



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2020년11월17일  
(11) 등록번호 10-2179552  
(24) 등록일자 2020년11월10일

(51) 국제특허분류(Int. Cl.)  
HO4M 1/725 (2006.01) G06K 9/00 (2006.01)  
G06T 7/11 (2017.01)  
(52) CPC특허분류  
HO4M 1/72522 (2013.01)  
G06K 9/00402 (2013.01)  
(21) 출원번호 10-2019-0056743  
(22) 출원일자 2019년05월15일  
심사청구일자 2019년05월15일  
(56) 선행기술조사문헌  
KR101462289 B1\*  
KR1020110099494 A\*  
KR1020190021066 A\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
주식회사 한컴위드  
경기도 성남시 분당구 대왕판교로644번길 49 ,  
9층(삼평동, 한컴타워)  
(72) 발명자  
김현수  
경기도 성남시 분당구 수내로 201, 408동 1002호  
(분당동, 샛별마을삼부아파트)  
이경수  
경기도 안산시 상록구 본오로 145, 105동 702호(  
본오동, 우성아파트)  
(뒷면에 계속)  
(74) 대리인  
김연권

전체 청구항 수 : 총 13 항

심사관 : 안병일

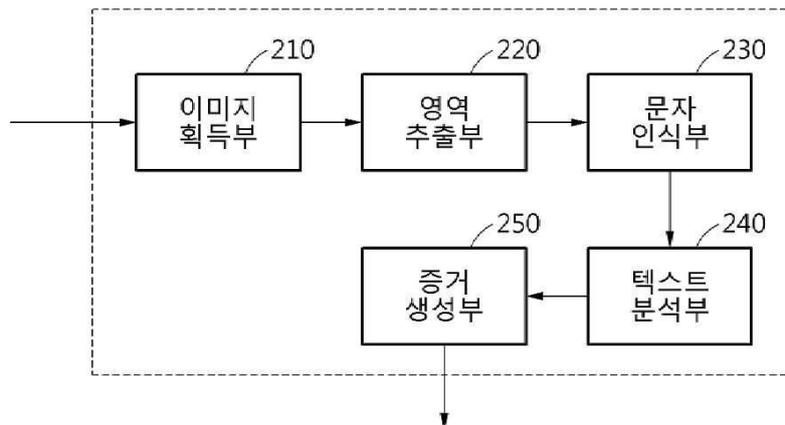
(54) 발명의 명칭 OCR 기반의 증거 수집 장치 및 그 방법

(57) 요약

증거 수집 장치 및 그 방법에 관한 것으로서, 일실시에 따른 증거 수집 장치는 스마트 기기의 메신저 앱(Application; App)으로부터 증거 이미지를 획득하는 이미지 획득부와, 획득된 증거 이미지로부터 복수의 유효 UI(User Interface) 영역을 추출하는 영역 추출부와, 추출된 유효 UI 영역에 대응되는 텍스트 데이터(Text Data)를 추출하기 위해 OCR(Optical Character Recognition)을 수행하는 문자 인식부 및 추출된 텍스트 데이터에 포함된 비정형(Unstructured) 텍스트를 텍스트 분석(Text Analysis)을 통해 정형 텍스트로 변환하는 텍스트 분석부를 포함한다.

대표도 - 도2

200



(52) CPC특허분류

*G06T 7/11* (2017.01)

*G06K 2209/01* (2013.01)

(72) 발명자

**이용우**

경기도 화성시 동탄대로22길 30, 609동 1902호(영  
천동, 동탄센트럴자이)

---

**드 파사 프라팀**

경기도 수원시 영통구 효원로 363, 131동 1303호(  
매탄동, 매탄 위브 하늘채)

## 명세서

### 청구범위

#### 청구항 1

스마트 기기의 메신저 앱(Application; App)으로부터 증거 이미지를 획득하는 이미지 획득부;

상기 획득된 증거 이미지로부터 복수의 유효 UI(User Interface) 영역을 추출하는 영역 추출부;

상기 추출된 유효 UI 영역에 대응되는 텍스트 데이터(Text Data)를 추출하기 위해 OCR(Optical Character Recognition)을 수행하는 문자 인식부 및

상기 추출된 텍스트 데이터에 포함된 비정형(Unstructured) 텍스트를 텍스트 분석(Text Analysis)을 통해 정형 텍스트로 변환하는 텍스트 분석부

를 포함하고,

상기 영역 추출부는,

상기 스마트 기기의 모델명 정보, 상기 스마트 기기의 운영체제(Operating System; OS) 정보 및 상기 메신저 앱의 종류에 관한 정보에 대응되는 상기 복수의 유효 UI 영역을 추출하며,

상기 복수의 유효 UI 영역은 수신 UI 영역 및 송신 UI 영역을 포함하는

증거 수집 장치.

#### 청구항 2

제1항에 있어서,

상기 이미지 획득부는,

상기 메신저 앱의 대화 내용을 촬영하여 상기 증거 이미지를 생성하는 카메라 장치와 유선 또는 무선 통신으로 연결되어, 상기 카메라 장치로부터 상기 증거 이미지를 획득하는

증거 수집 장치.

#### 청구항 3

제1항에 있어서,

상기 이미지 획득부는,

상기 스마트 기기와 유선 또는 무선 통신을 통해 연결되면, 상기 메신저 어플의 대화 내용을 자동으로 캡처하여 상기 증거 이미지를 획득하는

증거 수집 장치.

#### 청구항 4

삭제

#### 청구항 5

제1항에 있어서,

상기 텍스트 분석부는,

기저장된 모델에 구비된 복수의 패턴 정보 중 상기 비정형 텍스트에 대응되는 패턴 정보에 기초하여 상기 비정형 텍스트를 상기 정형 텍스트로 변환하는

증거 수집 장치.

**청구항 6**

제5항에 있어서,  
 상기 텍스트 분석부는,  
 상기 추출된 텍스트 데이터를 이용하여 상기 기저장된 모델에 포함된 복수의 패턴 정보를 트레이닝(Training)하는  
 증거 수집 장치.

**청구항 7**

제5항에 있어서,  
 상기 텍스트 분석부는,  
 상기 비정형 텍스트에 대응되는 패턴 정보와 리벤슈타인 거리(Levenshtein Distance) 알고리즘을 이용한 어휘화된 트리(Lexicalized Tree) 분석을 통해, 상기 비정형 텍스트를 상기 정형 텍스트로 변환하는  
 증거 수집 장치.

**청구항 8**

제1항에 있어서,  
 상기 변환된 정형 텍스트를 포함하는 텍스트 데이터에 기초하여 기설정된 형식의 증거 분석 데이터를 생성하는  
 증거 생성부  
 를 더 포함하는 증거 수집 장치.

**청구항 9**

제8항에 있어서,  
 상기 증거 생성부는,  
 상기 변환된 정형 텍스트를 포함하는 텍스트 데이터의 대화 내용을 분석하여, 상기 메신저 앱 사용자와 다른 사용자 사이의 관계 분석(Relationship Analysis) 결과가 반영된 상기 증거 분석 데이터를 생성하는  
 증거 수집 장치.

**청구항 10**

이미지 획득부에서, 스마트 기기의 메신저 앱(Application; App)으로부터 증거 이미지를 획득하는 단계;  
 영역 추출부에서, 상기 획득된 증거 이미지로부터 복수의 유효 UI(User Interface) 영역을 추출하는 단계;  
 문자 인식부에서, 상기 추출된 유효 UI 영역에 대응되는 텍스트 데이터를 추출하기 위해 OCR(Optical Character Recognition)을 수행하는 단계 및  
 텍스트 분석부에서, 상기 추출된 텍스트 데이터에 포함된 비정형(Unstructured) 텍스트를 텍스트 분석(Text Analysis)을 통해 정형 텍스트로 변환하는 단계  
 를 포함하고,  
 상기 복수의 유효 UI 영역을 추출하는 단계는,  
 상기 스마트 기기의 모델명 정보, 상기 스마트 기기의 운영체제(Operating System; OS) 정보 및 상기 메신저 앱의 종류에 관한 정보에 대응되는 상기 복수의 유효 UI 영역을 추출하며,  
 상기 복수의 유효 UI 영역은 수신 UI 영역 및 송신 UI 영역을 포함하는  
 증거 수집 방법.

**청구항 11**

삭제

**청구항 12**

제10항에 있어서,

상기 정형 텍스트로 변환하는 단계는,

기저장된 모델에 구비된 복수의 패턴 정보 중 상기 비정형 텍스트에 대응되는 패턴 정보에 기초하여 상기 비정형 텍스트를 상기 정형 텍스트로 변환하는

증거 수집 방법.

**청구항 13**

제12항에 있어서,

상기 정형 텍스트로 변환하는 단계는,

상기 추출된 텍스트 데이터를 이용하여 상기 기저장된 모델에 포함된 패턴 정보를 트레이닝(Training)하는

증거 수집 방법.

**청구항 14**

제12항에 있어서,

상기 정형 텍스트로 변환하는 단계는,

상기 비정형 텍스트에 대응되는 패턴 정보와 리벤슈타인 거리(Levenshtein Distance) 알고리즘을 이용한 어휘화된 트리(Lexicalized Tree) 분석을 통해, 상기 비정형 텍스트를 상기 정형 텍스트로 변환하는

증거 수집 방법.

