 광고가 탐지된 시간

해당 광고가 게시된 플랫폼

해당 광고를 게시한 계정 이름

탐지하는 키워드

선정된 유형이 입력되는 곳

해당 광고가 표시된 URL

08. 이전 글과 다음 글로 갈수 있는 링크 보이는 항목은 계정ID

02-2. 웹 사이트 구성요소 : 게시판 구성요소 - OCR

OCR탐지 현황

Simple Datatable

10 ▾ entries per page

Search...

NUM	ID	NICKNAME	PLATFORM	TIME	DATE
409	dpfls0014	홍예린	Twitter	21:16:01	2021-06-05
408	hzilqYFi4VfJuNk	박재준	Twitter	14:24:26	2021-06-05
407	DogeAdore	adoreDoge	Twitter	22:06:49	2021-06-10
406	hrBWTQeEASZV2jU	김현수	Twitter	07:10:39	2021-06-12
405	uesdioro	구글지도광고 바이럴마케팅광고 업체마케팅 홈페이지제작업체	Twitter	22:40:48	2021-06-10
404	dcfwcJnz	네이버지도마케팅 백링크광고 온라인광고종류 홍보아이디어	Twitter	07:38:30	2021-06-12
403	woo_sungmin	Sungmin WOO	Twitter	05:51:03	2021-06-12
402	hyhh98284R0Xuqj	김용재	Twitter	09:06:36	2021-06-12
401	vEde9OZMRjeUD	하이ㅎㅎ	Twitter	05:53:40	2021-06-12
400	cooky95111	채어	Twitter	08:55:47	2021-06-12

Showing 1 to 10 of 409 entries

01. 한 페이지에 볼 수 있는 게시물 개수를 지정하는 항목

02. 검색을 할 수 있는 항목

03. ID

04. NICKNAME

05. PLATFORM

06. TIME

07. DATE

08. PARTTEN

탐지된 계정의 ID

탐지된 계정의 닉네임

탐지된 플랫폼 ex) 트위터, 인스타그램, 페이스북 등등 기타 SNS명

탐지된 광고의 게시 날짜

탐지된 광고의 게시 시간

탐지현황에서 설정된 유형

02-2. 웹 사이트 구성요소 : 게시판 구성요소 - OCR

[Twitter]hziJqYFi4VfJuNk					
번호	408	작성일	2021-06-12 18:45:58	플랫폼	Twitter
계정	박재준	검색 키워드	금전 코로나	유형	-
URL	https://twitter.com/hziJqYFi4VfJuNk				



- 01. 번호
- 02. 작성일
- 03. 플랫폼
- 04. 계정
- 05. 검색 키워드
- 06. 유형
- 07. URL

탐지 현황의 게시글 번호
해당 광고가 탐지된 시간
해당 광고가 게시된 플랫폼
해당 광고를 게시한 계정 이름
탐지하는 키워드
선정된 유형이 입력되는 곳
해당 광고가 표시된 URL

