(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2025년06월30일
(11) 등록번호 10-2825920
(24) 등록일자 2025년06월23일

(51) 국제특허분류(Int. Cl.)
G06Q 50/10 (2012.01) *G06F 16/9032* (2019.01)
G06F 16/9038 (2019.01) *G06F 16/906* (2019.01)
G06F 18/2135 (2023.01) *G06F 40/279* (2020.01)
G06N 3/045 (2023.01) *G06N 3/0475* (2023.01)
G06N 3/096 (2023.01) *G06V 40/16* (2022.01)
(52) CPC특허분류
G06Q 50/10 (2015.01)
G06F 16/90332 (2019.01)
(21) 출원번호 10-2024-0101366
(22) 출원일자 2024년07월31일
심사청구일자 2024년07월31일
(56) 선행기술조사문헌
KR1020220088369 A*
(뒷면에 계속)

(73) 특허권자
재단법인 서울연구원
서울 서초구 서초동 391
서울특별시 여성가족재단
서울특별시 동작구 여의대방로54길 18 (대방동)
(72) 발명자
김준철
[redacted]
윤성범
[redacted]
이지애
[redacted]
(74) 대리인
특허법인이오

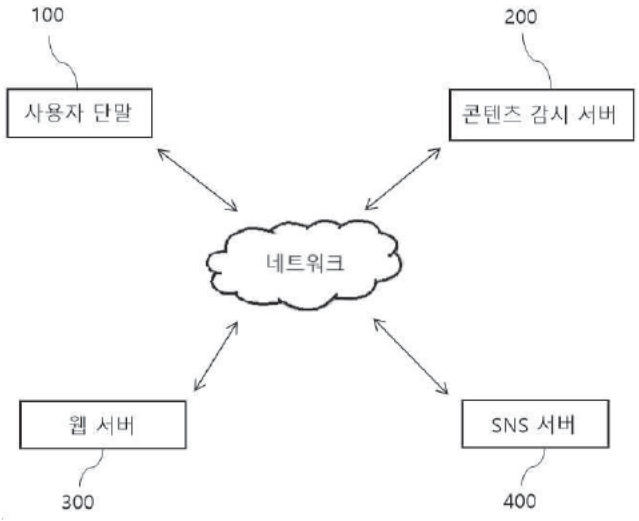
전체 청구항 수 : 총 8 항 심사관 : 나병윤

(54) 발명의 명칭 **인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템 및 방법**

(57) 요 약

인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템 및 방법을 개시한다. 본 발명은 온라인상에 게시된 불법 디지털 성범죄 콘텐츠를 추적, 검색 및 삭제되도록 관리하여 피해를 최소화시키고 피해 지원관의 업무 효율 개선과 정신적 부담을 감소시킬 수 있고, 피해자의 신고가 없는 아동/청소년 대상 불법 디지털 성범죄 콘텐츠에 대한 선제적 삭제를 수행함으로써, 해당 콘텐츠의 확산, 유포 및 재유포를 방지할 수 있다.

대 표 도 - 도1



(52) CPC특허분류
G06F 16/9038 (2019.01)
G06F 16/906 (2019.01)
G06F 18/2135 (2023.01)
G06F 40/279 (2020.01)
G06N 3/045 (2023.01)
G06N 3/0475 (2023.01)
G06N 3/096 (2023.01)
G06V 40/16 (2022.01)

(56) 선행기술조사문헌
KR1020230082801 A*
KR1020230109038 A*
KR1020240145784 A
KR1020240020787 A
KR1020210049387 A
KR102288326 B1
*는 심사관에 의하여 인용된 문헌

명 세 서

청구범위

청구항 1

사용자 단말(100)로부터 검색 대상 웹 서버(300) 및 SNS 서버(400), 임의의 프롬프트 데이터, 피해자에 대한 피해 영상물 데이터가 수신되면, 상기 프롬프트 데이터를 임베딩하여 복수의 키워드를 추출하고, 상기 키워드를 기반으로 추출되는 검색 대상 온라인 서비스 서버와, 상기 검색 대상 웹 서버(300) 및 SNS 서버(400)에 접속하여 임의의 콘텐츠를 수집하는 데이터 수집부(210);

상기 수집된 콘텐츠로부터 텍스트 데이터와 이미지 데이터를 추출하되, 추출된 단어 또는 문자열에 대한 오타와 정답을 구분하여 구분된 오타와 정답을 매칭시켜 교정 정보를 생성하고, 키워드별로 사용되는 단어 및 문자열을 설정하여 수집된 콘텐츠에 대한 키워드 정보를 추출하고, 상기 추출된 텍스트 데이터 및 이미지 데이터를 기반으로 게시글의 제목, 작성일, 작성빈도, 작성자 및 유포 빈도를 포함한 해당 콘텐츠의 데이터를 분석하여 유해 콘텐츠를 분류하는 데이터 검출부(220);

상기 유해 콘텐츠에 포함된 텍스트, 음성 및 이미지를 학습된 인공지능 기반의 분석 모델을 이용하여 상기 유해 콘텐츠와 상기 피해 영상물 데이터를 비교한 유사도 값을 기반으로 피해 영상물인지 분석하되, 상기 유해 콘텐츠의 이미지로부터 얼굴에 기반한 얼굴 형상, 눈, 코, 입의 좌우 대칭 특성, 각도 및 기울기 특성을 벡터화한 안면 인식 데이터를 추출하고, 상기 추출된 안면 인식 데이터의 특성 벡터 분류를 통해 나이와 성별을 예측하며, 상기 예측 결과에 따라 성인 대상 디지털 성범죄 콘텐츠 또는 아동/청소년 대상 디지털 성범죄 콘텐츠로 분류하는 데이터 분석부(230); 및

상기 유해 콘텐츠의 분석 결과를 데이터베이스(250)에 저장하되, 상기 유해 콘텐츠와 상기 피해 영상물 데이터의 유사도 값이 일정 값 이상이면 상기 유해 콘텐츠에 대한 이미지 캡처와, 상기 유해 콘텐츠와 관련된 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 중 하나 이상의 증거를 수집하여 채증하며, 상기 채증된 증거를 기반으로 상기 유해 콘텐츠에 대한 신고 데이터와 삭제 요청 정보를 생성하여 상기 유해 콘텐츠를 게시한 해당 온라인 서비스 서버로 출력하고, 상기 유해 콘텐츠에 대한 삭제 요청 정보를 기반으로 상기 삭제 요청에 대응한 해당 유해 콘텐츠의 삭제 결과, 재배포, 재확산 상태에 대한 모니터링을 수행하며, 상기 유해 콘텐츠가 아동/청소년 대상 디지털 성범죄 콘텐츠이면, 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 신고 데이터와 삭제 요청 정보를 추가 생성하여 상기 아동/청소년 대상 디지털 성범죄 콘텐츠를 게시한 온라인 서비스 서버와 사용자 단말(100)로 출력하는 채증/삭제 관리부(240);를 포함하는 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템.

청구항 2

제 1 항에 있어서,

상기 데이터 수집부(210)는 상기 프롬프트 데이터를 임베딩하여 키워드를 추출하되, GPT(Generative Pre-trained Transformer)를 이용한 군집 데이터 분석과 데이터 임베딩(Embedding)을 통해 복수의 키워드를 추출하는 것을 특징으로 하는 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템.

청구항 3

삭제

청구항 4

제 1 항에 있어서,

상기 채증/삭제 관리부(240)는 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 이미지 캡처와, 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 중 하나 이상의 증거를 수집하여 채증하고, 상기 채증된 증거를 기반으로 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 신고 데이터 및 삭제 요청 정보를 생성하는 것을 특징으로 하는 인공지능을 이용한 디지털 성범죄 콘텐츠

감시 시스템.

청구항 5

제 4 항에 있어서,

상기 채증/삭제 관리부(240)는 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 삭제 요청 정보를 기반으로 상기 삭제 요청에 대응한 해당 유해 콘텐츠 및 아동/청소년 대상 디지털 성범죄 콘텐츠의 삭제 결과, 재배포, 재확산 상태에 대한 추가 모니터링을 수행하는 것을 특징으로 하는 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템.