**청구항 15**

제10항에 있어서,

증거 생성부에서, 상기 변환된 정형 텍스트를 포함하는 텍스트 데이터에 기초하여 기설정된 형식의 증거 분석 데이터를 생성하는 단계

를 더 포함하는 증거 수집 방법.

**발명의 설명**

**기술 분야**

[0001] 증거 수집 장치 및 그 방법에 관한 것으로, 보다 상세하게는 OCR 기술에 기초하여 캡처된 이미지로부터 증거를 수집하는 기술적 사상에 관한 것이다.

**배경 기술**

[0002] 통신 및 인터넷 기술의 발달에 따라 스마트폰과 같은 스마트 기기가 대중화 되었으며, 이로 인해 스마트 기기를 통해 입수된 디지털 증거물 데이터를 법원에 제출하는 사례가 증가하고 있다.

[0003] 그러나, 스마트 기기는 암호화 프로토콜을 사용하는 앱이 많기 때문에 실제 현장에서 앱으로부터 디지털 증거물을 추출하는데 어려움을 겪고 있다.

[0004] 일례로, 암호화 프로토콜이 적용된 메신저 앱의 대화 내용을 디지털 증거물로 추출하기 위해서는 대상 스마트 기기에서 메신저 앱을 활성화 하여 대화 내용을 디스플레이 한 후, 디스플레이 화면을 카메라로 촬영하여 증거물 데이터를 획득해야 한다.

[0005] 그러나, 상술한 방법으로 획득된 증거물 데이터는 이미지 데이터이기 때문에 실제 디지털 증거물로서 필요한 유의미한 데이터(대화 내용)를 사람이 일일이 수기로 옮겨 적는 작업을 수행해야 한다는 번거로움이 있다.

**선행기술문헌**

**특허문헌**

[0006] (특허문헌 0001) 한국등록특허 제10-1462289호, "모바일 장치 시스템을 이용한 디지털 이미지 아카이빙 및 검색"

**발명의 내용**

**해결하려는 과제**

[0007] 본 발명은 스마트 기기의 모델명 정보, 스마트 기기의 운영체제 정보 및 메신저 앱의 종류에 관한 정보를 고려하여 유효 UI 영역을 추출함으로써, 유의미한 데이터를 보다 용이하게 획득할 수 있는 증거 수집 장치 및 그 방법을 제공하고자 한다.

[0008] 또한, 본 발명은 비정형 텍스트에 대응되는 패턴 정보와 리벤슈타인 거리 알고리즘을 이용한 어휘화된 트리 분석을 통해 비정형 텍스트 변환의 정확도를 향상시킬 수 있는 증거 수집 장치 및 그 방법을 제공하고자 한다.

[0009] 또한, 본 발명은 증거 이미지를 가공하여 기설정된 형식의 증거 분석 데이터를 자동 생성함으로써, 디지털 증거 수집의 편의성을 향상시킬 수 있는 증거 수집 장치 및 그 방법을 제공하고자 한다.

**과제의 해결 수단**

[0010] 일실시에에 따른 증거 수집 장치는 스마트 기기의 메신저 앱(Application; App)으로부터 증거 이미지를 획득하는 이미지 획득부와, 획득된 증거 이미지로부터 복수의 유효 UI(User Interface) 영역을 추출하는 영역 추출부와, 추출된 유효 UI 영역에 대응되는 텍스트 데이터(Text Data)를 추출하기 위해 OCR(Optical Character Recognition)을 수행하는 문자 인식부 및 추출된 텍스트 데이터에 포함된 비정형(Unstructured) 텍스트를 텍스트 분석(Text Analysis)을 통해 정형 텍스트로 변환하는 텍스트 분석부를 포함할 수 있다.

[0011] 일측에 따르면, 이미지 획득부는 메신저 앱의 대화 내용을 촬영하여 증거 이미지를 생성하는 카메라 장치와 유선 또는 무선 통신으로 연결되어, 카메라 장치로부터 증거 이미지를 획득할 수 있다.

[0012] 일측에 따르면, 이미지 획득부는 스마트 기기와 유선 또는 무선 통신을 통해 연결되면, 메신저 어플의 대화 내용을 자동으로 캡처하여 증거 이미지를 획득할 수 있다.

[0013] 일측에 따르면, 영역 추출부는 스마트 기기의 모델명 정보, 스마트 기기의 운영체제(Operating System; OS) 정보 및 메신저 앱의 종류에 관한 정보 중 적어도 하나 이상의 정보에 대응되는 유효 UI 영역을 추출할 수 있다.

[0014] 일측에 따르면, 텍스트 분석부는 기저장된 모델에 구비된 복수의 패턴 정보 중 비정형 텍스트에 대응되는 패턴 정보에 기초하여 비정형 텍스트를 정형 텍스트로 변환할 수 있다.

[0015] 일측에 따르면, 텍스트 분석부는 추출된 텍스트 데이터를 이용하여 기저장된 모델에 포함된 복수의 패턴 정보를 트레이닝(Training)할 수 있다.

[0016] 일측에 따르면, 텍스트 분석부는 비정형 텍스트에 대응되는 패턴 정보와 리벤슈타인 거리(Levenshtein Distance) 알고리즘을 이용한 어휘화된 트리(Lexicalized Tree) 분석을 통해, 비정형 텍스트를 정형 텍스트로 변환할 수 있다.

[0017] 일측에 따르면, 일실시에에 따른 증거 수집 장치는 변환된 정형 텍스트를 포함하는 텍스트 데이터에 기초하여 기설정된 형식의 증거 분석 데이터를 생성하는 증거 생성부를 더 포함할 수 있다.

[0018] 일측에 따르면, 증거 생성부는 변환된 정형 텍스트를 포함하는 텍스트 데이터의 대화 내용을 분석하여, 메신저 앱 사용자와 다른 사용자 사이의 관계 분석(Relationship Analysis) 결과가 반영된 증거 분석 데이터를 생성할 수 있다.

- [0019] 일실시예에 따른 증거 수집 방법은 이미지 획득부에서, 스마트 기기의 메신저 앱(Application; App)으로부터 증거 이미지를 획득하는 단계와, 영역 추출부에서, 획득된 증거 이미지로부터 복수의 유효 UI(User Interface) 영역을 추출하는 단계와, 문자 인식부에서, 추출된 유효 UI 영역에 대응되는 텍스트 데이터를 추출하기 위해 OCR(Optical Character Recognition)을 수행하는 단계 및 텍스트 분석부에서, 추출된 텍스트 데이터에 포함된 비정형(Unstructured) 텍스트를 텍스트 분석(Text Analysis)을 통해 정형 텍스트로 변환하는 단계를 포함할 수 있다.
- [0020] 일측에 따르면, 복수의 유효 UI 영역을 추출하는 단계는, 스마트 기기의 모델명 정보, 스마트 기기의 운영체제(Operating System; OS) 정보 및 메신저 앱의 종류에 관한 정보 중 적어도 하나 이상의 정보에 대응되는 유효 UI 영역을 추출할 수 있다.
- [0021] 일측에 따르면, 정형 텍스트로 변환하는 단계는, 기저장된 모델에 구비된 복수의 패턴 정보 중 비정형 텍스트에 대응되는 패턴 정보에 기초하여 비정형 텍스트를 정형 텍스트로 변환할 수 있다.
- [0022] 일측에 따르면, 정형 텍스트로 변환하는 단계는, 추출된 텍스트 데이터를 이용하여 기저장된 모델에 포함된 패턴 정보를 트레이닝(Training)할 수 있다.
- [0023] 일측에 따르면, 정형 텍스트로 변환하는 단계는, 비정형 텍스트에 대응되는 패턴 정보와 리벤슈타인 거리(Levenshtein Distance) 알고리즘을 이용한 어휘화된 트리(Lexicalized Tree) 분석을 통해, 비정형 텍스트를 정형 텍스트로 변환할 수 있다.
- [0024] 일측에 따르면, 일실시예에 따른 증거 수집 방법은 증거 생성부에서, 변환된 정형 텍스트를 포함하는 텍스트 데이터에 기초하여 기설정된 형식의 증거 분석 데이터를 생성할 수 있다.

**발명의 효과**

- [0025] 일실시예에 따르면, 스마트 기기의 모델명 정보, 스마트 기기의 운영체제 정보 및 메신저 앱의 종류에 관한 정보를 고려하여 유효 UI 영역을 추출함으로써, 유의미한 데이터를 보다 용이하게 획득할 수 있다.
- [0026] 일실시예에 따르면, 비정형 텍스트에 대응되는 패턴 정보와 리벤슈타인 거리 알고리즘을 이용한 어휘화된 트리 분석을 통해 비정형 텍스트 변환의 정확도를 향상시킬 수 있다.
- [0027] 일실시예에 따르면, 본 발명은 증거 이미지를 가공하여 기설정된 형식의 증거 분석 데이터를 자동 생성함으로써, 디지털 증거 수집의 편의성을 향상시킬 수 있다.

