|  |
| --- |
| **1팀 최종 보고서** |

****

|  |
| --- |
| 문서번호 : 2021Report.Final |
| 최종 작성일 : 2021년 06월 12일 |

**< 목 차 >**

**1. 금은방 프로젝트 개요3**

**1.1 프로젝트 명 및 기간3**

**1.2 프로젝트 배경3**

(1) 선정 이유3

(2) 현 실태 5

**1.3 프로젝트 목적6**

**1.2 프로젝트 기대 효과6**

**2. 프로젝트 조직7**

**2.1 프로젝트 조직도7**

**2.2 프로젝트 책임 및 역할7**

**3. 프로젝트 수행 일정(WBS)8**

**3.1 프로젝트 추진 일정8**

**3.2 프로젝트 세부 일정8**

**4. 프로젝트 소개9**

**4.1 인터뷰9**

**4.2 데이터 수집 및 알고리즘11**

(1) 데이터 크롤링11

(2) 워드 클라우드 13

(3) 데이터 셋 최신화13

**4.3 서버14**

(1) 홈페이지 구축14

(2) 유해광고 실태 현황 분석 18

**5. 성과18**

**5.1 정부 기관 및 공공기관 협업18**

(1) 프로젝트 산출물18

(2) 다시함께 상담센터18

**6. 향후 활용 방향성19**

**6.1 정부 기관 및 공공기관 협업19**

(1) 사이버 경찰청19

(2) 방송통신심의위원회19

(3) 기금조성총괄본부19

**1. 금은방 프로젝트 개요**

**1.1 프로젝트 명 및 기간**

▷ 프로젝트 명: 불법 성매매 광고 수집 및 분석 통계

▷ 프로젝트 코드: 2021-GoldStore-01

▷ 프로젝트 기간: 2021.04.01 ~ 2021.06.14 (10주)