청구항 6

a) 콘텐츠 감시 서버(200)가 사용자 단말(100)로부터 검색 대상 웹 서버(300) 및 SNS 서버(400), 임의의 프롬프트 데이터, 피해자에 대한 피해 영상물 데이터를 수신하는 단계;

b) 상기 콘텐츠 감시 서버(200)가 수신된 상기 프롬프트 데이터를 임베딩하여 복수의 키워드를 추출하고, 상기 키워드를 기반으로 추출되는 검색 대상 온라인 서비스 서버와, 상기 검색 대상 웹 서버(300) 및 SNS 서버(400)에 접속하여 임의의 콘텐츠를 수집하는 단계;

c) 상기 콘텐츠 감시 서버(200)가 수집된 콘텐츠로부터 텍스트 데이터와 이미지 데이터를 추출하되, 추출된 단어 또는 문자열에 대한 오타와 정답을 구분하여 구분된 오타와 정답을 매칭시켜 교정 정보를 생성하고, 키워드별로 사용되는 단어 및 문자열을 설정하여 수집된 콘텐츠에 대한 키워드 정보를 추출하며, 상기 추출된 텍스트 데이터 및 이미지 데이터를 기반으로 게시글의 제목, 작성일, 작성빈도, 작성자 및 유포 빈도를 포함한 해당 콘텐츠의 데이터를 분석하여 유해 콘텐츠를 분류하는 단계;

d) 상기 분류 결과, 성인 대상 불법 성범죄 콘텐츠이면, 콘텐츠 감시 서버(200)가 상기 유해 콘텐츠에 포함된 텍스트, 음성 및 이미지를 학습된 인공지능 기반의 분석 모델을 이용하여 상기 유해 콘텐츠와 상기 피해 영상물 데이터를 비교한 유사도 값을 기반으로 피해 영상물인지 분석하되, 상기 유해 콘텐츠의 이미지로부터 얼굴에 기반한 얼굴 형상, 눈, 코, 입의 좌우 대칭 특성, 각도 및 기울기 특성을 벡터화한 안면 인식 데이터를 추출하고, 상기 추출된 안면 인식 데이터의 특성 벡터 분류를 통해 나이와 성별을 예측하며, 상기 예측 결과에 따라 성인 대상 디지털 성범죄 콘텐츠 또는 아동/청소년 대상 디지털 성범죄 콘텐츠로 분류하는 단계; 및

e) 상기 콘텐츠 감시 서버(200)가 상기 유해 콘텐츠의 분석 결과를 데이터베이스(250)에 저장하되, 상기 유해 콘텐츠와 상기 피해 영상물 데이터의 유사도 값이 일정 값 이상이면 상기 유해 콘텐츠에 대한 이미지 캡처와, 상기 유해 콘텐츠와 관련된 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 중 하나 이상의 증거를 수집하여 채증하며, 상기 채증된 증거를 기반으로 상기 유해 콘텐츠에 대한 신고 데이터와 삭제 요청 정보를 생성하여 상기 유해 콘텐츠를 게시한 해당 온라인 서비스 서버로 출력하고, 상기 유해 콘텐츠에 대한 삭제 요청 정보를 기반으로 상기 삭제 요청에 대응한 해당 유해 콘텐츠의 삭제 결과, 재배포, 재확산 상태에 대한 모니터링을 수행하며, 상기 유해 콘텐츠가 아동/청소년 대상 디지털 성범죄 콘텐츠이면, 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 신고 데이터와 삭제 요청 정보를 추가 생성하여 상기 아동/청소년 대상 디지털 성범죄 콘텐츠를 게시한 온라인 서비스 서버와 사용자 단말(100)로 출력하는 단계;를 포함하는 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 방법.

청구항 7

제 6 항에 있어서,

상기 a) 단계는 상기 콘텐츠 감시 서버(200)가 상기 프롬프트 데이터를 임베딩하여 키워드를 추출하되, GPT(Generative Pre-trained Transformer)를 이용한 군집 데이터 분석과 데이터 임베딩(Embedding)을 통해 복수의 키워드를 추출하는 것을 특징으로 하는 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 방법.

청구항 8

삭제

청구항 9

제 6 항에 있어서,

상기 유해 콘텐츠가 아동/청소년 대상 디지털 성범죄 콘텐츠로 분류되면, 상기 콘텐츠 감시 서버(200)가 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 이미지 캡처와, 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 중 하나 이상의 증거를 수집하여 채증하고, 상기 채증된 증거를 기반으로 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 신고 데이터 및 삭제 요청 정보를 추가 생성하여 해당 온라인 서비스 서버로 출력하는 단계;를 더 포함하는 것을 특징으로 하는 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 방법.

청구항 10

제 9 항에 있어서,

상기 콘텐츠 감시 서버(200)가 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 삭제 요청 정보를 기반으로 상기 삭제 요청에 대응한 해당 유해 콘텐츠 및 아동/청소년 대상 디지털 성범죄 콘텐츠의 삭제 결과, 재배포, 재확산 상태에 대한 추가 모니터링을 수행하는 단계;를 더 포함하는 것을 특징으로 하는 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템 및 방법에 관한 발명으로서, 더욱 상세하게는 온라인상에 게시된 불법 디지털 성범죄 콘텐츠를 추적, 검색 및 삭제되도록 관리하여 피해를 최소화시키고 피해 지원관의 업무 효율 개선과 정신적 부담을 감소시킬 수 있는 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템 및 방법에 관한 것이다.

배 경 기 술

[0002] 컴퓨터 통신 및 이동 통신의 발달은 온라인을 이용한 웹 서비스, 소셜 네트워크 서비스(SNS) 등을 이용한 디지털 콘텐츠의 사용이 널리 확대될 수 있도록 하였다.

[0003] 이러한 웹 서비스와 소셜 네트워크 서비스를 통해 누구나 정보를 쉽게 얻을 수 있게 되었지만, 많은 수의 유해 콘텐츠가 웹 서비스와 소셜 네트워크 서비스를 통해 무분별하게 배포됨으로써 사회적 문제가 되고 있다.

[0004] 이러한 유해 콘텐츠 중에는 디지털 성범죄 콘텐츠가 포함될 수 있고, 웹 서비스와 소셜 네트워크 서비스를 통해 단기간에 급속히 전파되는 디지털 성범죄 콘텐츠는 피해자의 인권을 침해하고 국민 정서를 훼손하여 사회적인 문제로 대두 되고 있다.

[0005] 최근 들어서는, N 번방, 웰컴투 비디오 등 아동/청소년을 대상으로 하는 디지털 성범죄와, 이와 관련된 불법 디지털 성범죄 콘텐츠의 증가로 인해 학부모와 시민들의 불안이 더욱 가중되고 있다.

[0006] 더욱이, 무제한적 복제가 가능한 온라인 범죄의 특성상, 디지털 성범죄 피해자가 받는 2차, 3차적 고통 역시 심각한 상황이다.

[0007] 또한, 디지털 기술의 발전으로 아동 청소년에 대한 성착취 수법이 고도화되고 다양화되고 있으며 피해 규모도 급증하고 있다.

[0008] 2021년 7월 서울시 실태조사에 따르면 초/중/고생 4012명 중 약 21.3%(856명)이 채팅이나 SNS를 통해 디지털 성범죄에 직면한 경험이 있는 것으로 나타났다.

[0009] 또한, 조사 결과에 따르면 디지털 성범죄에 노출된 아동/청소년 중 56.4%가 성적인 메시지나 사진을 전송받은 적이 있다고 언급했으며, 이들의 27.2%는 온라인에서 지속적인 연락 및 만남 요구를 받았다고 응답했다.

[0010] 응답자들 중 4.8% 가 성적 이미지 유포 또는 유포 협박을 받았으며, 4.3%가 성적인 행위를 하면 돈을 주겠다는 제안을 받은 경험이 있다고 보고했다.

[0011] 이는 디지털 성범죄가 빠르게 확산되고 있으며 아동 청소년을 대상으로 한 범죄의 경우 특히, 선제적인 대응이 필요함을 시사한다.