**도면의 간단한 설명**

- [0028] 도 1a 내지 도 1b는 증거 이미지로부터 텍스트 데이터를 추출하는 예시를 설명하기 위한 도면이다.  
 도 2는 일실시예에 따른 증거 수집 장치를 설명하기 위한 도면이다.  
 도 3a 내지 도 3c는 일실시예에 따른 증거 수집 장치에서 유효 UI 영역을 추출하는 실시예를 설명하기 위한 도면이다.  
 도 4는 일실시예에 따른 증거 수집 장치에서 텍스트 분석을 통해 비정형 텍스트를 변환하는 실시예를 설명하기 위한 도면이다.  
 도 5a 내지 도 5c는 일실시예에 따른 증거 수집 장치에서 어휘화된 트리 분석을 통해 비정형 텍스트를 변환하는 실시예를 설명하기 위한 도면이다.  
 도 6a 내지 도 6b는 일실시예에 따른 증거 수집 장치에서 증거 분석 데이터를 생성하는 실시예를 설명하기 위한 도면이다.  
 도 7은 일실시예에 따른 증거 수집 방법을 설명하기 위한 도면이다.

**발명을 실시하기 위한 구체적인 내용**

- [0029] 이하, 본 문서의 다양한 실시 예들이 첨부된 도면을 참조하여 기재된다.
- [0030] 실시 예 및 이에 사용된 용어들은 본 문서에 기재된 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 해당 실시 예의 다양한 변경, 균등물, 및/또는 대체물을 포함하는 것으로 이해되어야 한다.

- [0031] 하기에서 다양한 실시 예들을 설명에 있어 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다.
- [0032] 그리고 후술되는 용어들은 다양한 실시 예들에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [0033] 도면의 설명과 관련하여, 유사한 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다.
- [0034] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다.
- [0035] 본 문서에서, "A 또는 B" 또는 "A 및/또는 B 중 적어도 하나" 등의 표현은 함께 나열된 항목들의 모든 가능한 조합을 포함할 수 있다.
- [0036] "제1," "제2," "첫째," 또는 "둘째," 등의 표현들은 해당 구성요소들을, 순서 또는 중요도에 상관없이 수식할 수 있고, 한 구성요소를 다른 구성요소와 구분하기 위해 사용될 뿐 해당 구성요소들을 한정하지 않는다.
- [0037] 어떤(예: 제1) 구성요소가 다른(예: 제2) 구성요소에 "(기능적으로 또는 통신적으로) 연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 어떤 구성요소가 다른 구성요소에 직접적으로 연결되거나, 다른 구성요소(예: 제3 구성요소)를 통하여 연결될 수 있다.
- [0038] 본 명세서에서, "~하도록 구성된(또는 설정된)(configured to)"은 상황에 따라, 예를 들면, 하드웨어적 또는 소프트웨어적으로 "~에 적합한," "~하는 능력을 가지는," "~하도록 변경된," "~하도록 만들어진," "~를 할 수 있는," 또는 "~하도록 설계된"과 상호 호환적으로(interchangeably) 사용될 수 있다.
- [0039] 어떤 상황에서는, "~하도록 구성된 장치"라는 표현은, 그 장치가 다른 장치 또는 부품들과 함께 "~할 수 있는" 것을 의미할 수 있다.
- [0040] 예를 들면, 문구 "A, B, 및 C를 수행하도록 구성된(또는 설정된) 프로세서"는 해당 동작을 수행하기 위한 전용 프로세서(예: 임베디드 프로세서), 또는 메모리 장치에 저장된 하나 이상의 소프트웨어 프로그램들을 실행함으로써, 해당 동작들을 수행할 수 있는 범용 프로세서(예: CPU 또는 application processor)를 의미할 수 있다.
- [0041] 또한, '또는' 이라는 용어는 배타적 논리합 'exclusive or' 이기보다는 포함적인 논리합 'inclusive or' 를 의미한다.
- [0042] 즉, 달리 언급되지 않는 한 또는 문맥으로부터 명확하지 않는 한, 'x가 a 또는 b를 이용한다' 라는 표현은 포함적인 자연 순열들(natural inclusive permutations) 중 어느 하나를 의미한다.
- [0044] 상술한 구체적인 실시예들에서, 발명에 포함되는 구성 요소는 제시된 구체적인 실시 예에 따라 단수 또는 복수로 표현되었다.
- [0045] 그러나, 단수 또는 복수의 표현은 설명의 편의를 위해 제시한 상황에 적합하게 선택된 것으로서, 상술한 실시 예들이 단수 또는 복수의 구성 요소에 제한되는 것은 아니며, 복수로 표현된 구성 요소라 하더라도 단수로 구성되거나, 단수로 표현된 구성 요소라 하더라도 복수로 구성될 수 있다.
- [0046] 한편 발명의 설명에서는 구체적인 실시 예에 관해 설명하였으나, 다양한 실시 예들이 내포하는 기술적 사상의 범위에서 벗어나지 않는 한도 내에서 여러 가지 변형이 가능함은 물론이다.
- [0047] 그러므로 본 발명의 범위는 설명된 실시 예에 국한되어 정해져서는 아니되며 후술하는 청구범위뿐만 아니라 이 청구범위와 균등한 것들에 의해 정해져야 한다.
- [0049] 도 1a 내지 도 1b는 증거 이미지로부터 텍스트 데이터를 추출하는 예시를 설명하기 위한 도면이다.
- [0050] 도 1a 내지 도 1b를 참조하면, 참조부호 110은 일반적인 증거 수집 장치를 이용하여 증거 이미지로부터 텍스트 데이터를 추출하는 예시를 나타내고, 참조부호 120은 일실시예에 따른 증거 수집 장치를 이용하여 증거 이미지로부터 텍스트 데이터를 추출하는 예시를 나타낸다.
- [0051] 참조부호 110에 따르면, 일반적인 증거 수집 장치는 증거 이미지에 대하여 OCR(Optical Character Recognition)을 수행하여 텍스트 데이터(Text Data)를 추출할 수 있다.

- [0052] 예를 들면, 증거 이미지는 메신저 앱의 대화 내용을 포함하는 스마트 기기의 화면을 촬영한 이미지일 수 있다.
- [0053] 구체적으로, 일반적인 증거 수집 장치는 유의미한 데이터(대화 내용) 뿐만 아니라, 스마트 기기의 로고 및 이모티콘을 비롯하여 불필요한 UI(User Interface) 영역도 OCR이 수행되어, 유의미한 데이터를 선별하여 추출하는데 어려움이 있다.
- [0054] 그러나, 참조부호 120에 도시된 일실시예에 따른 증거 수집 장치는 증거 이미지로부터 복수의 유효 UI 영역을 추출하고, 복수의 유효 UI 영역에 대응되는 텍스트 데이터를 추출할 수 있다.
- [0055] 또한, 일실시예에 따른 증거 수집 장치는 추출된 텍스트 데이터에 포함된 비정형 텍스트를 텍스트 분석(Text Analysis)을 통해 정형 텍스트로 변환하고, 변환된 정형 텍스트를 포함하는 텍스트 데이터에 기초하여 기설정된 형식의 증거 분석 데이터를 생성하여 출력할 수 있다.
- [0056] 일실시예에 따른 증거 수집 장치는 이후 도 2를 통해 보다 구체적으로 설명하기로 한다.
- [0058] 도 2는 일실시예에 따른 증거 수집 장치를 설명하기 위한 도면이다.
- [0059] 도 2를 참조하면, 일실시예에 따른 증거 수집 장치(200)는 스마트 기기의 모델명 정보, 스마트 기기의 운영체제 정보 및 메신저 앱의 종류에 관한 정보를 고려하여 UI 영역을 추출함으로써, 유의미한 데이터를 보다 용이하게 획득할 수 있다.
- [0060] 또한, 증거 수집 장치(200)는 비정형 텍스트에 대응되는 패턴 정보와 리벤슈타인 거리 알고리즘을 이용한 어휘화된 트리 분석을 통해 비정형 텍스트 변환의 정확도를 향상시킬 수 있다.
- [0061] 또한, 증거 수집 장치(200)는 증거 이미지를 가공하여 기설정된 형식의 증거 분석 데이터를 자동 생성함으로써, 디지털 증거 수집의 편의성을 향상시킬 수 있다.
- [0062] 이를 위해, 일실시예에 따른 증거 수집 장치(200)는 이미지 획득부(210), 영역 추출부(220), 문자 인식부(230) 및 텍스트 분석부(240)를 포함할 수 있다.
- [0063] 또한, 증거 수집 장치(200)는 증거 생성부(250)를 더 포함할 수도 있다.
- [0064] 일실시예에 따른 이미지 획득부(210)는 스마트 기기의 메신저 앱(Application; App)으로부터 증거 이미지를 획득할 수 있다.
- [0065] 예를 들면, 스마트 기기는 스마트 폰 및 테블릿 PC 중 적어도 하나일 수 있으나, 일실시예에 따른 스마트 기기는 이에 한정되지 않고, 다양한 전자기기들을 포함할 수 있다.
- [0066] 또한, 메신저 앱은 스마트 기기에서 동작하는 기본 메신저 앱, 시그널(Signal), 텔레그램(Telegram), 라인(Line) 및 카카오톡(Kakao Talk) 중 적어도 하나일 수 있으나, 일실시예에 따른 메신저 앱은 이에 한정되지 않고, 스마트 기기에서 동작하는 다양한 메신저 앱을 포함할 수 있다.
- [0067] 일측에 따르면, 이미지 획득부(210)는 메신저 앱의 대화 내용을 촬영하여 증거 이미지를 생성하는 카메라 장치와 유선 또는 무선 통신으로 연결되어, 카메라 장치로부터 증거 이미지를 획득할 수 있다.
- [0068] 또한, 이미지 획득부(210)는 스마트 기기와 유선 또는 무선 통신을 통해 연결되면, 메신저 어플의 대화 내용을 자동으로 캡처하여 증거 이미지를 획득할 수도 있다.
- [0069] 예를 들면, 일실시예에 따른 증거 수집 장치(200)는 스마트 기기와 연결되어 스마트 기기의 현재 화면에 디스플레이된 메신저 어플의 대화 내용을 자동으로 캡처하여 증거 이미지를 획득할 수 있다.
- [0070] 일실시예에 따른 영역 추출부(220)는 이미지 획득부(210)로부터 획득된 증거 이미지로부터 복수의 유효 UI(User Interface) 영역을 추출할 수 있다.
- [0071] 예를 들면, 유효 UI 영역은 스마트 기기의 로고 및 이모티콘을 포함하는 불필요한 UI를 제외한 유의미한 데이터(예를 들면, 대화 내용)가 포함된 영역일 수 있다.
- [0072] 일측에 따르면, 영역 추출부(220)는 스마트 기기의 모델명 정보, 스마트 기기의 운영체제(Operating System; OS) 정보 및 메신저 앱의 종류에 관한 정보 중 적어도 하나 이상의 정보에 대응되는 유효 UI 영역을 추출할 수 있다.