이전글	DogeAdore
다음글	dpfls0014

- 08. 이전 글과 다음 글로 갈수 있는 링크 보이는 항목은 계정ID
- 09. 10p의 유형선정 과정과 동일

02. 웹 사이트 구성요소 : 통계페이지 구성요소

01. 유형별 누적 검출 현황

탐지된 광고 유형들을 그래프로 표현한 것

02. 누적 신고통계

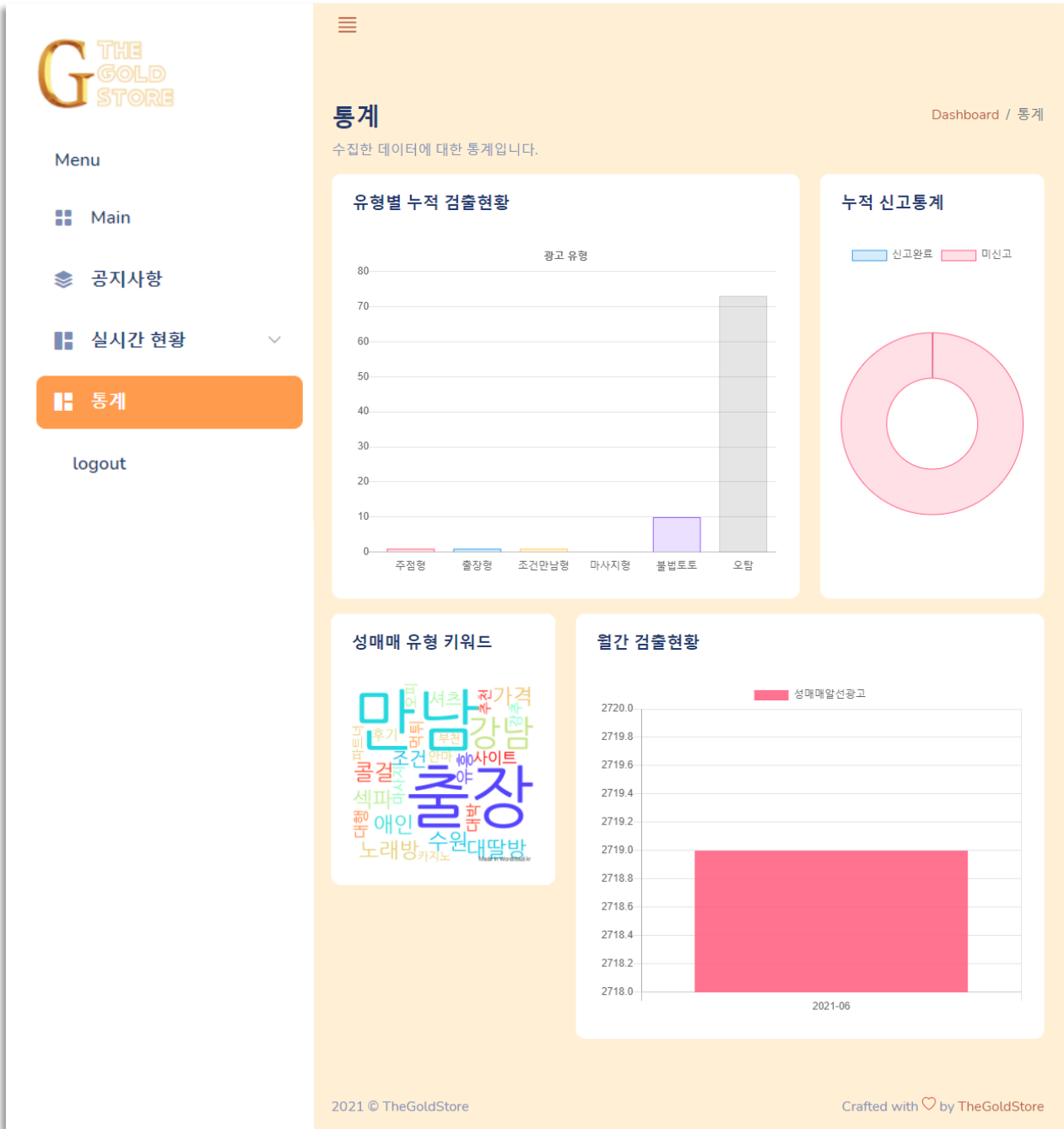
탐지된 성매매 알선 광고의 총 탐지 건수 및 신고 건수 통계 그래프

03. 월간 검출현황

탐지된 전체 성매매 알선광고를 월별로 그래프를 통해 표현

04. 성매매 유형 키워드

탐지된 성매매 알선 광고에서 빈도수가 높은 키워드를 찾아 시각화



03

프로그램 소개

03-1. 프로그램 소개 : 테이블 정의서

01. data_country

테이블 명		data_country			테이블 설명		도시 이름	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	number	번호	int	11	NotNull			AUTO_INCREMENT
2	country_name	도시 이름	varchar	45	NotNull			

02. data_normal

테이블 명		data_normal			테이블 설명		일반적인 키워드(OCR 이용)	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	number	번호	int	11	NotNull			
2	main_set	메인 키워드	varchar	45	NotNull			
3	sub_set	서브 키워드	varchar	45	NotNull			

03. data_set

테이블 명		data_set			테이블 설명		크롤링 키워드	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	number	번호	int	11	NotNull			AUTO_INCREMENT
2	main_set	메인 키워드	varchar	45	NotNull			
3	sub_set	서브 키워드	varchar	45	NotNull			

04. data_stopword

테이블 명		data_country			테이블 설명		명사 추출을 위한 필터링 용어	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	number	번호	int	11	NotNull			
2	stopword	불용어	varchar	45	NotNull			