[0012] 그러나 성인과 아동/청소년을 대상으로 하는 디지털 성범죄의 발생률이 지속적으로 상승하고 있는 상황에서, 이러한 범죄를 미리 차단하고 피해를 최소화하기 위해서는 지속적인 온라인 콘텐츠 감시 기술이 요구되고 있지만, 다양한 온라인 서비스 제공자(OSP)에서의 디지털 성범죄 콘텐츠를 효과적으로 탐지 및 분류하지 못하는 문제점이 있다.

선행기술문헌

특허문헌

[0013] (특허문헌 0001) 한국 공개특허공보 공개번호 제10-2021-0098651호, 공개일 2021.08.11. (발명의 명칭: 음란물 유포 감시 장치 및 감시 방법)

발명의 내용

해결하려는 과제

[0014] 이러한 문제점을 해결하기 위하여, 본 발명은 온라인상에 게시된 불법 디지털 성범죄 콘텐츠를 추적, 검색 및 삭제되도록 관리하여 피해를 최소화시키고 피해 지원관의 업무 효율 개선과 정신적 부담을 감소시킬 수 있는 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템 및 방법을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0015] 본 발명의 일 실시 예는 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템으로서, 사용자 단말로부터 임의의 검색 대상 온라인 서비스 서버 정보, 임의의 프롬프트 데이터, 피해자에 대한 피해 영상물 데이터가 수신되면, 상기 프롬프트 데이터를 임베딩하여 복수의 키워드를 추출하고, 상기 추출된 키워드를 기반으로 상기 검색 대상 온라인 서비스 서버에 접속하여 임의의 콘텐츠를 수집하는 데이터 수집부; 상기 수집된 콘텐츠로부터 텍스트 데이터와 이미지 데이터를 추출하고, 상기 추출된 텍스트 데이터 및 이미지 데이터를 기반으로 게시글의 제목, 작성일, 작성빈도, 작성자 및 유포 빈도를 포함한 해당 콘텐츠의 데이터를 분석하여 유해 콘텐츠를 분류하는 데이터 검출부; 상기 유해 콘텐츠에 포함된 텍스트, 음성 및 이미지를 인공지능 기반의 분석 모델을 이용하여 상기 유해 콘텐츠와 상기 피해 영상물 데이터를 비교한 유사도 값을 기반으로 피해 영상물인지 분석하는 데이터 분석부; 및 상기 유해 콘텐츠의 분석 결과를 데이터베이스에 저장하되, 상기 유사도 값이 일정 값 이상이면 상기 유해 콘텐츠에 대한 이미지 캡처와, 상기 유해 콘텐츠와 관련된 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 중 하나 이상의 증거를 수집하여 채증하고, 상기 채증된 증거를 기반으로 상기 유해 콘텐츠에 대한 신고 데이터와 삭제 요청 정보를 생성하여 해당 온라인 서비스 서버로 출력하는 채증/삭제 관리부;를 포함한다.

[0016] 또한, 상기 실시 예에 따른 데이터 수집부는 상기 입력된 프롬프트 데이터를 임베딩하여 키워드를 추출하되, GPT(Generative Pre-trained Transformer)를 이용한 군집 데이터 분석과 데이터 임베딩(Embedding)을 통해 복수의 키워드를 추출하는 것을 특징으로 한다.

[0017] 또한, 상기 실시 예에 따른 데이터 분석부는 상기 유해 콘텐츠의 이미지로부터 피해자의 얼굴을 인식하고, 상기 피해자의 얼굴에 기반한 얼굴 형상, 눈, 코, 입의 좌우 대칭 특성, 각도 및 기울기 특성에 따른 분류를 통해 피해자의 나이와 성별을 예측하고, 상기 예측 결과에 따라 성인 대상 디지털 성범죄 콘텐츠 또는 아동/청소년 대상 디지털 성범죄 콘텐츠로 분류하는 것을 특징으로 한다.

[0018] 또한, 상기 실시 예에 따른 채증/삭제 관리부는 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 이미지 캡처와, 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 중 하나 이상의 증거를 수집하여 채증하고, 상기 채증된 증거를 기반으로 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 신고 데이터 및 삭제 요청 정보를 추가 생성하여 해당 온라인 서비스 서버로 출력하는 것을 특징으로 한다.

[0019] 또한, 상기 실시 예에 따른 채증/삭제 관리부는 상기 유해 콘텐츠 및 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 삭제 요청 정보를 기반으로 상기 삭제 요청에 대응한 해당 유해 콘텐츠 및 아동/청소년 대상 디지털 성범죄 콘텐츠의 삭제 결과, 재배포, 재확산 상태에 대한 추가 모니터링을 수행하는 것을 특징으로 한다.

[0020] 또한, 본 발명의 일 실시 예는 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 방법으로서, a) 콘텐츠 감시 서버가 사용자 단말로부터 임의의 검색 대상 온라인 서비스 서버 정보, 임의의 프롬프트 데이터, 피해자에 대한 피해 영상물 데이터를 수신하는 단계; b) 상기 콘텐츠 감시 서버가 수신된 상기 프롬프트 데이터를 임베딩하여 복수의 키워드를 추출하고, 상기 추출된 키워드를 기반으로 상기 검색 대상 온라인 서비스 서버에 접속하여 임의의 콘텐츠를 수집하는 단계; c) 상기 콘텐츠 감시 서버가 수집된 콘텐츠로부터 텍스트 데이터와 이미지 데이터를 추출하고, 상기 추출된 텍스트 데이터 및 이미지 데이터를 기반으로 게시글의 제목, 작성일, 작성빈도, 작성자 및 유포 빈도를 포함한 해당 콘텐츠의 데이터를 분석하여 유해 콘텐츠를 분류하는 단계; d) 상기 분류 결과, 성인 대상 불법 성범죄 콘텐츠이면 상기 콘텐츠 감시 서버가 상기 유해 콘텐츠에 포함된 텍스트, 음성 및 이미지를 인공지능 기반의 분석 모델을 이용하여 상기 유해 콘텐츠와 상기 피해 영상물 데이터를 비교한 유사도 값을 기반으로 피해 영상물인지 분석하는 단계; 및 e) 상기 콘텐츠 감시 서버가 상기 유해 콘텐츠의 분석 결과를 데이터베이스에 저장하되, 상기 유사도 값이 일정 값 이상이면 상기 유해 콘텐츠에 대한 이미지 캡처와, 상기 유해 콘텐츠와 관련된 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 중 하나 이상의 증거를 수집하여 채증하고, 상기 채증된 증거를 기반으로 상기 유해 콘텐츠에 대한 신고 데이터와 삭제 요청 정보를 생성하여 해당 온라인 서비스 서버로 출력하는 단계;를 포함한다.

[0021] 또한, 상기 실시 예에 따른 a) 단계는 상기 콘텐츠 감시 서버가 상기 입력된 프롬프트 데이터를 임베딩하여 키워드를 추출하되, GPT(Generative Pre-trained Transformer)를 이용한 군집 데이터 분석과 데이터 임베딩(Embedding)을 통해 복수의 키워드를 추출하는 것을 특징으로 한다.

[0022] 또한, 상기 실시 예에 따른 c) 단계는 상기 콘텐츠 감시 서버가 분류된 상기 유해 콘텐츠의 이미지로부터 피해자의 얼굴을 인식하고, 상기 피해자의 얼굴에 기반한 얼굴 형상, 눈, 코, 입의 좌우 대칭 특성, 각도 및 기울기 특성에 따른 분류를 통해 피해자의 나이와 성별을 예측하고, 상기 예측 결과에 따라 성인 대상 디지털 성범죄 콘텐츠 또는 아동/청소년 대상 디지털 성범죄 콘텐츠로 분류하는 단계;를 더 포함하는 것을 특징으로 한다.

[0023] 또한, 상기 실시 예에 따른 유해 콘텐츠가 아동/청소년 대상 디지털 성범죄 콘텐츠로 분류되면

[0024] d) 단계는 상기 콘텐츠 감시 서버가 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 이미지 캡처와, 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 중 하나 이상의 증거를 수집하여 채증하고, 상기 채증된 증거를 기반으로 상기 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 신고 데이터 및 삭제 요청 정보를 추가 생성하여 해당 온라인 서비스 서버로 출력하는 단계;를 더 포함하는 것을 특징으로 한다.