- [0073] 예를 들면, 스마트 기기의 모델명 정보는 스마트 기기의 제조사 정보를 포함할 수 있다.
- [0074] 구체적으로, 증거 이미지 상에서의 유효 UI 영역은 스마트 기기의 모델명, 제조사, 운영체제 버전 및 메신저 앱의 종류에 따라 서로 다르게 선정될 수 있다.
- [0075] 즉, 일실시예에 따른 증거 수집 장치(200)는 스마트 기기의 모델명, 제조사, 운영체제 버전 및 메신저 앱의 종류에 따라 유효 UI 영역을 정확하게 추출하기 위한 정보를 사전에 수집 및 저장할 수 있다.
- [0076] 영역 추출부(220)에서 유효 UI 영역을 추출하는 예시는 이후 실시예 도 3a 내지 도 3c를 통해 보다 구체적으로 설명하기로 한다.
- [0077] 일실시예에 따른 문자 인식부(230)는 영역 추출부(220)를 통해 추출된 유효 UI 영역에 대응되는 텍스트 데이터(Text Data)를 추출하기 위해 OCR(Optical Character Recognition)을 수행할 수 있다.
- [0078] 예를 들면, 문자 인식부(230)는 일반적으로 사용되는 OCR 기술을 통해 유효 UI 영역에 대응되는 텍스트 데이터를 추출할 수 있다.
- [0079] 즉, 본 발명을 이용하면, 유효 UI 영역에 포함된 유의미한 데이터만을 OCR하여 텍스트 데이터를 추출함으로써, 불필요한 UI 영역을 자동 필터링할 수 있으며, 이를 통해 유의미한 데이터를 보다 용이하게 획득할 수 있다.
- [0080] 일실시예에 따른 텍스트 분석부(240)는 문자 인식부(230)를 통해 추출된 텍스트 데이터에 포함된 비정형(Unstructured) 텍스트를 텍스트 분석(Text Analysis)을 통해 정형 텍스트로 변환할 수 있다.
- [0081] 일측에 따르면, 텍스트 분석부(240)는 기저장된 모델에 구비된 복수의 패턴 정보 중 비정형 텍스트에 대응되는 패턴 정보에 기초하여 비정형 텍스트를 정형 텍스트로 변환할 수 있다.
- [0082] 예를 들면, 기설정된 모델은 공통 모델(Common Models)과 구성 모델(Configuration Models)로 구성될 수 있고, 패턴 정보는 헤더(Header) 패턴, 통화 기록(Call History) 패턴, 메시지(Message) 패턴 및 날짜/시간(Date/Time) 패턴으로 구성될 수 있으나, 일실시예에 따른 기설정된 모델 및 패턴 정보는 이에 한정되지 않고 다양한 형태로 구현될 수 있다.
- [0083] 일측에 따르면, 비정형 텍스트에 대응되는 패턴 정보는 비정형 텍스트를 입력으로 수신하여 비정형 텍스트를 정형 텍스트로 변환하는 모델일 수 있다.
- [0084] 또한, 텍스트 분석부(240)는 도메인 특정 엔진(Domain Specific Engine)을 구비할 수 있으며, 도메인 특정 엔진에서 비정형 텍스트에 대응되는 패턴 정보에 기초하여 비정형 텍스트를 정형 텍스트로 변환할 수도 있다.
- [0085] 보다 구체적인 예를 들면, 텍스트 분석부(240)는 추출된 텍스트 데이터가 헤더(Header) 패턴, 통화 기록(Call History) 패턴, 메시지(Message) 패턴 및 날짜/시간(Date/Time) 패턴 중 어느 하나에 대응되는지 여부를 판단하고, 판단된 대응 여부에 기초하여 비정형 텍스트를 정형 텍스트로 변환할 수 있다.
- [0086] 다시 말해, 텍스트 분석부(240)는 비정형 텍스트를 포함하는 텍스트 데이터가 날짜/시간 패턴에 대응되면, 비정형 텍스트를 날짜/시간 패턴정보에 기초하여 정형 텍스트로 변환할 수 있다.
- [0087] 일측에 따르면, 텍스트 분석부(240)는 추출된 텍스트 데이터를 이용하여 기저장된 모델에 포함된 복수의 패턴 정보를 트레이닝(Training)할 수 있다.
- [0088] 예를 들면, 텍스트 분석부(240)의 도메인 특정 엔진(Domain Specific Engine)은 추출된 텍스트 데이터를 입력으로 수신하여 기저장된 모델에 포함된 복수의 패턴 정보를 트레이닝할 수 있다.
- [0089] 보다 구체적인 예를 들면, 도메인 특정 엔진은 기계학습(Machine Learning)을 통해 복수의 패턴 정보를 트레이닝할 수도 있다.
- [0090] 일측에 따르면, 텍스트 분석부(240)는 비정형 텍스트에 대응되는 패턴 정보와 리벤슈타인 거리(Levenshtein Distance) 알고리즘을 이용한 어휘화된 트리(Lexicalized Tree) 분석을 통해, 비정형 텍스트를 정형 텍스트로 변환할 수 있다.
- [0091] 리벤슈타인 거리 알고리즘은 러시아 과학자 Vladimir Levenshtein의 이름에서 유래된 것으로 편집 거리 알고리즘이라고도 하고, 두 문자열의 유사도를 측정하기 위하여 고안된 알고리즘으로 철자검사, 음성인식, 표절검사 등에 이용될 수 있다.
- [0092] 리벤슈타인 거리 알고리즘은 2차원 배열을 이용하여 두 문자열을 비교하며 삽입, 삭제, 변경을 문자열 한부분마

다 진행하여 최소편집거리 값을 구할 수 있으며, 부분마다 구한 최소편집거리의 누적된 값이 두 문자열의 최종 편집거리 값이 되며, 그 값을 통하여 유사한 정도를 구하는 판단의 척도로 사용할 수 있다.

[0093] 일측에 따르면, 텍스트 분석부(240)는 비정형 텍스트가 영문 텍스트인지 아닌지에 따라 서로 다른 리벤슈타인 거리 알고리즘을 적용할 수 있다.

[0094] 예를 들면, 텍스트 분석부(240)는 비정형 텍스트가 영문 텍스트이면, 다이렉트 리벤슈타인 거리(Direct LD) 알고리즘을 적용하고, 비정형 텍스트가 영문 텍스트가 아니면 리벤슈타인 우선 거리(Levenshtein Prioritized Distance) 알고리즘을 적용할 수 있다.

[0095] 보다 구체적으로, 리벤슈타인 거리 알고리즘을 이용한 최종편집거리 값은 하기 수학적 식 1을 통해 도출될 수 있다.

[0096] [수학적 식 1]

$$lev_{a,b}(i, j) = \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0, \\ \min \begin{cases} lev_{a,b}(i-1, j) + 1 \\ lev_{a,b}(i, j-1) + 1 \\ lev_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

[0097]

[0098] 여기서, a와 i는 2차원 배열의 x축 방향의 값, b와 j는 y축 방향의 값일 수 있으며, a, b, i, j는 0 이상의 정수일 수 있다.

[0099] 보다 구체적인 예를 들면, 텍스트 분석부(240)는 비정형 텍스트인 'Saturday'라는 텍스트를 입력으로 수신하면, 기저장된 모델에 구비된 복수의 패턴 정보 중에서 비정형 텍스트에 대응되는 패턴 정보에 기초하여 비정형 텍스트인 'Saturday'를 'Sunday' 및 'Saturday'와 각각 비교할 수 있다.

[0100] 여기서, 'Saturday'와 'Sunday'라는 텍스트 사이의 최종편집거리 값은 하기 표 1을 통해 결정될 수 있다.