▷ 프로젝트 수행 팀: 금은방팀

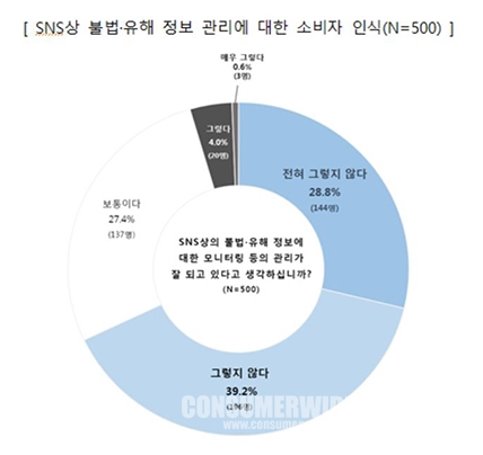
* 1. **프로젝트 배경**

1. 선정 이유

* 인터넷 이용자의 상당 수가 소셜 네트워크 서비스의 불법, 유해 광고에 노출되어있지만 불법 유해광고 모니터링 등의 관리가 미비한 상황

테이블이(가) 표시된 사진

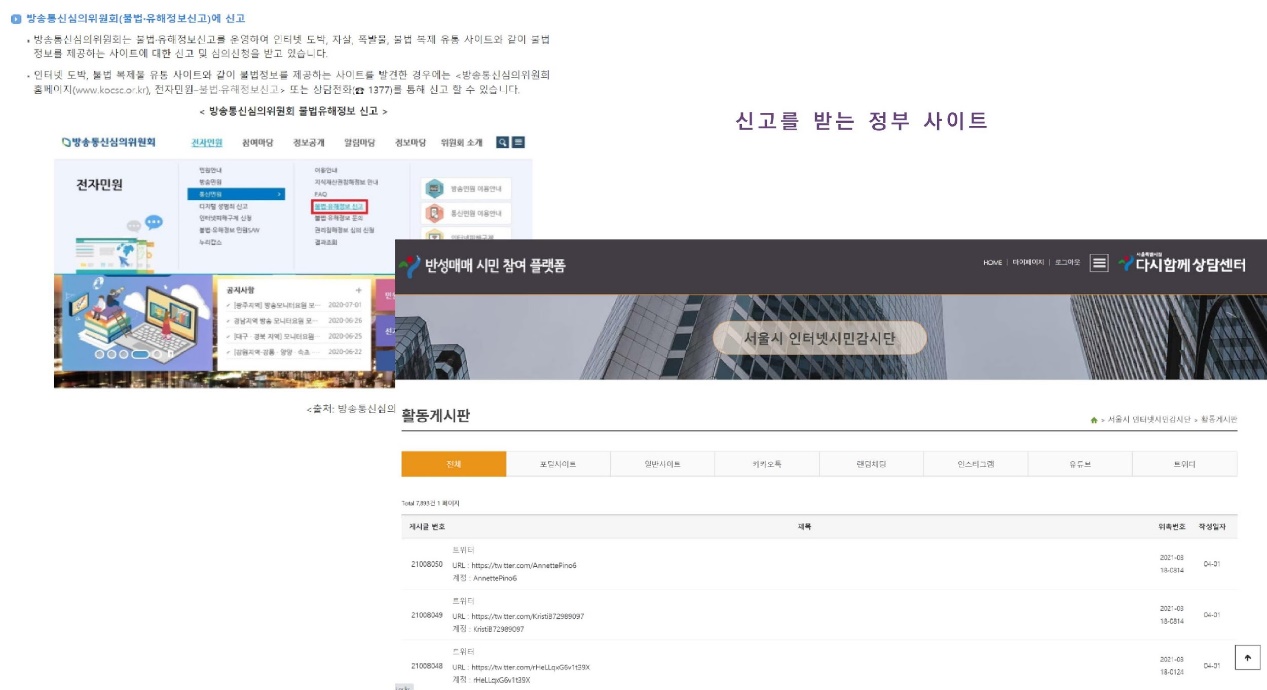
자동 생성된 설명



* 현재 이러한 불법 유해 광고 관리를 위해 매년 1000명의 인터넷 시민 감시단을 모집하고 있으며, 모든 관리는 시민들의 **수작업**을 통해 이루어지고 있는 상태-

텍스트이(가) 표시된 사진

자동 생성된 설명

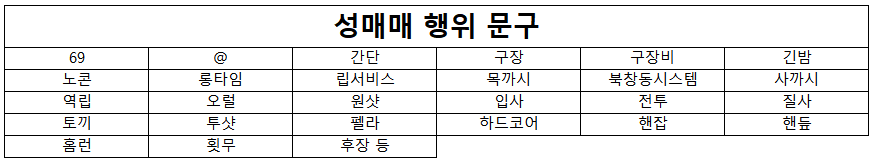
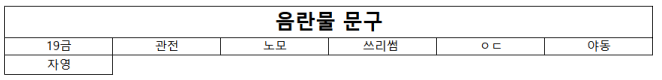


* 모든 과정이 수동으로 진행되는 현 불법 유해 광고 모니터링은 **인력과 시간의 측면**에서 **비효율적**
* 따라서 자동 불법 광고 탐지 프로그램을 만들어 효율적인 불법 유해 광고 모니터링을 실현하고, 사용자의 **안전한 인터넷 환경 조성**

**현 실태**

* 블로그나 SNS 등 온라인상에서 청소년을 대상으로 하는 불법 성매매 광고가 폭발적으로 증가
* 텍스트이(가) 표시된 사진

  자동 생성된 설명페이스북이나 인스타그램 등 청소년들이 자주 이용하는 온라인 SNS 매체를 통해 성매매 정보에 쉽게 노출 가능

테이블이(가) 표시된 사진

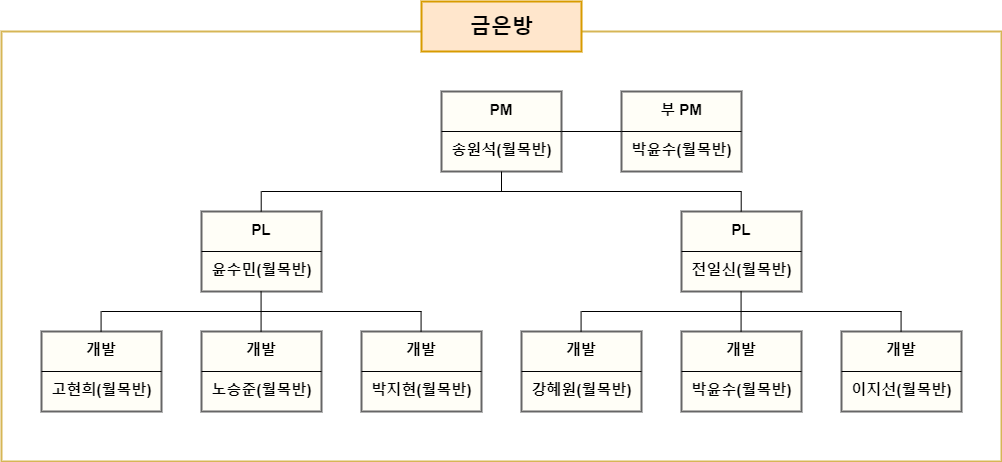
자동 생성된 설명

* 특정 키워드를 통해 쉽게 접근이 가능하며, 매크로를 이용해 지속적으로 불법유해 광고를 SNS에 노출

* 1. **프로젝트 목적**
* SNS를 통해 퍼져 나가는 **불법 성매매 광고**를 **자동으로 수집 및 분석**하여 모니터링의 효율성 증대
* 정보통신망법 제44조의7 (불법정보의 유통금지), 제 50조의8 (불법행위를 위한 광고성 정보 전송금지) 법률에 의거하여 소셜 미디어에서 퍼져 나가고 있는 유해광고를 수집하고 분석
* 나아가 수집된 정보들을 분석한 결과를 통계치로 작성하여 현 유해광고의 실태 현황을 분석할 수 있는 자료 제작
  1. **프로젝트 기대 효과**
* 기존의 수작업 방식에서 벗어나 불법 유해 광고를 자동으로 수집 및 분석하는 프로그램을 제작하여 인력과 비용을 절감
* 불법 유해 광고 자동 수집 및 분석 프로그램을 통해 빅데이터를 수집하고 다양한 연구에 활용
* 서울시 인터넷 시민 감시단 등 국가 유해 정보 감시 기관과 협업하여 건전한 온라인 환경 조성