[0025] 또한, 상기 실시 예에 따른 d) 단계는 상기 콘텐츠 감시 서버가 상기 유해 콘텐츠 및 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 삭제 요청 정보를 기반으로 상기 삭제 요청에 대응한 해당 유해 콘텐츠 및 아동/청소년 대상 디지털 성범죄 콘텐츠의 삭제 결과, 재배포, 재확산 상태에 대한 추가 모니터링을 수행하는 단계;를 더 포함하는 것을 특징으로 한다.

발명의 효과

[0026] 본 발명은 온라인상에 게시된 불법 디지털 성범죄 콘텐츠를 추적, 검색 및 삭제되도록 관리하여 피해를 최소화시키고 피해 지원관의 업무 효율 개선과 정신적 부담을 감소시킬 수 있는 장점이 있다.

[0027] 또한, 본 발명은 피해자의 신고가 없는 아동/청소년 대상 불법 디지털 성범죄 콘텐츠에 대한 선제적 삭제를 수행함으로써, 해당 콘텐츠의 확산, 유포 및 재유포를 방지할 수 있는 장점이 있다.

도면의 간단한 설명

[0028] 도 1은 본 발명의 일 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템을 개략적으로 나타낸 블록도.

도 2는 도1의 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템의 콘텐츠 감시 서버 구성을 설명하기 위해 나타낸 블록도.

도 3은 도 2의 실시 예에 따른 콘텐츠 감시 서버의 데이터 분석부 구성을 설명하기 위해 나타낸 블록도.

도 4는 도 2의 실시 예에 따른 콘텐츠 감시 서버의 채증/삭제 관리부 구성을 설명하기 위해 나타낸 블록도.

도 5는 본 발명의 일 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 방법을 설명하기 위해 나타

낸 흐름도.

도 6은 도 5의 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 방법의 증거 채증 및 삭제 과정을 설명하기 위해 나타낸 흐름도.

발명을 실시하기 위한 구체적인 내용

[0029] 이하에서는 본 발명의 바람직한 실시 예 및 첨부하는 도면을 참조하여 본 발명을 상세히 설명하되, 도면의 동일한 참조부호는 동일한 구성요소를 지칭함을 전제하여 설명하기로 한다.

[0030] 본 발명의 실시를 위한 구체적인 내용을 설명하기에 앞서, 본 발명의 기술적 요지와 직접적 관련이 없는 구성에 대해서는 본 발명의 기술적 요지를 흐뜨리지 않는 범위 내에서 생략하였음에 유의하여야 할 것이다.

[0031] 또한, 본 명세서 및 청구범위에 사용된 용어 또는 단어는 발명자가 자신의 발명을 최선의 방법으로 설명하기 위해 적절한 용어의 개념을 정의할 수 있다는 원칙에 입각하여 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야 할 것이다.

[0032] 본 명세서에서 어떤 부분이 어떤 구성요소를 "포함"한다는 표현은 다른 구성요소를 배제하는 것이 아니라 다른 구성요소를 더 포함할 수 있다는 것을 의미한다.

[0033] 또한, "··부", "··기", "··모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어, 또는 그 둘의 결합으로 구분될 수 있다.

[0034] 또한, "적어도 하나의" 라는 용어는 단수 및 복수를 포함하는 용어로 정의되고, 적어도 하나의 라는 용어가 존재하지 않더라도 각 구성요소가 단수 또는 복수로 존재할 수 있고, 단수 또는 복수를 의미할 수 있음은 자명하다 할 것이다.

[0035] 또한, 각 구성요소가 단수 또는 복수로 구비되는 것은, 실시 예에 따라 변경가능하다 할 것이다.

[0036] 이하, 첨부된 도면을 참조하여 본 발명의 일 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템 및 방법의 바람직한 실시 예를 상세하게 설명한다.

[0037] 도 1은 본 발명의 일 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템을 개략적으로 나타낸 블록도이고, 도 2는 도1의 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템의 콘텐츠 감시 서버 구성을 설명하기 위해 나타낸 블록도이며, 도 3은 도 2의 실시 예에 따른 콘텐츠 감시 서버의 데이터 분석부 구성을 설명하기 위해 나타낸 블록도이고, 도 4는 도 2의 실시 예에 따른 콘텐츠 감시 서버의 채증/삭제 관리부 구성을 설명하기 위해 나타낸 블록도이다.

[0038] 도 1 내지 도 4에 나타낸 바와 같이, 본 발명의 일 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 시스템은 온라인상에 게시된 불법 디지털 성범죄 콘텐츠를 추적, 검색 및 삭제되도록 관리하여 피해가 최소될 수 있도록 사용자 단말(100)과, 콘텐츠 검색 서버(200)와, 검색 대상인 웹 서버(300)와 SNS 서버(400)를 포함하여 구성될 수 있다.

[0039] 사용자 단말(100)은 네트워크를 통해 콘텐츠 검색 서버(200)와 접속하여 온라인상에 게시된 불법 디지털 성범죄 콘텐츠를 추적, 검색 및 삭제의 요청이 가능하도록 애플리케이션 프로그램의 설치가 가능한 데스크탑 PC, 노트북 PC, 태블릿 PC, 스마트 폰과 같은 모바일 단말로 구성될 수 있다.

[0040] 또한, 사용자 단말(100)은 피해자의 신고를 통해 신고된 피해 영상물을 콘텐츠 검색 서버(200)로 전송할 수 있다.

[0041] 또한, 사용자 단말(100)은 검색 대상인 웹 서버(300) 및 SNS 서버(400) 등을 포함한 온라인 서비스 서버의 정보를 입력받아 콘텐츠 검색 서버(200)로 전송할 수 있다.

[0042] 또한, 사용자 단말(1000은 피해 영상물과 관련하여 검색하고 싶은 내용을 포함한 프롬프트 데이터(Prompt data)를 텍스트 데이터로 입력받을 수 있다.

[0043] 여기서, 프롬프트 데이터는 예를 들어, 피해 영상물 관련 사물, 언어, 신조어, 채팅 데이터, 웹 사이트, SNS 데이터 등을 포함할 수 있다.

[0044] 또한, 사용자 단말(100)은 콘텐츠 검색 서버(200)로부터 전송되는 피해 영상물 관련 검색 결과 및 증거 정보, 성인 대상 디지털 성범죄 콘텐츠 및 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 검색 결과 및 해당 콘텐츠

들의 증거 정보와 삭제 관련 처리 정보를 수신하여 디스플레이되도록 한다.

[0045] 콘텐츠 검색 서버(200)는 온라인상에 게시된 불법 디지털 성범죄 콘텐츠를 추적, 검색 및 삭제하여 관리하는 구성으로서, 데이터 수집부(210)와, 데이터 검출부(220)와, 데이터 분석부(230)와, 채증/삭제 관리부(240)와, 데이터베이스(250)를 포함하여 구성될 수 있다.

[0046] 데이터 수집부(210)는 사용자 단말(100)로부터 임의의 검색 대상 온라인 서비스 서버 정보, 임의의 프롬프트 데이터, 피해자에 대한 피해 영상물 데이터를 수신할 수 있다.

[0047] 데이터 수집부(210)는 수신된 검색 대상 웹 서버(300) 및 SNS 서버(400)를 대상으로 웹 크롤링(Web crawling)을 통해 유해 콘텐츠 및 피해 영상물 관련 콘텐츠를 수집할 수 있다.

[0048] 또한, 데이터 수집부(210)는 입력된 프롬프트 데이터를 임베딩하여 키워드를 추출하되, GPT(Generative Pre-trained Transformer)를 이용한 군집 데이터 분석과 데이터 임베딩(Embedding)을 통해 복수의 키워드를 추출할 수 있다.

[0049] 데이터 수집부(210)는 프롬프트 데이터 중에서, 아동, 청소년에 대한 내용이 포함되어 있으면, 예를 들어, '책', '교복', '인형' 등의 사물 관련 키워드와 청소년들이 주로 사용하는 언어, 신조어 등을 키워드로 추출하여 더욱 많은 유해 콘텐츠 및 피해 관련 영상 콘텐츠를 수집할 수 있도록 한다.