[0101] [표 1]

		S	a	t	u	n	d	a	y
	0	1	2	3	4	5	6	7	8
S	1	0	1	2	3	4	5	6	7
u	2	1	1	2	2	3	4	5	6
n	3	2	2	2	3	2	3	4	5
d	4	3	3	3	3	3	2	3	4
a	5	4	3	4	4	4	3	2	3
y	6	5	4	4	5	5	4	3	2

[0102]

[0103] 즉, 표 1 및 수학적 식 1에 따르면, 'Saturday'와 'Sunday'라는 텍스트를 비교하여 최종편집거리 값을 연산하는 리벤슈타인 거리 알고리즘은 비교 대상인 두 글자가 같은 값인 경우는 왼쪽 대각선 위쪽의 값을 가져오고 다른 값인 경우는 위쪽, 왼쪽, 왼쪽 대각선 위쪽의 값 중 가장 작은 값에 1을 더한 값을 가져오며, 알고리즘 동작이 종료되면 오른쪽 가장아래 부분의 값이 두 문자열의 최종편집거리 값으로 결정될 수 있다.

[0104] 즉, 'Saturday'와 'Sunday'라는 텍스트의 최종편집거리 값은 '2'로 결정될 수 있다. 마찬가지로 방법으로 'Saturday'와 'Saturday'라는 텍스트의 최종편집거리 값은 '1'로 결정될 수 있다.

[0105] 결국, 텍스트 분석부(240)는 'Saturday'와 'Sunday' 및 'Saturday'와 'Saturday' 라는 텍스트의 최종편집거리 값에 기초하여, 비정형 텍스트인 'Saturday'를 정형 텍스트인 'Saturday'로 변환할 수 있다.

[0106] 한편, 텍스트 분석부(240)는 비정형 텍스트인 '모주'라는 텍스트를 입력으로 수신하면, 하기 표 2와 같이 기저장된 모델에 구비된 복수의 패턴 정보 중에서 비정형 텍스트에 대응되는 패턴 정보에 기초하여 비정형 텍스트인

'모주'를 '오전' 및 '오후'와 각각 비교할 수 있다.

[0107] [표 2]

		□	⊥	⊗	⊤								
	0	2	3	4	5		0	2	3	4	5	6	
○	2	1	2	3	4		○	2	1	2	3	4	5
⊥	3	2	1	2	3		⊥	3	2	1	2	3	4
⊗	4	3	2	2	3		⊗	4	3	2	1	2	3
⊤	5	4	3	3	2		⊤	5	4	3	2	3	4
							⊥	6	5	4	3	4	5

Prioritized value

○	⊗	□	⊥	?		⊗	○	⊗	⊥	⊥	?
0	1	1	1	2		0	1	1	1	1	2

- [0108]
- [0109] 표 2 및 수학적 식 1에 따르면, '모주'와 '오전'이라는 텍스트의 최종편집거리 값은 '5'로 결정되고, '모주'와 '오후'라는 텍스트의 최종편집거리 값은 '2'로 결정될 수 있다.
- [0110] 결국, 텍스트 분석부(240)는 '모주'와 '오전' 및 '모주'와 '오후'라는 텍스트의 최종편집거리 값에 기초하여 비정형 텍스트인 '모주'를 정형 텍스트인 '오후'로 변환할 수 있다.
- [0111] 일실시예에 따른 텍스트 분석부(240)에서 비정형 텍스트를 변환하는 예시는 이후 실시예 도 4 내지 5c를 통해 보다 구체적으로 설명하기로 한다.
- [0112] 일측에 따르면, 증거 생성부(250)는 변환된 정형 텍스트를 포함하는 텍스트 데이터에 기초하여 기설정된 형식의 증거 분석 데이터를 생성할 수 있다.
- [0113] 다시 말해, 증거 생성부(250)는 텍스트 분석부(240)로부터 출력되는 정형 텍스트를 포함하는 텍스트 데이터를 수신할 수 있다.
- [0114] 예를 들면, 기설정된 형식의 증거 분석 데이터는 표 형식, 그래프 형식, 텍스트 형식 및 이미지 형식 중 어느 하나의 형식일 수 있으나, 일실시예에 따른 기설정된 형식의 증거 분석 데이터는 이에 한정되지 않고 다양한 형태의 보고서 형식이 적용될 수 있다.
- [0115] 즉, 증거 생성부(250)는 정형 텍스트를 포함하는 텍스트 데이터를 가공하여 기설정된 형식의 증거 분석 데이터를 자동 생성함으로써, 디지털 증거 수집의 편의성을 향상시킬 수 있다.
- [0116] 일측에 따르면, 증거 생성부(250)는 정형 텍스트를 포함하는 텍스트 데이터의 대화 내용을 분석하여, 메신저 앱 사용자와 다른 사용자 사이의 관계 분석(Relationship Analysis) 결과가 반영된 증거 분석 데이터를 생성할 수 있다.
- [0117] 일실시예에 따른 증거 생성부(250)에서 증거 분석 데이터를 생성하는 예시는 이후 실시예 도 6a 내지 도 6c를 통해 보다 구체적으로 설명하기로 한다.
- [0119] 도 3a 내지 도 3c는 일실시예에 따른 증거 수집 장치에서 유효 UI 영역을 추출하는 실시예를 설명하기 위한 도면이다.
- [0120] 다시 말해, 도 3a 내지 도 3c를 통해 설명하는 실시예는 도 2를 통해 설명한 영역 추출부에서 수행될 수 있다.
- [0121] 도 3a 내지 도 3c를 참조하면, 참조부호 310은 제1 내지 제3 메신저 앱(311 내지 313)의 UI(User Interface) 디자인을 나타내고, 참조부호 320은 제1 메신저 앱(311)에서 복수의 유효 UI 영역을 추출하는 예시를 나타내며, 참조부호 330은 제2 메신저 앱(312)에서 복수의 유효 UI 영역을 추출하는 예시를 나타낸다.
- [0122] 참조부호 310에 따르면, 서로 다른 종류의 제1 내지 제3 메신저(310 내지 330)는 서로 다른 형태의 UI 디자인이 적용되는 것을 알 수 있다.

- [0123] 즉, 일실시예에 따른 증거 수집 장치는 증거 이미지에서 복수의 유효 UI 영역을 보다 정확하게 추출하기 위해, 메신저 앱의 종류에 관한 정보에 따라 서로 다른 유효 UI 영역을 설정할 수 있다.
- [0124] 한편, UI 디자인은 스마트 기기의 모델 정보, 제조사 정보 및 운영체제에 따라 변경될 수도 있다. 즉, 일실시예에 따른 증거 수집 장치는 스마트 기기의 모델명 정보 및/또는 스마트 기기의 운영체제 정보에 따라 서로 다른 유효 UI 영역을 설정할 수도 있다.
- [0125] 참조부호 320에 따르면, 제1 메신저의 수신 UI 영역은 메시지 내용 및 메시지 내용 우측 하단에 수신 시각이 배치되는 UI 디자인을 나타내고, 송신 UI 영역은 메시지 내용 및 메시지 내용 우측 하단에 수신 확인 여부 및 발신 시각이 배치되는 UI 디자인을 나타낼 수 있다.
- [0126] 즉, 일실시예에 따른 증거 수집 장치는 제1 메신저의 UI 디자인을 고려하여, 수신 UI 영역과 송신 UI 영역을 유효 UI 영역으로 설정할 수 있다.
- [0127] 또한, 제1 메신저의 수신 UI 영역은 세부적으로 메시지 내용 영역 및 수신 시각 영역으로 구분될 수 있으며, 제1 메신저의 송신 UI 영역은 세부적으로 메시지 내용 영역, 발신 시각 영역 및 수신 확인 여부 영역으로 구분될 수 있다.
- [0128] 참조부호 330에 따르면, 제2 메신저의 수신 UI 영역은 메시지 내용과, 메시지 내용의 우측 하단에 수신 시각이 배치되는 UI 디자인을 나타내고, 송신 UI 영역은 메시지 내용과, 메시지 내용의 좌측 하단에 수신 확인 여부 및 발신 시각이 배치되는 UI 디자인을 나타낼 수 있다.
- [0129] 즉, 일실시예에 따른 증거 수집 장치는 제2 메신저의 UI 디자인을 고려하여, 수신 UI 영역과 송신 UI 영역을 유효 UI 영역으로 설정할 수 있다.
- [0130] 또한, 제2 메신저의 수신 UI 영역은 세부적으로 메시지 내용 영역 및 수신 시각 영역으로 구분될 수 있으며, 제2 메신저의 송신 UI 영역은 세부적으로 메시지 내용 영역, 발신 시각 영역 및 수신 확인 여부 영역으로 구분될 수 있다.
- [0132] 도 4는 일실시예에 따른 증거 수집 장치에서 텍스트 분석을 통해 비정형 텍스트를 변환하는 실시예를 설명하기 위한 도면이다.
- [0133] 다시 말해, 도 4를 통해 설명하는 실시예는 도 2를 통해 설명한 텍스트 분석부에서 수행될 수 있다.
- [0134] 도 4를 참조하면, 일실시예에 따른 증거 수집 장치는 수신한 증거 이미지(410)의 유효 UI 영역에 대응되는 텍스트 데이터를 추출하고, 추출된 텍스트 데이터에 대하여 OCR을 수행할 수 있다.
- [0135] 또한, 증거 수집장치는 OCR을 통해 추출된 텍스트 데이터에 포함된 비정형 텍스트(420)를 텍스트 분석을 통해 정형 텍스트(430)로 변환할 수 있다.
- [0137] 도 5a 내지 도 5c는 일실시예에 따른 증거 수집 장치에서 어휘화된 트리 분석을 통해 비정형 텍스트를 변환하는 실시예를 설명하기 위한 도면이다.
- [0138] 다시 말해, 도 5a 내지 도 5c를 통해 설명하는 실시예는 도 2를 통해 설명한 텍스트 분석부에서 수행될 수 있다.
- [0139] 도 5a 내지 도 5c를 참조하면, 참조부호 510은 '오후 1:35'라는 비정형 텍스트를 어휘화된 트리 분석을 통해 정형 텍스트로 변환하는 예시를 나타내고, 참조부호 520은 '오후 1137 7/'이라는 비정형 텍스트를 어휘화된 트리 분석을 통해 정형 텍스트로 변환하는 예시를 나타낸다.
- [0140] 또한, 참조부호 530은 '7월 31일 화요일'이라는 비정형 텍스트를 어휘화된 트리 분석을 통해 정형 텍스트로 변환하는 예시를 나타낸다.
- [0141] 구체적으로, 참조부호 510에 따르면, 일실시예에 따른 증거 수집 장치는 '오후 1:35'라는 비정형 텍스트를 포함하는 텍스트 데이터 기초하여 비정형 텍스트를 'Time' 및 'Receive' 영역으로 분류할 수 있으며, 기저장된 모델에 구비된 복수의 패턴 정보 중 분류된 영역 정보에 대응되는 패턴 정보를 추출할 수 있다.
- [0142] 또한, 일실시예에 따른 증거 수집 장치는 추출된 패턴 정보와 리벤슈타인 거리(Levenshtein distance) 알고리즘을 이용한 어휘화된 트리(Lexicalized Tree) 분석을 통해, 비정형 텍스트인 '오후 1:35'를 정형 텍스트로 변환할 수 있다.