**2. 프로젝트 조직**

**2.1 프로젝트 조직도**



**[그림2-1] 프로젝트 조직도**

**2.2 프로젝트 책임 및 역할**

|  |  |  |
| --- | --- | --- |
| **책임** | **역할** | **담당자명** |
| PM | 프로젝트 총괄 책임자 | 송원석 |
| PL | 서버 개발 관리 | 윤수민 |
| 데이터 셋 개발 관리 | 전일신 |
| 개발\_1팀 | Developer | 이지선 |
| Developer | 박윤수 |
| Developer | 강혜원 |
| 개발\_2팀 | Developer | 고현희 |
| Developer | 박지현 |
| Developer | 노승준 |

**[표2-1] 책임 및 역할**

**3. 프로젝트 수행 일정**

**3.1 프로젝트 추진 일정**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **프로젝트 일정** | | **2021.04.01~ 2021.06.18 (10주)** | | | | | | | | | |
| **단계** | | **4월** | | | | **5월** | | | | **6월** | |
| **1주차** | **2주차** | **3주차** | **4주차** | **1주차** | **2주차** | **3주차** | **4주차** | **1주차** | **2주차** |
| 프로젝트 계획 및 설계 | |  |  |  |  |  |  |  |  |  |  |
| 개발 | 자료 수집 및 학습 |  |  |  |  |  |  |  |  |  |  |
| 웹 서버 구축 |  |  |  |  |  |  |  |  |  |  |
| 웹 사이트 개발 |  |  |  |  |  |  |  |  |  |  |
| 데이터 수집 |  |  |  |  |  |  |  |  |  |  |
| 데이터 시각화 |  |  |  |  |  |  |  |  |  |  |
| 보고 |  |  |  |  |  |  |  |  |  |  |
| 테스트 | 단위테스트 |  |  |  |  |  |  |  |  |  |  |
| 통합테스트 |  |  |  |  |  |  |  |  |  |  |
| 최종테스트 |  |  |  |  |  |  |  |  |  |  |
| 발표 | 최종발표준비 |  |  |  |  |  |  |  |  |  |  |

**[표4-1] 프로젝트 추진 일정**

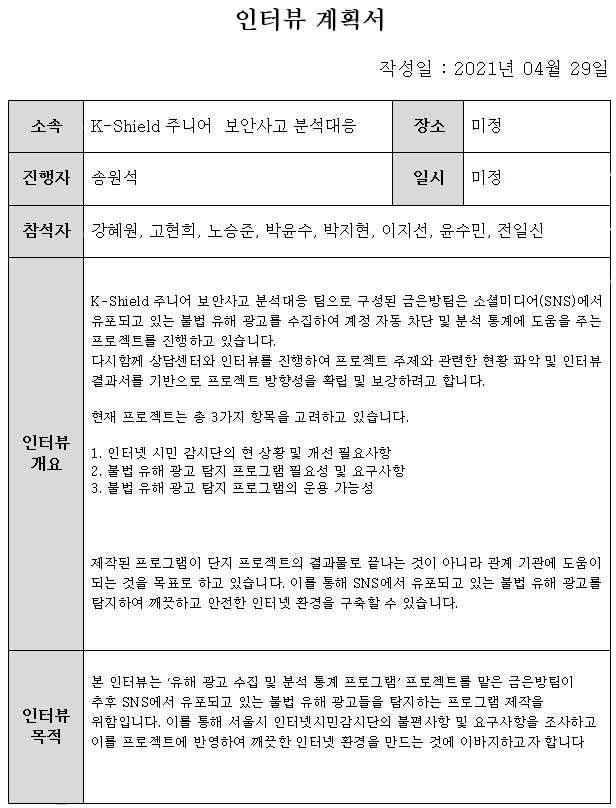
**3.2 프로젝트 세부 일정**

|  |  |  |
| --- | --- | --- |
| **일정** | **단계** | **작업** |
| 2021.04.01 ~ 2021.04.05 | **프로젝트계획** | 프로젝트 범위 확정  프로젝트 일정 확정  프로젝트 진행 방향 확정 |
| 2021.04.05 ~  2021.05.22 | **개발** | 자료 수집 및 관련 공부 학습  서버 구축  유해 광고 데이터 수집  유해 단어셋 시각화  웹 사이트 개발  보고 |
|
| 2021.04.26 ~  2021.05.31 | **테스트 / 이행** | 단위테스트  통합테스트  최종테스트 |
| 2021.06.07 ~  2021.06.17 | **마무리** | 사용자 매뉴얼 작성  최종발표 준비 |

**[표4-2] 단계별 세부 일정**

**4. 프로젝트 소개  
4.1 인터뷰**

(1) 인터뷰 계획서

* 다시함께 상담센터는 한국 YMCA전국연맹이 서울시로부터 성매매피해자의 인권을 보호하고, 성매매를 예방, 축소, 감시하기 위한 활동을 하고 있으며 성매매 신고 감시단을 운영
* 불법 성매매 광고를 근절 시키기 위해 금은방 팀은 다시함께 상담센터와 함께 SNS에서 노출되고 있는 성매매 관련 불법 광고 탐지 및 차단프로젝트 계획, 이를 위해 사전 인터뷰를 진행