[0050] 또한, 데이터 수집부(210)는 사용자 단말(100)로부터 입력되는 검색 대상 온라인 서비스 서버 이외에, 추출된 키워드를 기반으로 검색 대상 온라인 서비스 서버의 정보를 추출하고, 해당 온라인 서비스에 접속하여 웹 크롤링을 통해 유해 콘텐츠 및 피해 관련 영상 콘텐츠를 수집할 수 있다.

[0051] 데이터 검출부(220)는 웹 크롤링을 통해 수집된 콘텐츠로부터 텍스트 데이터와 이미지 데이터를 추출할 수 있다.

[0052] 또한, 데이터 검출부(220)는 추출된 텍스트 데이터 및 이미지 데이터를 기반으로 게시글의 제목, 작성일, 작성빈도, 작성자 및 유포 빈도를 포함한 해당 콘텐츠의 데이터를 분석하여 성인 대상 디지털 성범죄 콘텐츠와 아동/청소년 대상 디지털 성범죄 콘텐츠 등으로 유해 콘텐츠를 분류할 수 있다.

[0053] 또한, 데이터 검출부(220)는 수집된 콘텐츠 관련 데이터에서 잘못 표기된 단어 또는 문자열을 미리 설정된 교정룰(Rule)에 기반하여 교정하고, 해당 콘텐츠를 텍스트 데이터와 이미지 데이터로 분류할 수 있다.

[0054] 여기서, 교정 룰은 추출된 단어 또는 문자열에 대한 오타와 정답을 구분하고, 구분된 오타와 정답을 매칭시켜 교정 정보를 생성할 수 있다.

[0055] 또한, 데이터 검출부(220)는 키워드별로 사용되는 단어 및 문자열 등을 설정하고, 수집된 콘텐츠 관련 데이터에서 자연어 언어 모델인 BERT 모델을 이용하여 해당 콘텐츠에 대한 키워드 정보를 추출할 수도 있다.

[0056] 데이터 분석부(230)는 유해 콘텐츠에 포함된 텍스트, 음성 및 이미지를 인공지능 기반의 분석 모델을 이용하여 유해 콘텐츠와 피해 영상물 데이터를 비교하여 유사도 값을 산출할 수 있다.

[0057] 또한, 데이터 분석부(230)는 유해 콘텐츠와 피해 영상물 데이터를 비교한 유사도 값을 기반으로 피해 영상물인지 분석할 수 있으며, 유사도 분석부(231)를 포함하여 구성될 수 있다.

[0058] 유사도 분석부(231)는 광학문자인식(OCR)을 통해 유해 콘텐츠에 포함된 텍스트 데이터를 인식할 수 있고, 인식된 유해 콘텐츠의 텍스트 데이터에 대하여 텍스트 값들을 벡터 공간에 위치시켜 벡터 값으로 변환한 임베딩 벡터를 산출할 수 있다.

[0059] 또한, 유사도 분석부(231)는 산출된 임베딩 벡터를 각 벡터의 위치에 기반하여 키워드 정보와의 행렬 연산을 통해 유사도에 따른 유사도 값을 산출할 수 있다.

[0060] 또한, 유사도 분석부(231)는 텍스트 데이터에 대한 유사도 분석을 통해 게시글의 작성 빈도, 게시글 작성자, 유포 빈도와 같은 게시글의 패턴을 분석할 수도 있다.

[0061] 또한, 유사도 분석부(231)는 학습된 인공지능 기반의 분석 모델을 이용하여 유해 콘텐츠에 포함된 음성을 분석하고, 분석된 음성의 특징 벡터를 추출하여 추출된 음성의 특징 벡터로부터 음향적 특성과 언어적 특성에 기반한 피해 영상물 데이터와의 유사도에 따른 유사도 값을 산출할 수 있다.

[0062] 또한, 유사도 분석부(231)는 학습된 인공지능 기반의 분석 모델을 이용하여 유해 콘텐츠에 포함된 이미지를 분

석하고, 분석된 이미지의 특징 벡터를 추출하여 추출된 이미지의 특징 벡터로부터 이미지 특성에 기반한 피해 영상물 데이터와의 유사도에 따른 유사도 값을 산출할 수 있다.

[0063] 또한, 데이터 분석부(230)는 성인 대상 디지털 성범죄 콘텐츠 또는 아동/청소년 대상 디지털 성범죄 콘텐츠로 분류할 수 있다.

[0064] 이를 위해, 데이터 분석부(230)는 유해 콘텐츠의 이미지로부터 피해자의 얼굴을 인식하고, 인식된 피해자의 얼굴에 기반한 얼굴 형상, 눈, 코, 입의 좌우 대칭 특성, 각도 및 기울기 특성에 따른 분류를 통해 피해자의 나이와 성별을 예측할 수 있도록 예측 분석부(232)를 포함하여 구성될 수 있다.

[0065] 예측 분석부(232)는 유해 콘텐츠에 포함된 이미지로부터 피해자의 얼굴을 인식한 안면 인식 데이터를 추출할 수 있고, 안면 인식 데이터는 사용자의 얼굴 형상, 눈, 코, 입의 좌우 대칭 특성, 각도 및 기울기 등을 포함할 수 있고, 추출된 안면 인식 데이터의 특성은 벡터화될 수 있다..

[0066] 예측 분석부(232)는 나이대별, 성별에 따른 얼굴 모습을 획득하여 레이블이 지정된 학습 데이터 셋을 기반으로 CNN(Convolutional Neural Network) 알고리즘을 이용한 기계 학습을 통해 나이대별 및 성별의 분류를 학습한 인공지능 기반의 분석 모델을 이용할 수 있다.

[0067] 여기서, 인공지능 기반의 분석 모델은 CNN(Convolutional Neural Network) 기반의 딥러닝 모델을 이용하여 이미지에 기반한 학습 데이터로부터 나이대별 및 성별에 대한 분류를 학습할 수 있다.

[0068] 또한, 인공지능 기반의 분석 모델은 기계 학습(Machine Learning) 중에서 딥러닝(Deep learning)이라는 방법을 통해 만들어진 분석 모델들의 종류라고 볼 수 있다.

[0069] 따라서, 인공지능 기반의 분석 모델은 딥러닝 모델 또는 딥러닝 분석 모델의 표현으로 사용될 수도 있다.

[0070] 또한, 기계 학습은 복잡한 시스템이 명시적으로 프로그래밍되지 않고서, 경험으로부터 자동으로 학습하고 개선할 수 있게 하는 인공 지능의 응용이다.

[0071] 또한, 기계 학습 모델들의 정확도 및 유효성은 그들 모델들을 훈련시키는 데 사용되는 데이터에 부분적으로 의존할 수 있다.

[0072] 또한, 인공지능 기반의 분석 모델은 분류를 수행하는 과정에서 수집된 피해자의 나이대별 및 성별에 대한 분류 데이터를 이용하여 분석 모델의 재학습과 이에 따른 성능 검증을 통해 분석 모델의 수정을 추가 수행할 수도 있다.

[0073] 즉, 예측 분석부(232)는 분석 모델의 성능을 향상시키기 위해 피처 엔지니어링(Feature Engineering), 하이퍼 파라미터(Hyper parameter) 최적화, 챔피언 모델(Champion Model)을 이용한 오토 ML(Auto Machine Learning)을 추가 수행함으로써, 분석가의 분석 없이도 분석을 수행할 수 있고, 사용 편의성을 향상시킬 수 있도록 한다.

[0074] 여기서, 피처 엔지니어링은 특성 엔지니어링, 기존 데이터 변수들에서 모델 개선에 사용할 수 있는 의미 있는 변수들을 선택하고, 변형하거나 예측력에 영향을 미치는 새로운 변수를 생성하여 모델 성능을 높이는 방법이다.

[0075] 또한, 하이퍼 파라미터 최적화는 인공 신경망 훈련 시 가장 우수한 성능을 도출할 수 있는 하이퍼 파라미터를 찾아내는 기술을 의미하며, 하이퍼 파라미터로는 학습률, 학습류 스케줄링 방법, 손실 함수, 훈련 반복횟수, 가중치 초기화 방법, 정규화방법, 적층할 계층의 수 등이 고려될 수 있다.