05. data_strange

테이블 명		data_strange			테이블 설명		탐지 우회 단어	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	number	번호	int	30	NotNull			
2	words	우회 단어	varchar	30	NotNull			

06. data_type

테이블 명		data_type			테이블 설명		광고 유형별 검출 현황	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	tpk	번호	int	30	NotNull			AUTO_INCREMENT
2	pattern	광고 유형	char	50	NotNull			
3	stats	탐지수	int	100	NotNull			

07. OCR

테이블 명		ocr			테이블 설명		OCR 결과	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	num	번호	int	11	NotNull			
2	id	아이디	varchar	100	NotNull			
3	nickname	닉네임	varchar	80	NotNull			
4	date	날짜	date		NotNull			
5	time	시간	time		NotNull			
6	url	url	varchar	100	NotNull			
7	search_word	검색 단어	varchar	100	NotNull			
8	snstype	플랫폼	varchar	45	NotNull			
9	modify_time	크롤링시간	timestamp		NotNull		current_time stamp()	ON UPDATE CURRENT_TIME STAMP()
10	media_url	미디어링크	varchar	100	NotNull			
11	ocr_text	ocr텍스트	varchar		NotNull			

08. report

테이블 명		report			테이블 설명		신고 게시판	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	num	번호	int	11	NotNull			
2	id	아이디	varchar	100	NotNull			
3	nickname	닉네임	varchar	80	NotNull			
4	contents	내용	text		NotNull			
5	date	날짜	date		NotNull			
6	time	시간	time		NotNull			
7	url	url	varchar	100	NotNull			
8	search_word	검색 단어	varchar	100	NotNull			
9	ko_text	허수 값	text					
10	modify_time	크롤링시간	timestamp		NotNull		current_time stamp()	ON UPDATE CURRENT_TIME STAMP()
11	snstype	플랫폼	varchar	45	NotNull			
12	pattern	광고유형	char	50				
13	timetest		timestamp					
14	id_str	Id 값	varchar	45				
15	media_url	사진 url	text		NotNull			24

테이블 명		toto			테이블 설명		토토 게시판	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	num	번호	int	11	NotNull			
2	id	아이디	varchar	100	NotNull			
3	nickname	닉네임	varchar	80	NotNull			
4	contents	내용	text		NotNull			
5	date	날짜	date		NotNull			
6	time	시간	time		NotNull			
7	url	url	varchar	100	NotNull			
8	search_word	검색 단어	varchar	100	NotNull			
9	ko_text	허수 값	text					
10	modify_time	크롤링시간	timestamp		NotNull		current_time stamp()	ON UPDATE CURRENT_TIME STAMP()
11	snstype	플랫폼	varchar	45	NotNull			
12	pattern	광고유형	char	50				
13	timetest		timestamp					
14	id_str	Id 값	varchar	45				
15	media_url	사진 url	text		NotNull			25

10. twitter

테이블 명		twitter			테이블 설명		트위터 크롤링 데이터	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	num	번호	int	11	NotNull			
2	id	아이디	varchar	100	NotNull			
3	nickname	닉네임	varchar	80	NotNull			
4	contents	내용	text		NotNull			
5	date	날짜	date		NotNull			
6	time	시간	time		NotNull			
7	url	url	varchar	100	NotNull			
8	search_word	검색 단어	varchar	100	NotNull			
9	ko_text	허수 값	text					
10	modify_time	크롤링시간	timestamp		NotNull		current_time stamp()	ON UPDATE CURRENT_TIME STAMP()
11	snstype	플랫폼	varchar	45	NotNull			
12	pattern	광고유형	char	50				
13	timetest		timestamp					
14	id_str	Id 값	varchar	45				
15	media_url	사진 url	text		NotNull			26

11. users

테이블 명		users			테이블 설명		사용자 정보	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	Uid	번호	int	11	notNull			AUTO_INCREMENT
2	user_id	아이디	varchar	50	notNull			
3	user_name	이름	varchar	50	notNull			
4	password	패스워드	varchar	255	notNull			
5	admin_level	관리자 레벨	int				0	

12. words_cloud

테이블 명		Words_cloud			테이블 설명		워드 클라우드	
NO	칼럼 ID	칼럼 명	타입	길이	NULL	Key	default	비고
1	words	단어	varchar	30	notNull			AUTO_INCREMENT
2	words_size	빈도 수	int	11	Null			