**(2) 1차 인터뷰**

* + 일시 : 2021. 05. 14 14:00-15:30
  + 참여자 : 박윤수, 이지선
  + 요구사항  
    1) 글은 성매매 알선 키워드와 관련 없으나 첨부된 이미지에 광고가 있는 경우, 탐지 기능  
    2) 트위터 자체 신고 기능  
    3) 게시글에 알선 포털사이트가 업로드되어 있는 경우, 계정 신고 뿐만 아니라 알선 사이트 신고 기능  
    4) 신고 후, 처리 결과 확인 기능 (계정 삭제, 아직 유통 중 등)  
    5) 홈페이지 디자인이 당근마켓처럼 귀여운 디자인이면 좋을 것  
    6) 실시간 탐지보다는 실시간 신고현황이 더 잘 드러나면 좋을 것  
    7) 불법광고는 성매매 알선광고, 깨끗한 인터넷 환경이 아닌 성 평등한 인터넷 환경으로 어휘 변경
  + 현재 기획중인 프로그램이 완성될 경우, 인터넷시민감시단에서 활용할 의향 있음
  + 현재 인터넷시민감시단이 수집한 데이터 분석은 SNS 별 건수와 광고 유형 파악에만 활용 중
  + 가장 중요한 것은 트위터 자체 신고



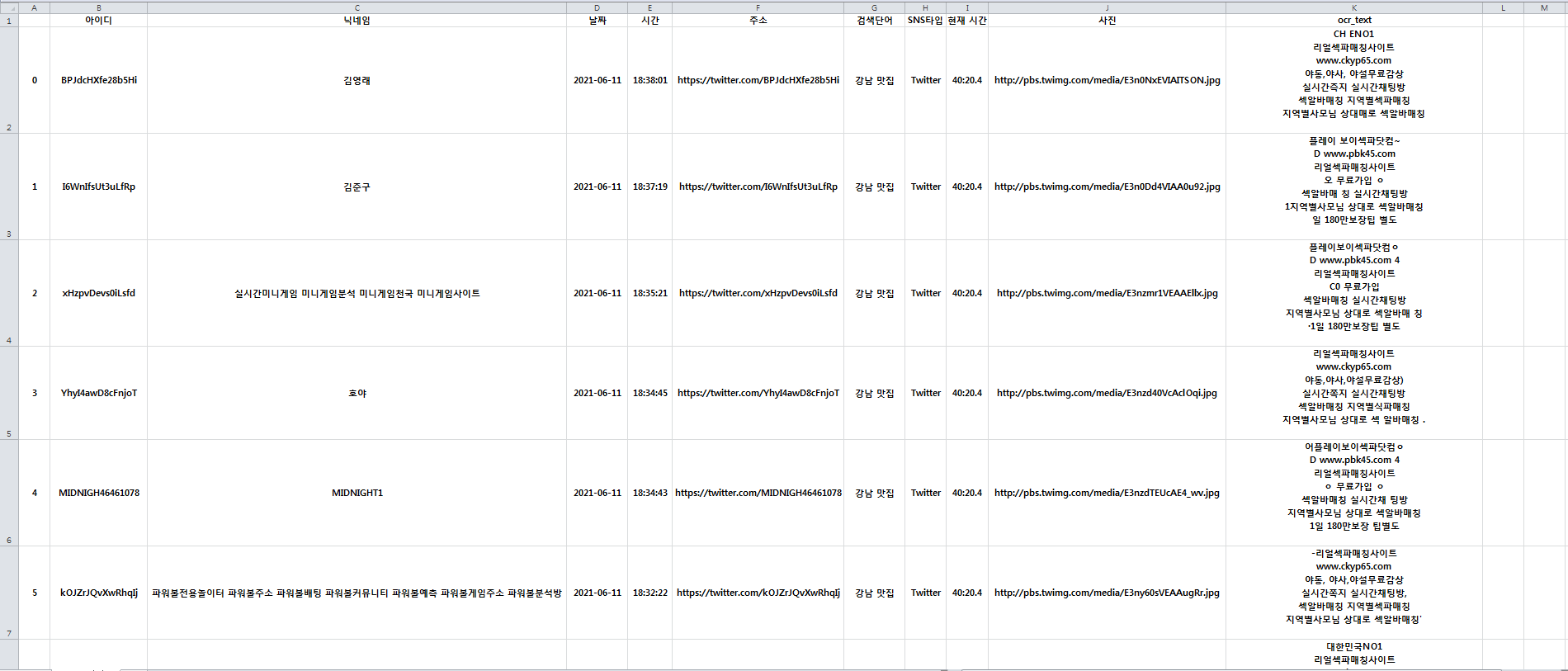


**(3) 2차 인터뷰**

* 일시 : 2021. 06. 09 14:00-15:00
* 참여자 : 고현희, 이지선
* 요구사항  
  1) 로그인한 사용자만 관리할 수 있게 접근권한 부여  
  2) 왜 신고됐는지 증거가 필요하기 때문에 게시판에 신고한 캡처본 업로드 기능  
  3) 계정 URL에 접속하지 않아도 계정 삭제가 됐는지(유통이 되고 있는지) 확인 기능
* 본 프로젝트의 프로그램을 사용할 의사는 논의가 더 필요함
* 관련 증빙서류는 상부와 논의 후, 추후 안내 예정
  1. **데이터 수집 및 알고리즘**