- [0143] 예를 들면, 'Time' 영역에 대응되는 패턴 정보는 오전/오후 패턴 값, 시간(Hours) 패턴 값, 분(Minutes) 패턴 값 및 초(Seconds) 패턴 값으로 구성될 수 있다. 또한, 오전/오후 정보는 오전, 오후, AM 및 PM을 포함할 수 있다.
- [0144] 보다 구체적으로, 일실시예에 따른 증거 수집 장치는 비정형 텍스트인 '오후 1:35'를 '오후', '1', '35'로 분할하고, 분할된 값 각각을 토큰화(Tokenize)하며, 토큰화된 값들에 기초하여 추출된 패턴 정보와 리벤슈타인 거리 알고리즘을 이용한 분석을 수행함으로써, 비정형 텍스트인 '오후 1:35'를 정형 텍스트인 '13:35'로 변환할 수 있다.
- [0145] 일측에 따르면, 일실시예에 따른 증거 수집 장치는 복수의 패턴 정보를 트레이닝(Training)할 수 있으며, 트레이닝된 패턴 정보(Training data)와 리벤슈타인 거리 알고리즘에 기초한 퍼펙트 매치(Perfect match)를 통해 최종편집거리 값을 결정할 수 있다.
- [0146] 참조부호 520에 따르면, 일실시예에 따른 증거 수집 장치는 '오후 1137 7/'라는 비정형 텍스트를 포함하는 텍스트 데이터 기초하여 비정형 텍스트를 'Time' 및 'Send' 영역으로 분류할 수 있으며, 기저장된 모델에 구비된 복수의 패턴 정보 중 분류된 영역 정보에 대응되는 패턴 정보를 추출할 수 있다.
- [0147] 또한, 일실시예에 따른 증거 수집 장치는 추출된 패턴 정보와 리벤슈타인 거리(Levenshtein distance) 알고리즘을 이용한 어휘화된 트리(Lexicalized Tree) 분석을 통해, 비정형 텍스트인 '오후 1:35'를 정형 텍스트로 변환할 수 있다.
- [0148] 보다 구체적으로, 일실시예에 따른 증거 수집 장치는 비정형 텍스트인 '오후 1137 7/'를 '오후', '11', '37' 및 '7/'로 분할하여 토큰화하고, 토큰화된 값들에 기초하여 추출된 패턴 정보와 리벤슈타인 거리 알고리즘을 이용한 분석을 수행할 수 있다.
- [0149] 또한, 일실시예에 따른 증거 수집 장치는 비정형 텍스트인 '오후 1137 7/'를 '오후', '1', '37' 및 '7/'로 분할하여 토큰화하고, 토큰화된 값들에 기초하여 추출된 패턴 정보와 리벤슈타인 거리 알고리즘을 이용한 분석을 수행할 수 있다.
- [0150] 다시 말해, 일실시예에 따른 증거 수집 장치는 시간이 올바른 순서로 감지되지 않으면, 추출된 패턴 정보에 포함된 사전 제한 또한 사후 제한 정보에 따라 시간을 다시 분석하여 어휘화된 트리 분석을 재수행하는 순차적인 모델링(Sequential Modeling)을 수행할 수 있으며, 재수행 시에도 여전히 시간이 감지되지 않으면 오류 플래그를 출력할 수 있다.
- [0151] 일실시예에 따른 증거 수집 장치는 순차적인 모델링을 통해 비정형 텍스트인 '오후 1137 7/'를 정형 텍스트인 '23:37'로 변환할 수 있다.
- [0152] 참조부호 530에 따르면, 일실시예에 따른 증거 수집 장치는 '7월 31일 화요일'이라는 비정형 텍스트를 포함하는 텍스트 데이터 기초하여 비정형 텍스트를 'Date/Time' 영역으로 분류할 수 있으며, 기저장된 모델에 구비된 복수의 패턴 정보 중 분류된 영역 정보에 대응되는 패턴 정보를 추출할 수 있다.
- [0153] 또한, 일실시예에 따른 증거 수집 장치는 추출된 패턴 정보와 리벤슈타인 거리(Levenshtein distance) 알고리즘을 이용한 어휘화된 트리(Lexicalized Tree) 분석을 통해, 비정형 텍스트인 '7월 31일 화요일'을 정형 텍스트인 '2018/07/31'로 변환할 수 있다.
- [0154] 일측에 따르면, 일실시예에 따른 증거 수집 장치는 'Date/Time'에 대응되는 비정형 데이터의 날짜가 현재 날짜 이후의 시점이면 현재 연도 보다 이전 연도 데이터로 수정할 수 있다.
- [0155] 일측에 따르면, 참조부호 530에서 일실시예에 따른 증거 수집 장치는 참조부호 510 내지 520을 통해 설명한 방법과 동일한 방법으로 비정형 텍스트를 정형 텍스트로 변환할 수도 있다.
- [0157] 도 6a 내지 도 6b는 일실시예에 따른 증거 수집 장치에서 증거 분석 데이터를 생성하는 실시예를 설명하기 위한 도면이다.
- [0158] 다시 말해, 도 6a 내지 도 6b를 통해 설명하는 실시예는 도 2를 통해 설명한 증거 생성부에서 수행될 수 있다.
- [0159] 도 6a 내지 도 6b를 참조하면, 참조부호 610은 표의 형식으로 생성된 증거 분석 데이터를 나타내고, 참조부호 620은 관계 분석(Relationship Analysis) 결과가 반영된 증거 분석 데이터를 나타낸다.
- [0160] 구체적으로, 참조부호 610에 따르면, 일실시예에 따른 증거 수집 장치는 도 2의 텍스트 분석부로부터 수신한 텍

스트 데이터를 표 형식의 증거 분석 데이터로 가공하여 출력할 수 있다.