[0076] 또한, 예측 분석부(232)는 데이터 증강(Data Augmentation)을 통해 추가 학습 데이터 셋을 생성하고, 생성된 추가 학습 데이터 셋을 이용하여 학습할 수도 있다.

[0077] 또한, 예측 분석부(232)는 추출된 안면 인식 데이터의 특성 벡터 분류를 통해 피해자의 나이 및 성별을 예측한 결과값을 획득할 수 있다.

[0078] 이를 통해 예측 분석부(232)는 피해자의 나이가 예를 들어, 10대 청소년 또는 아동으로 예측되면, 아동/청소년 대상 디지털 성범죄 콘텐츠로 분류할 수 있다.

[0079] 채증/삭제 관리부(240)는 데이터 분석부(230)에서 분석한 유해 콘텐츠의 분석 결과를 데이터베이스(250)에 저장되도록 한다.

[0080] 또한, 채증/삭제 관리부(240)는 증거 추출부(241)와, 삭제 감시부(242)를 포함하여 구성될 수 있다.

[0081] 증거 추출부(241)는 성인 대상 불법 성범죄 콘텐츠로 분류되고, 데이터 분석부(230)에서 분석된 유해 콘텐츠와

피해 영상물 데이터의 유사도 값이 미리 설정된 일정 값 이상이면, 해당 유해 콘텐츠에 대한 이미지의 캡처를 수행할 수 있다.

[0082] 또한, 증거 추출부(241)는 해당 유해 콘텐츠와 관련된 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 등의 증거를 수집하여 채증을 수행할 수 있다.

[0083] 또한, 증거 추출부(241)는 채증을 통해 수집한 증거를 기반으로 해당 유해 콘텐츠에 대한 신고 데이터와 해당 유해 콘텐츠의 삭제 요청 정보를 생성하고, 생성된 신고 데이터와 삭제 요청 정보를 유해 콘텐츠를 게시한 온라인 서비스 서버와 사용자 단말(100)로 출력한다.

[0084] 또한, 증거 추출부(241)는 데이터 분석부(230)에서 아동/청소년 대상 디지털 성범죄 콘텐츠로 분류되면, 해당 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 이미지의 캡처를 수행할 수 있다.

[0085] 또한, 증거 추출부(241)는 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 등의 증거를 수집하여 채증을 수행할 수 있다.

[0086] 또한, 증거 추출부(241)는 채증을 통해 수집한 증거를 기반으로 해당 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 신고 데이터와 해당 아동/청소년 대상 디지털 성범죄 콘텐츠의 삭제 요청 정보를 추가 생성하여 유해 콘텐츠를 게시한 온라인 서비스 서버와 사용자 단말(100)로 출력한다.

[0087] 삭제 감시부(242)는 성인 대상 디지털 성범죄 콘텐츠와 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 삭제 요청 정보를 기반으로 해당 삭제 요청에 대응한 성인 대상 디지털 성범죄 콘텐츠와 아동/청소년 대상 디지털 성범죄 콘텐츠의 삭제가 수행되었는지에 대한 삭제 결과와, 콘텐츠의 재배포 및 재확산 상태에 대한 모니터링을 수행한다.

[0088] 이를 통해, 삭제 감시부(242)는 피해자의 신고가 없는 아동/청소년 대상 불법 디지털 성범죄 콘텐츠에 대한 선제적인 삭제가 이루어질 수 있도록 동작함으로써, 해당 콘텐츠의 확산, 유포 및 재유포를 방지할 수 있다.

[0089] 데이터베이스(250)는 피해 영상물 관련 검색 결과 및 증거 정보, 성인 대상 디지털 성범죄 콘텐츠 및 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 검색 결과 및 해당 콘텐츠들의 증거 정보와 삭제 관련 처리 정보를 저장한다.

[0091] 다음은 본 발명의 일 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 방법을 설명한다.

[0092] 도 5는 본 발명의 일 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 방법을 설명하기 위해 나타낸 흐름도이고, 도 6은 도 5의 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 방법의 증거 채증 및 삭제 과정을 설명하기 위해 나타낸 흐름도이다.

[0093] 도 1 내지 도6을 참조하면, 본 발명의 일 실시 예에 따른 인공지능을 이용한 디지털 성범죄 콘텐츠 감시 방법은 콘텐츠 감시 서버(200)가 사용자 단말(100)로부터 임의의 검색 대상 온라인 서비스 서버 정보, 임의의 프롬프트 데이터, 피해자에 대한 피해 영상물 데이터를 수신(S100)한다.

[0094] 콘텐츠 감시 서버(200)는 수신된 프롬프트 데이터를 임베딩하여 복수의 키워드를 추출하고, 추출된 키워드를 기반으로 검색 대상 온라인 서비스 서버에 접속하여 임의의 콘텐츠를 수집(S200)한다.

[0095] S200 단계에서, 콘텐츠 감시 서버(200)는 수신된 검색 대상 웹 서버(300) 및 SNS 서버(400)를 대상으로 웹 크롤링을 통해 유해 콘텐츠 및 피해 영상물 관련 콘텐츠를 수집할 수 있다.

[0096] 또한, S200 단계에서 콘텐츠 감시 서버(200)는 프롬프트 데이터를 임베딩하여 키워드를 추출하되, GPT(Generative Pre-trained Transformer)를 이용한 군집 데이터 분석과 데이터 임베딩(Embedding)을 통해 복수의 키워드를 추출함으로써, 더욱 많은 유해 콘텐츠 및 피해 관련 영상 콘텐츠를 수집할 수 있다.

[0097] 또한, S200 단계에서 콘텐츠 감시 서버(200)는 사용자 단말(100)로부터 입력되는 검색 대상 온라인 서비스 서버 이외에, 추출된 키워드를 기반으로 검색 대상 온라인 서비스 서버의 정보를 추출하고, 해당 온라인 서비스에 접속하여 웹 크롤링을 통해 유해 콘텐츠 및 피해 관련 영상 콘텐츠를 수집할 수 있다.

[0098] 계속해서, 콘텐츠 감시 서버(200)는 수집된 콘텐츠로부터 텍스트 데이터와 이미지 데이터를 추출하고, 추출된 텍스트 데이터 및 이미지 데이터를 기반으로 게시글의 작성빈도, 작성자 및 유포 빈도를 포함한 해당 콘텐츠의 패턴을 분석하여 유해 콘텐츠로 분류되는 성인 대상 디지털 성범죄 콘텐츠 또는 아동/청소년 대상 디지털 성범죄 콘텐츠를 검출(S300)한다.

[0099] 즉, S300 단계에서, 콘텐츠 감시 서버(200)는 웹 크롤링을 통해 수집된 콘텐츠로부터 텍스트 데이터와 이미지 데이터를 추출하고, 추출된 텍스트 데이터 및 이미지 데이터를 기반으로 게시글의 작성빈도, 작성자 및 유포 빈도를 포함한 해당 콘텐츠의 패턴을 분석하여 유해 콘텐츠를 분류할 수 있다.

[0100] S300 단계에서, 콘텐츠 감시 서버(200)는 유해 콘텐츠의 이미지로부터 피해자의 얼굴을 인식하고, 피해자의 얼굴에 기반한 얼굴 형상, 눈, 코, 입의 좌우 대칭 특성, 각도 및 기울기 특성에 따른 분류를 통해 피해자의 나이와 성별을 예측할 수 있다.

[0101] 이를 통해, 콘텐츠 감시 서버(200)는 예측 결과에 따라 성인 대상 디지털 성범죄 콘텐츠 또는 아동/청소년 대상 디지털 성범죄 콘텐츠로 분류할 수 있다.

[0102] 계속해서, S300 단계에서 분류된 유해 콘텐츠가 성인 대상 디지털 성범죄 콘텐츠이면, 콘텐츠 감시 서버(200)는 유해 콘텐츠에 포함된 텍스트, 음성 및 이미지를 인공지능 기반의 분석 모델을 이용하여 해당 유해 콘텐츠와 피해 영상물 데이터를 비교한 유사도 값을 기반으로 신고된 피해 영상물인지 분석(S400)할 수 있다.