**(1) 데이터 크롤링**

* 데이터 크롤링을 통해 불법 유해 광고물들이 SNS에 얼마나 많이 유포가 되고 있는지를 데이터화
* Tweepy API를 사용하여 트위터에 존재하는 불법 성매매 광고물의 ID, 내용, 날짜 등의 수집이 가능하고 이를 문서화하여 다양한 자료로서 사용 가능
* Google OCR API를 사용하여 사진에 담겨있는 각종 불법 성매매 광고 단어들을 탐지 및 수집하여 이를 문서화 후 다양한 자료로 사용 가능
* mysql을 사용하여 미리 키워드들을 저장함으로써 데이터 크롤링을 할 때 키워드들이 중복 검색되지 않도록 하며 저장된 키워드와 관련된 자료들을 탐지 및 분석한다.
* 이를 통해 아이디, 닉네임, 날짜, 시간, 사진URL 등의 내용들을 문서화 할 수 있다.



[데이터 크롤링 후 문서화된 내용]

**(2) 워드 클라우드**

* 분석에 용이하게 하기 위해 수집한 데이터를 통하여 불법 성매매 광고에서 사용되고 있는 단어들을 시각화
* 이를 통해 최근 대두되고 있는 불법 성매매 광고 단어가 무엇인지 파악하고 새롭게 등장한 단어들을 확인 가능

텍스트이(가) 표시된 사진

자동 생성된 설명

**(3) 데이터 셋 최신화**

* 워드 클라우드를 통해 분석된 단어들을 정리하고 수집하여 기존에 없었던 새로운 불법 성매매 광고 단어를 데이터 크롤링 DB 셋에 추가시켜 자동으로 데이터 최신화
* 이를 통해 시시각각 변하고 있는 불법 광고 현황에 대응 가능
* 한글처럼 쓰이는 글자들을 데이터 크롤링 DB 셋에 추가시켜 다양한 데이터 수집 가능
* 이를 통해 시시각각 변하고 있는 불법 광고 현황에 대응 가능

**4.3 서버**

**(1) 홈페이지 구축**

* 앞서 수집한 불법 광고 정보를 시민 감시단에게 빠르게 제공하기 위해 홈페이지를 구축
* 이를 통해 불법 광고 정보가 제공되는 게시판을 운용과 빠른 정보탐색 및 신고 가능