- [0161] 예를 들면, 텍스트 분석부로부터 수신한 텍스트 데이터는 복수의 메시지 내용(Content) 각각에 대응되는 인덱스(Index) 정보, 카테고리(Category) 정보, 타입(Type) 정보, Date/Time 정보, 파트너(Partner) 정보 및 통화 시간(Call Duration) 정보를 포함할 수 있다.
- [0162] 참조부호 620에 따르면, 일실시예에 따른 증거 수집 장치는 변환된 정형 텍스트를 포함하는 텍스트 데이터의 대화 내용을 분석하여, 메신저 앱 사용자와 다른 사용자 사이의 관계 분석(Relationship Analysis) 결과가 반영된 증거 분석 데이터를 생성할 수 있다.
- [0163] 즉, 일실시예에 따른 증거 수집 장치는 텍스트 분석부로부터 수신한 텍스트 데이터를 통해 인식된 대화 내용을 분석하여 메신저 앱 사용자의 이름(또는 아이디)와, 다른 사용자의 이름(또는, 아이디)간의 관계 분석 결과가 반영된 관계도를 증거 분석 데이터로서 출력할 수 있다.
- [0165] 도 7은 일실시예에 따른 증거 수집 방법을 설명하기 위한 도면이다.
- [0166] 다시 말해, 도 7은 도 1a 내지 도 6b를 통해 설명한 일실시예에 따른 증거 수집 장치를 이용하여 증거를 수집하는 방법을 설명하는 도면으로, 이후 도 7을 통해 설명하는 내용 중 일실시예에 따른 증거 수집 장치를 통해 설명한 내용과 중복되는 설명은 생략하기로 한다.
- [0167] 도 7을 참조하면, 710 단계에서 일실시예에 따른 증거 수집 방법은 이미지 획득부에서 스마트 기기의 메신저 앱(Application; App)으로부터 증거 이미지를 획득할 수 있다.
- [0168] 다음으로, 720 단계에서 일실시예에 따른 증거 수집 방법은 영역 추출부에서 획득된 증거 이미지로부터 복수의 유효 UI(User Interface) 영역을 추출할 수 있다.
- [0169] 일측에 따르면, 720 단계에서 일실시예에 따른 증거 수집 방법은 스마트 기기의 모델명 정보, 스마트 기기의 운영체제(Operating System; OS) 정보 및 메신저 앱의 종류에 관한 정보 중 적어도 하나 이상의 정보에 대응되는 유효 UI 영역을 추출할 수 있다.
- [0170] 다음으로, 730 단계에서 일실시예에 따른 증거 수집 방법은 문자 인식부에서 추출된 유효 UI 영역에 대응되는 텍스트 데이터를 추출하기 위해 OCR(Optical Character Recognition)을 수행할 수 있다.
- [0171] 다음으로, 740 단계에서 일실시예에 따른 증거 수집 방법은 텍스트 분석부에서 추출된 텍스트 데이터에 포함된 비정형(Unstructured) 텍스트를 텍스트 분석(Text Analysis)을 통해 정형 텍스트로 변환할 수 있다.
- [0172] 일측에 따르면, 740 단계에서 일실시예에 따른 증거 수집 방법은 기저장된 모델에 구비된 복수의 패턴 정보 중 비정형 텍스트에 대응되는 패턴 정보에 기초하여 비정형 텍스트를 정형 텍스트로 변환할 수 있다.
- [0173] 일측에 따르면, 740 단계에서 일실시예에 따른 증거 수집 방법은 추출된 텍스트 데이터를 이용하여 기저장된 모델에 포함된 패턴 정보를 트레이닝(Training)할 수 있다.
- [0174] 또한, 740 단계에서 일실시예에 따른 증거 수집 방법은 비정형 텍스트에 대응되는 패턴 정보와 리벤슈타인 거리(Levenshtein distance) 알고리즘을 이용한 어휘화된 트리(Lexicalized Tree) 분석을 통해, 비정형 텍스트를 정형 텍스트로 변환할 수 있다.
- [0175] 일측에 따르면, 750 단계에서 일실시예에 따른 증거 수집 방법은 증거 생성부에서 변환된 정형 텍스트를 포함하는 텍스트 데이터에 기초하여 기설정된 형식의 증거 분석 데이터를 생성할 수 있다.
- [0177] 결국, 본 발명을 이용하면, 스마트 기기의 모델명 정보, 스마트 기기의 운영체제 정보 및 메신저 앱의 종류에 관한 정보를 고려하여 UI 영역을 추출함으로써, 유의미한 데이터를 보다 용이하게 획득할 수 있다.
- [0178] 또한, 비정형 텍스트에 대응되는 패턴 정보와 리벤슈타인 거리 알고리즘을 이용한 어휘화된 트리 분석을 통해 비정형 텍스트 변환의 정확도를 향상시킬 수 있다.
- [0179] 또한, 증거 이미지를 가공하여 기설정된 형식의 증거 분석 데이터를 자동 생성함으로써, 디지털 증거 수집의 편의성을 향상시킬 수 있다.
- [0181] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴

퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0182]

[0183] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embody)될 수 있다.

[0184] 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0185]

[0186] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

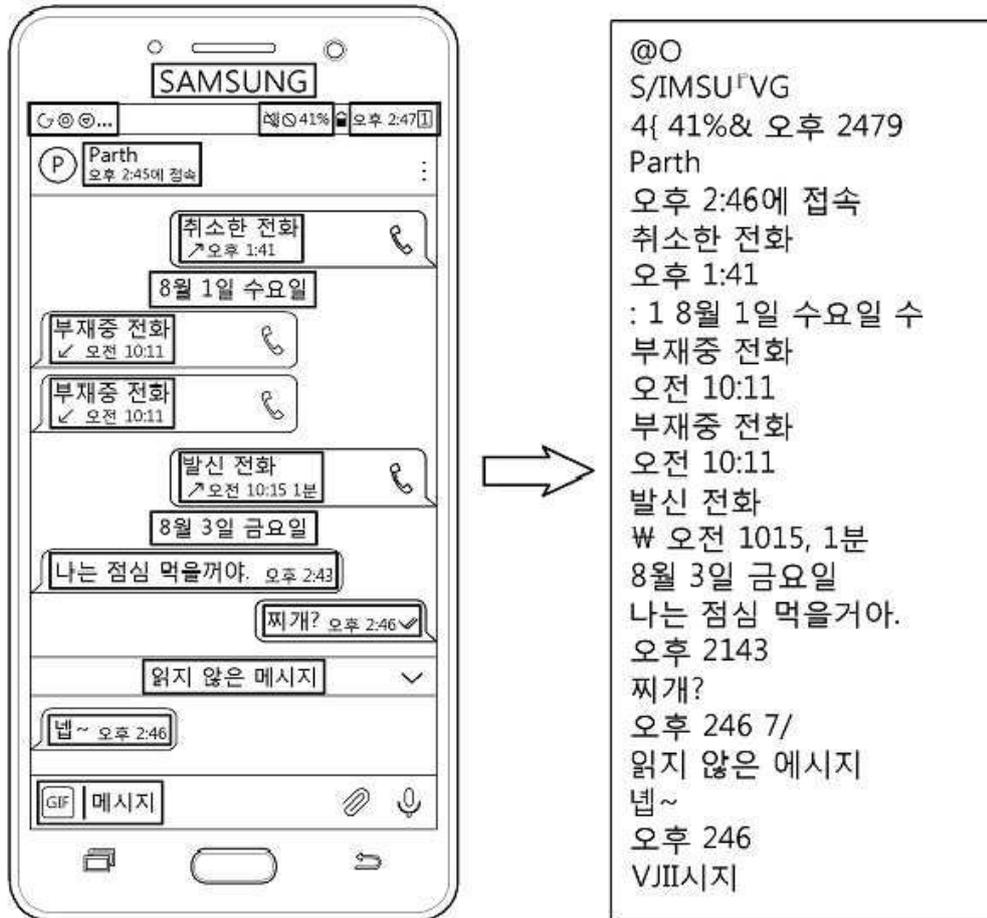
**부호의 설명**

- |        |               |              |
|--------|---------------|--------------|
| [0187] | 200: 증거 수집 장치 | 210: 이미지 획득부 |
|        | 220: 영역 추출부   | 230: 문자 인식부  |
|        | 240: 텍스트 분석부  | 250: 증거 생성부  |