[0103] S400 단계에서 콘텐츠 감시 서버(200)는 광학문자인식(OCR)을 통해 유해 콘텐츠에 포함된 텍스트 데이터를 인식하고, 인식된 유해 콘텐츠의 텍스트 데이터에 대하여 텍스트 값들을 벡터 공간에 위치시켜 벡터 값으로 변환한 임베딩 벡터를 각 벡터의 위치에 기반하여 키워드 정보와의 행렬 연산을 통해 유사도에 따른 유사도 값을 산출할 수 있다.

[0104] S400 단계에서 콘텐츠 감시 서버(200)는 텍스트 데이터에 대한 유사도 분석을 통해 게시글의 작성 빈도, 게시글 작성자, 유포 빈도와 같은 게시글의 패턴을 분석할 수도 있다.

[0105] S400 단계에서 콘텐츠 감시 서버(200)는 학습된 인공지능 기반의 분석 모델을 이용하여 유해 콘텐츠에 포함된 음성을 분석하고, 분석된 음성의 특징 벡터를 추출하여 추출된 음성의 특징 벡터로부터 음향적 특성과 언어적 특성에 기반한 피해 영상물 데이터와의 유사도에 따른 유사도 값을 산출할 수 있다.

[0106] S400 단계에서 콘텐츠 감시 서버(200)는 학습된 인공지능 기반의 분석 모델을 이용하여 유해 콘텐츠에 포함된 이미지를 분석하고, 분석된 이미지의 특징 벡터를 추출하여 추출된 이미지의 특징 벡터로부터 이미지 특성에 기반한 피해 영상물 데이터와의 유사도에 따른 유사도 값을 산출할 수 있다.

[0107] 계속해서, 콘텐츠 감시 서버(200)는 유해 콘텐츠의 분석 결과를 데이터베이스(250)에 저장하고, 해당 유해 콘텐츠에 대한 증거 채증과 삭제가 이루어질 수 있도록 관리(S500)한다.

[0108] S500 단계에서, 콘텐츠 감시 서버(200)는 산출된 유사도 값이 일정 값 이상이면 유해 콘텐츠에 대한 이미지 캡처와, 해당 유해 콘텐츠와 관련된 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 중 하나 이상의 증거를 수집하여 채증(S510)한다.

[0109] 또한, 콘텐츠 감시 서버(200)는 채증을 통해 수집한 증거를 기반으로 해당 유해 콘텐츠에 대한 신고 데이터와 해당 유해 콘텐츠의 삭제 요청 정보를 생성하고, 생성된 신고 데이터와 삭제 요청 정보를 유해 콘텐츠를 게시한 온라인 서비스 서버와 사용자 단말(100)로 출력(S520)한다.

[0110] 또한, 콘텐츠 감시 서버(200)는 성인 대상 디지털 성범죄 콘텐츠에 대한 삭제 요청 정보를 기반으로 해당 삭제 요청에 대응한 성인 대상 디지털 성범죄 콘텐츠의 삭제가 수행되었는지에 대한 삭제 결과와, 콘텐츠의 재배포 및 재확산 상태에 대한 모니터링을 수행(S530)한다.

[0111] 또한, 콘텐츠 감시 서버(200)는 모니터링 결과를 분석하여 삭제 상태, 재배포 및 재확산 상태에 대한 결과를 임의의 포맷에 따라 결과 보고서를 생성하여 사용자 단말(100)로 리포팅(S540)할 수 있다.

[0112] 한편, S300 단계에서, 유해 콘텐츠가 아동/청소년 대상 디지털 성범죄 콘텐츠로 분류되면, 콘텐츠 감시 서버(200)는 해당 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 이미지의 캡처를 수행할 수 있다.

[0113] 또한, 콘텐츠 감시 서버(200)는 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 게시제목, 피해 영상물, URL, 표준시각, 게시자, 게시일 등의 증거를 수집하여 채증을 수행할 수 있다.

[0114] 또한, 콘텐츠 감시 서버(200)는 채증을 통해 수집한 증거를 기반으로 해당 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 신고 데이터와 해당 아동/청소년 대상 디지털 성범죄 콘텐츠의 삭제 요청 정보를 추가 생성하여 유해 콘텐츠를 게시한 온라인 서비스 서버와 사용자 단말(100)로 출력한다.

[0115] 또한, 콘텐츠 감시 서버(200)는 아동/청소년 대상 디지털 성범죄 콘텐츠에 대한 삭제 요청 정보를 기반으로 해

당 삭제 요청에 대응한 아동/청소년 대상 디지털 성범죄 콘텐츠의 삭제가 수행되었는지에 대한 삭제 결과와, 콘텐츠의 재배포 및 재확산 상태에 대한 모니터링을 수행하고, 모니터링 결과를 분석하여 삭제 상태, 재배포 및 재확산 상태에 대한 결과를 임의의 포맷에 따라 결과 보고서를 생성하여 사용자 단말(100)로 리포팅할 수 있다.

[0116] 따라서, 온라인상에 게시된 불법 디지털 성범죄 콘텐츠를 추적, 검색 및 삭제되도록 관리하여 피해를 최소화시키고 피해 지원관의 업무 효율 개선과 정신적 부담을 감소시킬 수 있다.

[0117] 또한, 피해자의 신고가 없는 아동/청소년 대상 불법 디지털 성범죄 콘텐츠에 대한 선제적 삭제를 수행함으로써, 해당 콘텐츠의 확산, 유포 및 재유포를 방지할 수 있다.

[0118] 상기와 같이, 본 발명의 바람직한 실시 예를 참조하여 설명하였지만 해당 기술 분야의 숙련된 당업자라면 하기의 특허청구범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

[0119] 또한, 본 발명의 특허청구범위에 기재된 도면번호는 설명의 명료성과 편의를 위해 기재한 것일 뿐 이에 한정되는 것은 아니며, 실시예를 설명하는 과정에서 도면에 도시된 선들의 두께나 구성요소의 크기 등은 설명의 명료성과 편의상 과장되게 도시되어 있을 수 있다.

[0120] 또한, 상술된 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례에 따라 달라질 수 있으므로, 이러한 용어들에 대한 해석은 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.

[0121] 또한, 명시적으로 도시되거나 설명되지 아니하였다 하여도 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명의 기재사항으로부터 본 발명에 의한 기술적 사상을 포함하는 다양한 형태의 변형을 할 수 있음은 자명하며, 이는 여전히 본 발명의 권리범위에 속한다.

[0122] 또한, 첨부하는 도면을 참조하여 설명된 상기의 실시예들은 본 발명을 설명하기 위한 목적으로 기술된 것이며 본 발명의 권리범위는 이러한 실시예에 국한되지 아니한다.