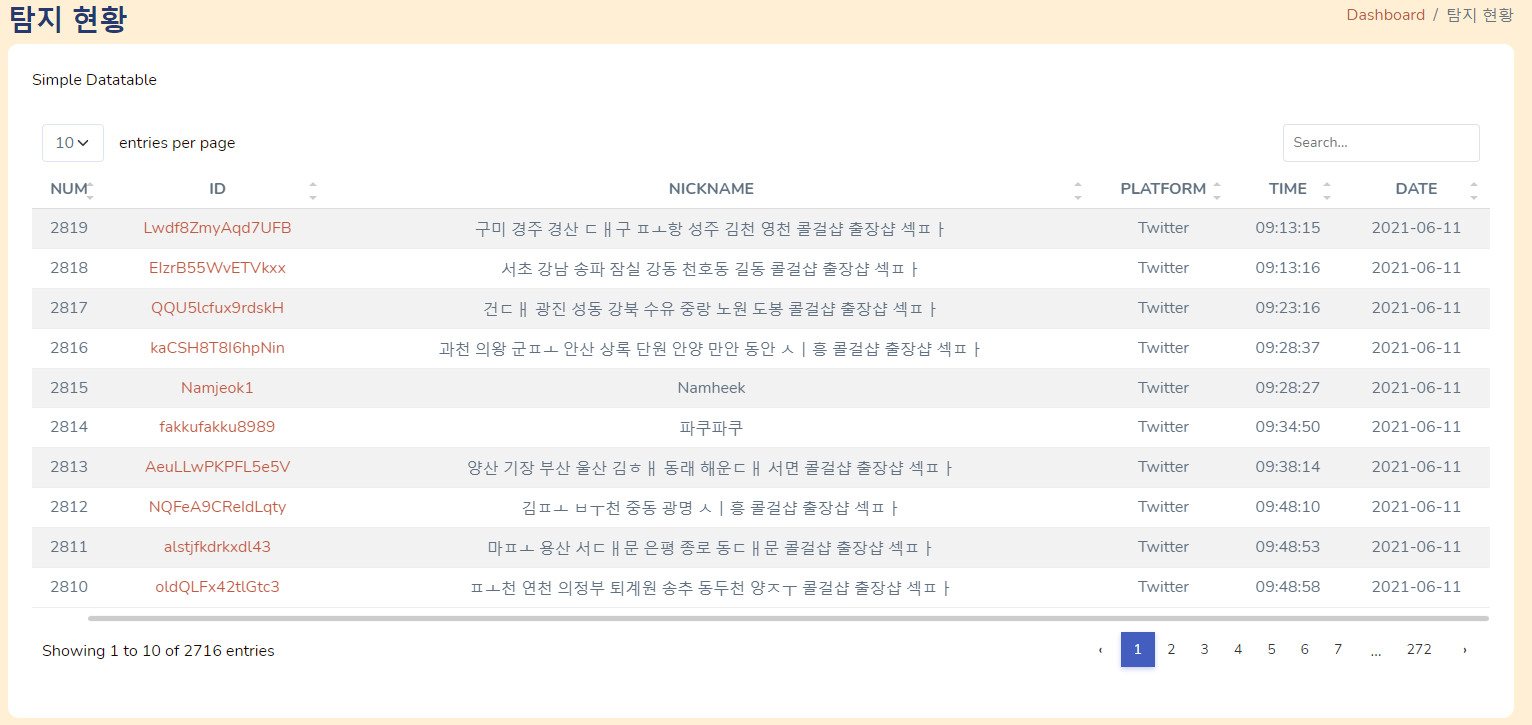
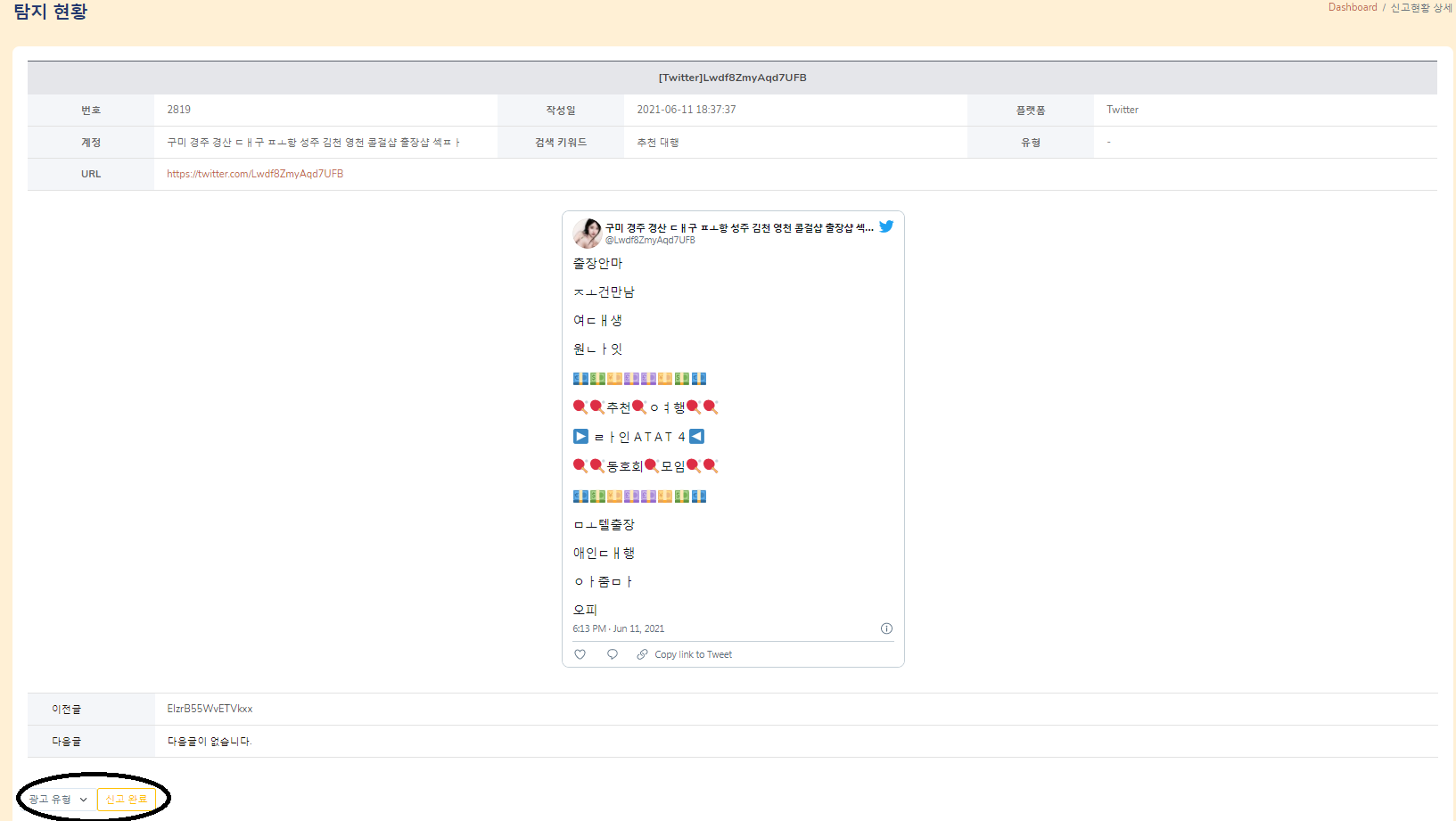
**▶ 메인 페이지**