부호의 설명

[0123] 100 : 사용자 단말

200 : 콘텐츠 감시 서버

210 : 데이터 수집부

220 : 데이터 검출부

230 : 데이터 분석부

231 : 유사도 분석부

232 : 예측 분석부

240 : 채증/삭제 관리부

241 : 증거 추출부

242 : 삭제 감시부

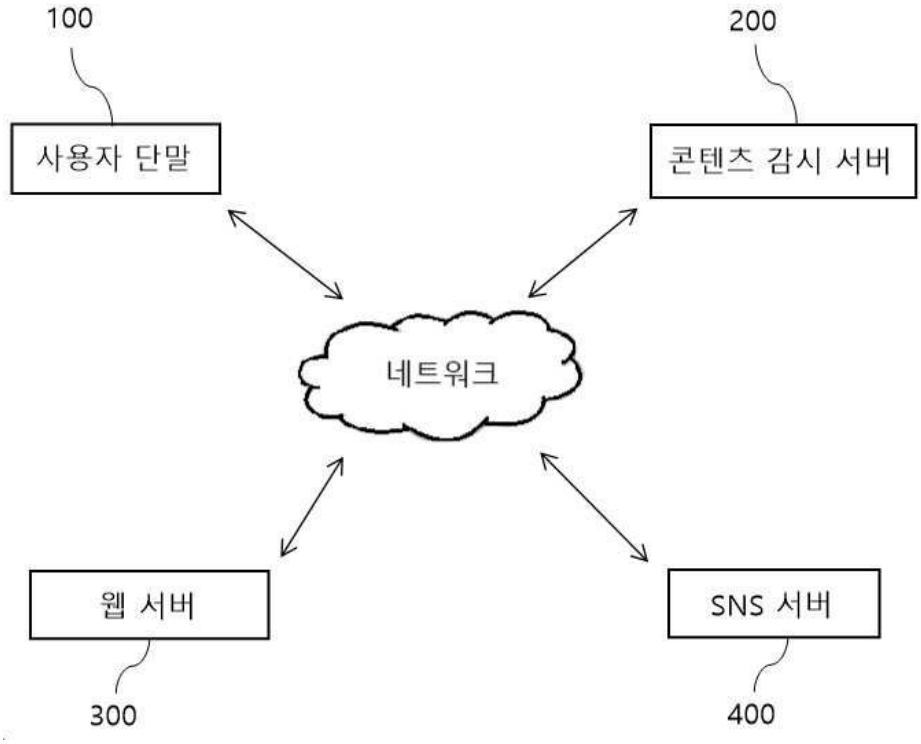
250 : 데이터 베이스

300 : 웹 서버

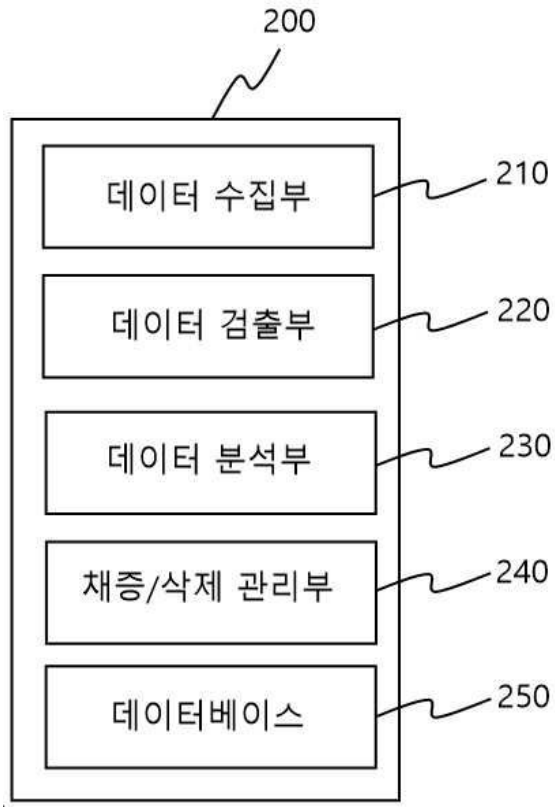
400 : SNS 서버

도면

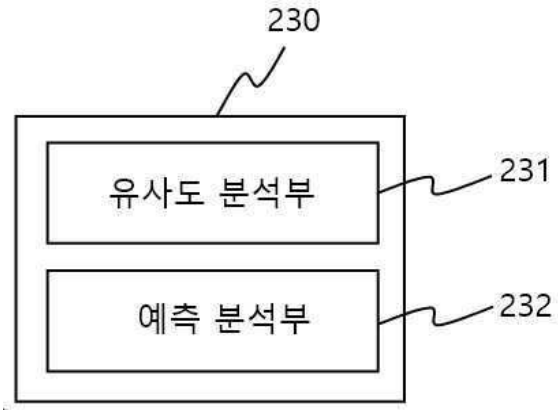
도면1


100
사용자 단말
200
콘텐츠 감시 서버
네트워크
웹 서버
300
SNS 서버
400

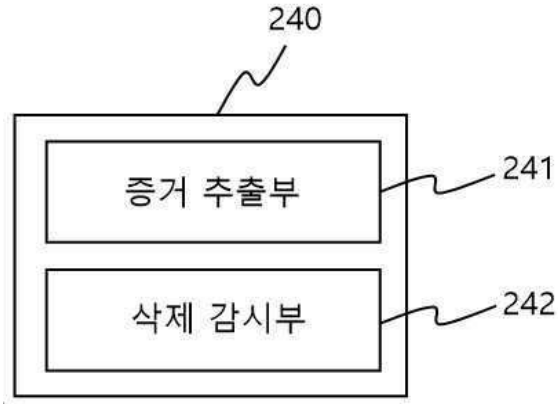
도면2


200
데이터 수집부
210
데이터 검출부
220
데이터 분석부
230
채증/삭제 관리부
240
데이터베이스
250

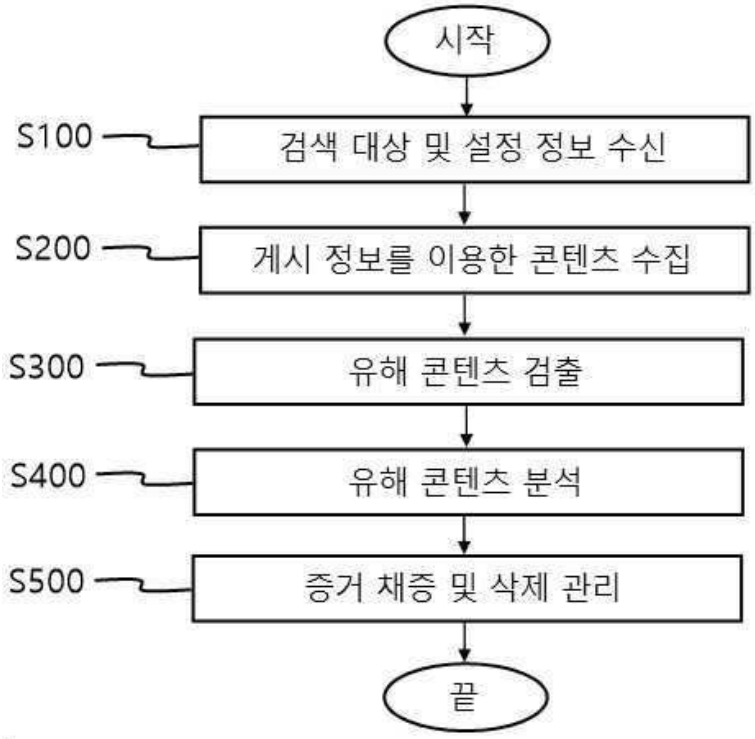
도면3


230
유사도 분석부
231
예측 분석부
232

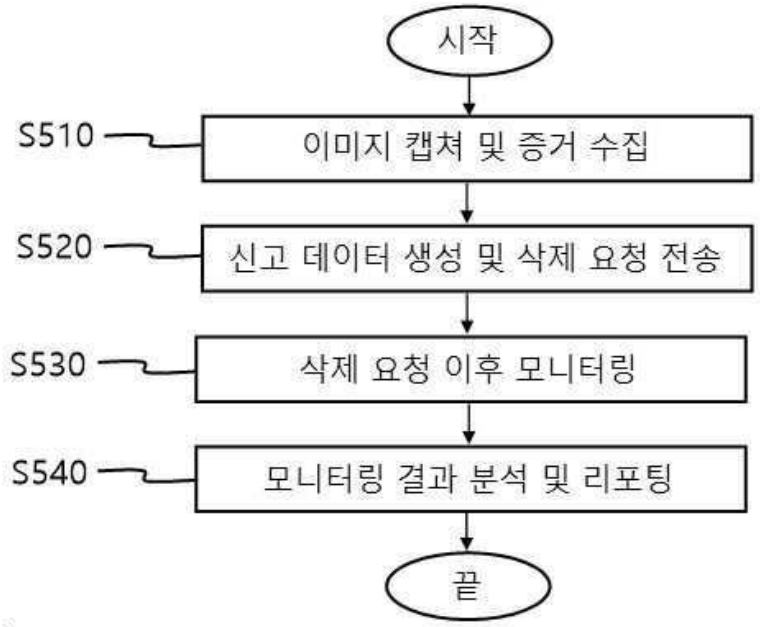
도면4


240
증거 추출부
241
삭제 감시부
242

도면5


시작
S100
검색 대상 및 설정 정보 수신
S200
게시 정보를 이용한 콘텐츠 수집
S300
유해 콘텐츠 검출
S400
유해 콘텐츠 분석
S500
증거 채증 및 삭제 관리
끝

도면6


시작
S510
이미지 캡쳐 및 증거 수집
S520
신고 데이터 생성 및 삭제 요청 전송
S530
삭제 요청 이후 모니터링
S540
모니터링 결과 분석 및 리포팅
끝