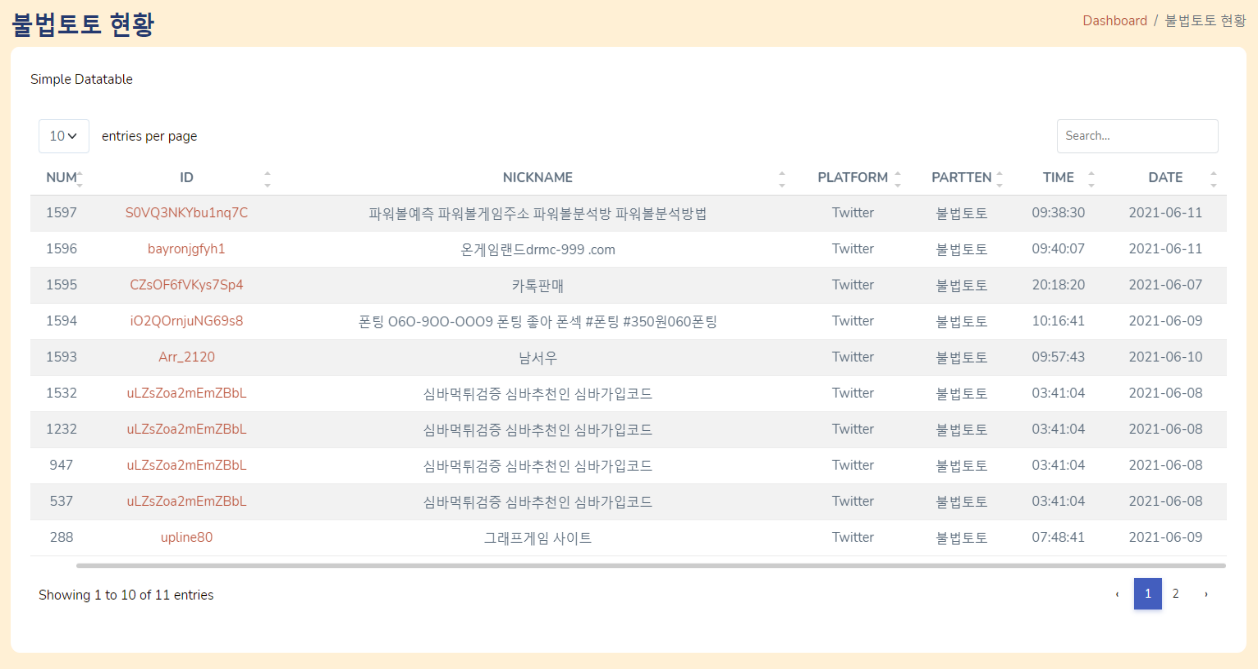


1) 금일 탐지된 불법 광고, 금일 신고한 광고, 누적 탐지된 불법광고, 누적 신고한 광고 수를 한눈에 쉽게 보여줌  
2) 시간별 탐지 현황을 그래프로 보여줘 어느 일자, 어느 시간대에 광고가 많이 탐지되는지 파악 가능  
3) 실시간으로 신고한 현황과 도넛 그래프로 누적 신고비율을 보여줌

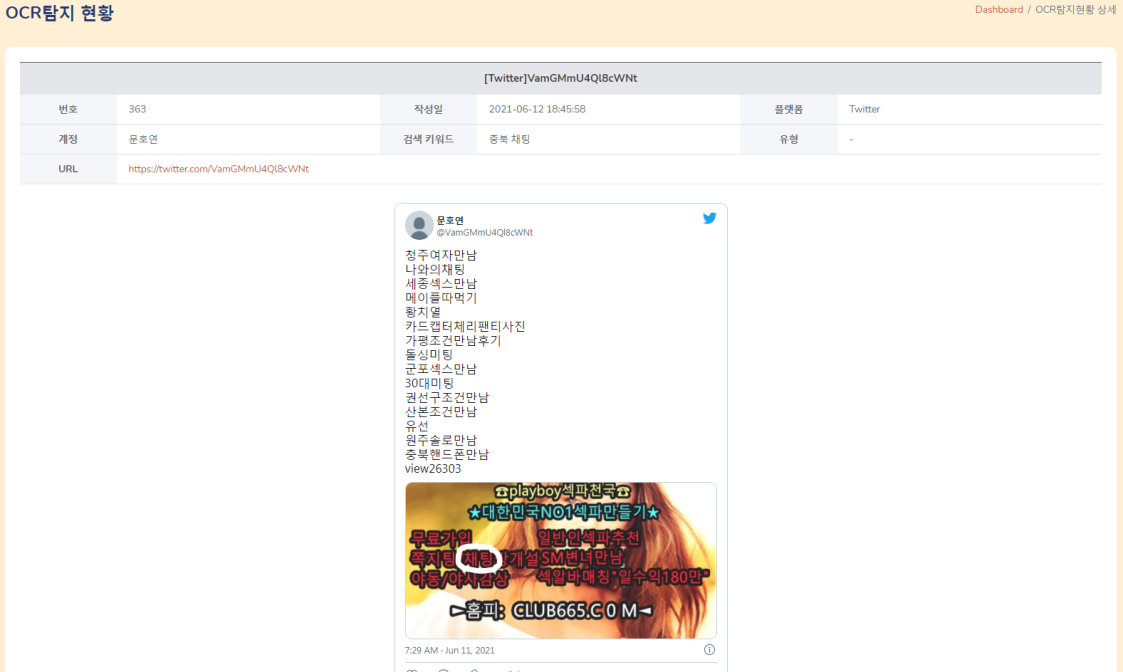
**▶실시간 현황: 실시간으로 크롤링을 통해 탐지되는 불법 광고들을 테이블 형식으로**

**보여주는 탐지 현황 게시판과 실시간으로 신고가 완료된 신고 현황 게시판, 불법 토토 광고를 나타내는 불법 토토 게시판, 이미지 안에 불법 성매매 광고 단어들이 포함된 게시글을 보여주는 OCR 게사판으로 나뉘어져 있다.**

1. **탐지 현황** **:** 크롤링을 통해 탐지된 불법 광고들을 테이블 형식으로 보여줌. ID는 해당 광고의 SNS 계정 ID, 닉네임은 SNS 계정의 닉네임을 나타내며, PLATFORM은 불법 광고의 유통 플랫폼, TIME은 해당 게시글이 작성된 시간을 DATE는 게시글 게시 날짜를 나타냄.
2. **신고 현황** **:** 탐지 현황 테이블에서 게시글을 선택하면, 해당 게시글의 상세 내용 확인 가능함. 제목은 광고의 플랫폼과 ID를 나타내며, 크롤링에 사용된 검색 키워드 및 계정의 URL 확인이 가능함. 게시글을 신고한 후, 하단의 광고 유형을 선택하여 신고 완료 버튼을 누르면 신고 현황 테이블로 넘어감.
3. **불법토토 :** 현재 성매매 알선 광고를 위주로 크롤링하고 있으나, 불법 토토 광고를 해당 페이지에 따로 나타냄



1. **OCR :** 크롤링을 통해 탐지된 이미지들 중 이미지 안에도 불법 성매매 광고 단어들이 있는 게시글을 보여줌



**▶ 통계**

- 크롤링한 데이터들을 그래프로 나타내 통계분석이 가능하도록 함.  
- 유형별 누적 검출현황은 탐지 현황에서 신고 후, 광고 유형을 분류한 데이터를 토대로 나타내며, 현재 주점형, 출장형, 조건만남형, 마사지형, 불법토토 등이 있음. 향후 유형은 업데이트 될 예정임  
- 누적 신고 통계는 메인 페이지와 같이 탐지되어 아직 신고하지 않은 불법 광고와 신고 완료된 불법광고 비율을 나타냄

* 워드클라우드는 탐지된 불법 광고에서 사용되고 있는 단어들을 시각화하여 보여줌으로써

가장 많이 사용되는 단어 파악을 용이하게 함.

* 월간 검출현황은 월별 탐지된 광고 수를 나타냄

**(2) 유해광고 실태 현황 분석**

* 크롤링을 통해 모은 자료들을 종합하여 유해광고의 현 상태 도식화
* 불법 성매매 광고에 대한 현황, 통계 등 다양한 정보를 파악할 수 있도록 현황 분석표 제작 이를 기관, 기업, 개인이 논문 및 통계자료로 이용 가능

**5. 성과**

**5.1 프로젝트 산출물**

**(1) 성매매 알선광고 자동 수집 및 분석 웹사이트 ‘Trawler’**

- 본 문서의 4절 프로젝트 소개와 같이 SNS에서 성매매 알선 광고를 위주로 불법 광고를 자동으로 검색 및 수집하여 광고 유형을 분류하고, 탐지된 광고의 통계를 나타냄.

**(2)** **사용자 매뉴얼**

- Trawler 웹 사이트의 주요 기능 및 페이지 구성요소를 포함하여 사용자가 쉽게 활용할 수 있도록 가이드 제공

**5.2 정부 기관 및 공공기관 협업**

**(1) 다시함께 상담센터**

가. 성과

- 두 번의 인터뷰 과정을 통해 다시함께 상담센터의 인터넷시민감시단에서 본 프로젝트 산출물인 Trawler의 필요성과 활용가능성을 검토함

나. 기대효과

1) 성매매 목적의 신고체계 구축으로 신고 작업이 더욱 원활해짐

2) 신고대상 수집 시간 절감

3) 수집 및 신고에 소모되는 인적 자원 절감

**6. 향후 활용 방향성**

**6.1 정부기관 및 공공기관 협업**

**(1) 사이버 경찰청**

* 성인광고 제재를 통한 성범죄 발생률 저하
* 성매매 알선광고 뿐만 아니라 불법 토토 등 불법광고 제재를 통한 범죄 예방
* 유해환경 감시를 통해 건전한 온라인 환경과 이용자의 안전과 정서를 보장
* 현재는 성매매와 관련된 데이터만 수집하지만 추후 수집하는 키워드를

불법도박까지 확대하는 방향으로 하여 경찰청과 협업

**(2) 기금조성총괄본부**



* 현재 민간단체, 모니터링단 등과 함께 불법 스포츠 도박의 근절을 위해 합동 집중 단속을 실시하는 단체
* 수집하는 키워드를 불법 도박까지 확대하는 방향으로 기금조성총괄본부와 협업

**(3) 방송통신심의위원회**



* 성매매 및 음란, 디지털 성범죄정보에 관한 규제를 담당하는 단체
* 동일한 알고리즘으로 스팸문자까지 단속할 수 있도록 모바일 확장을 고려함