도면

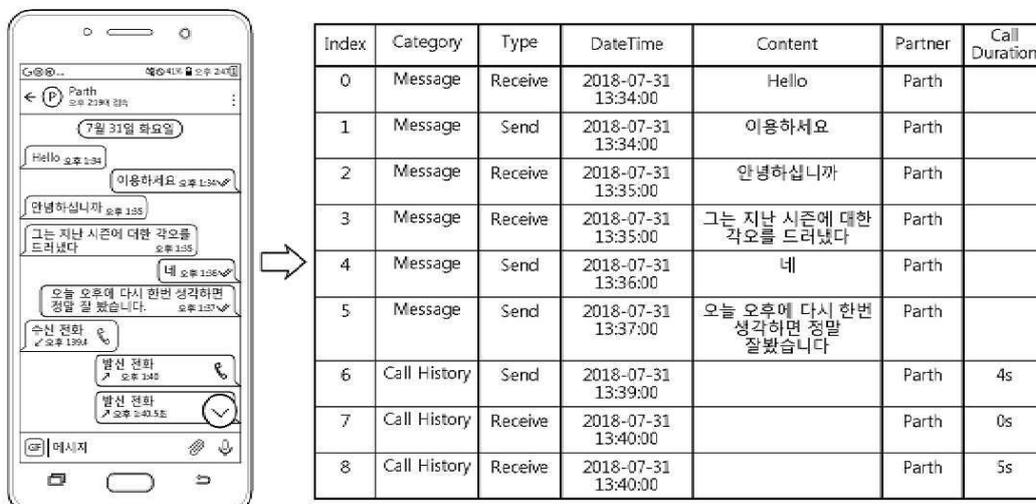
도면1a

110

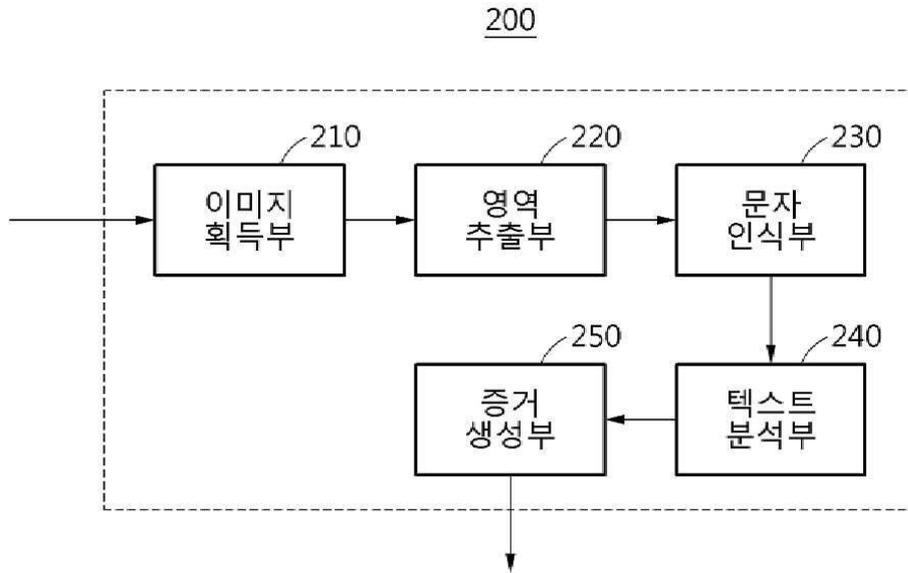


도면1b

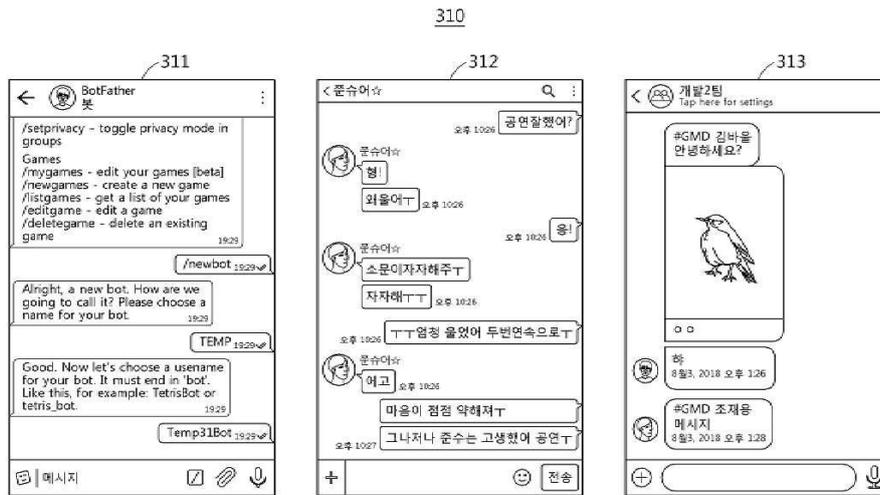
120



도면2

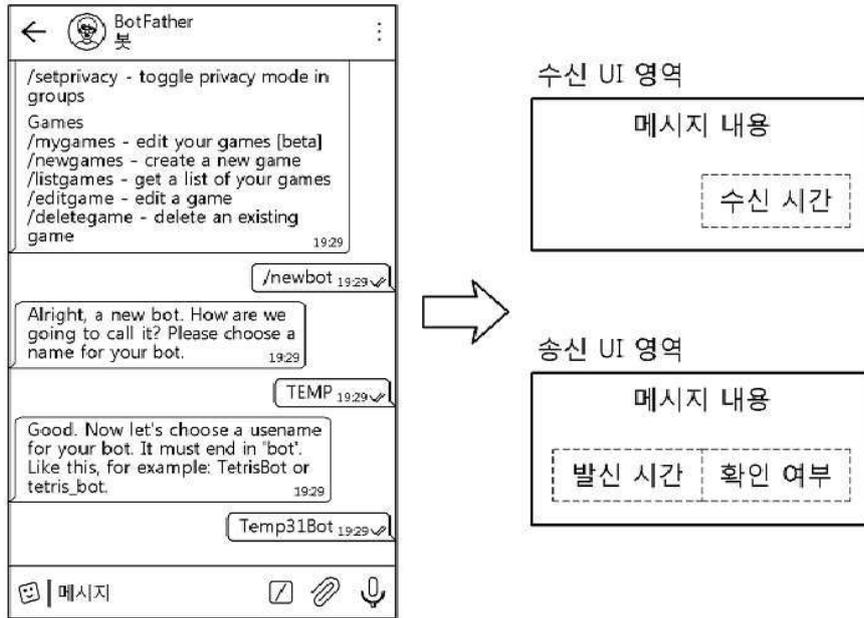


도면3a



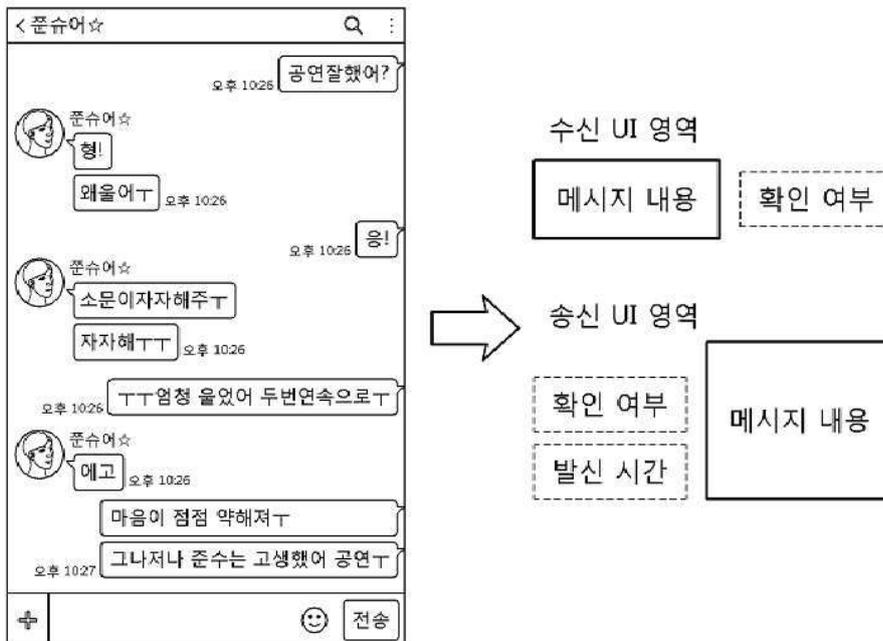
도면3b

320

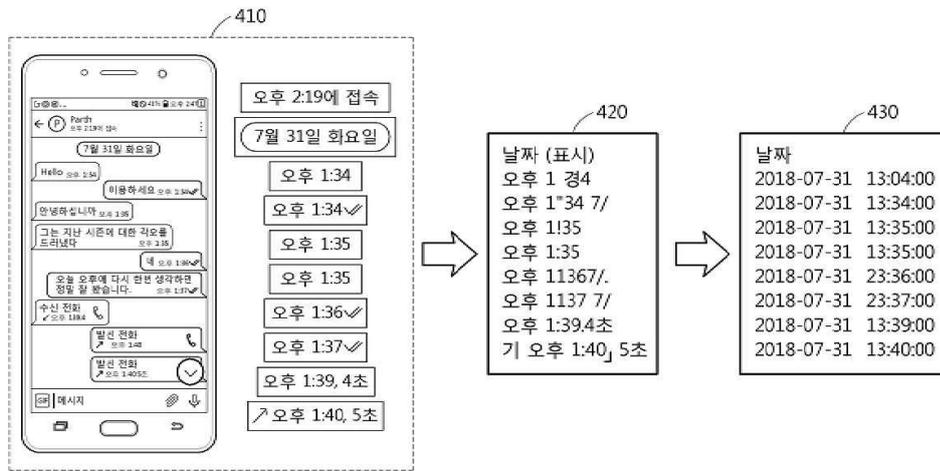


도면3c

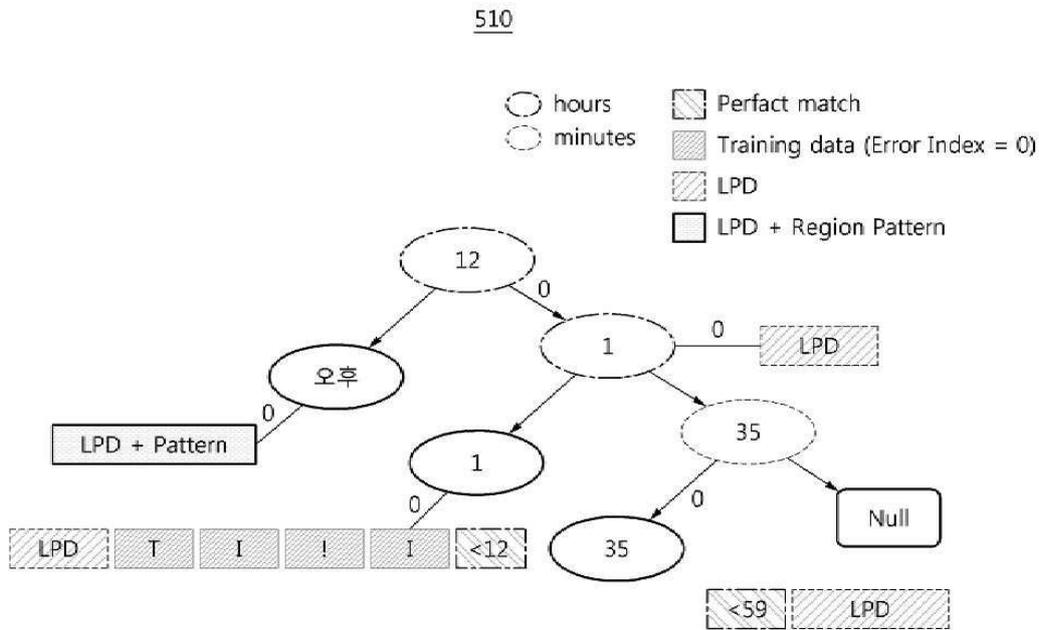
330



도면4

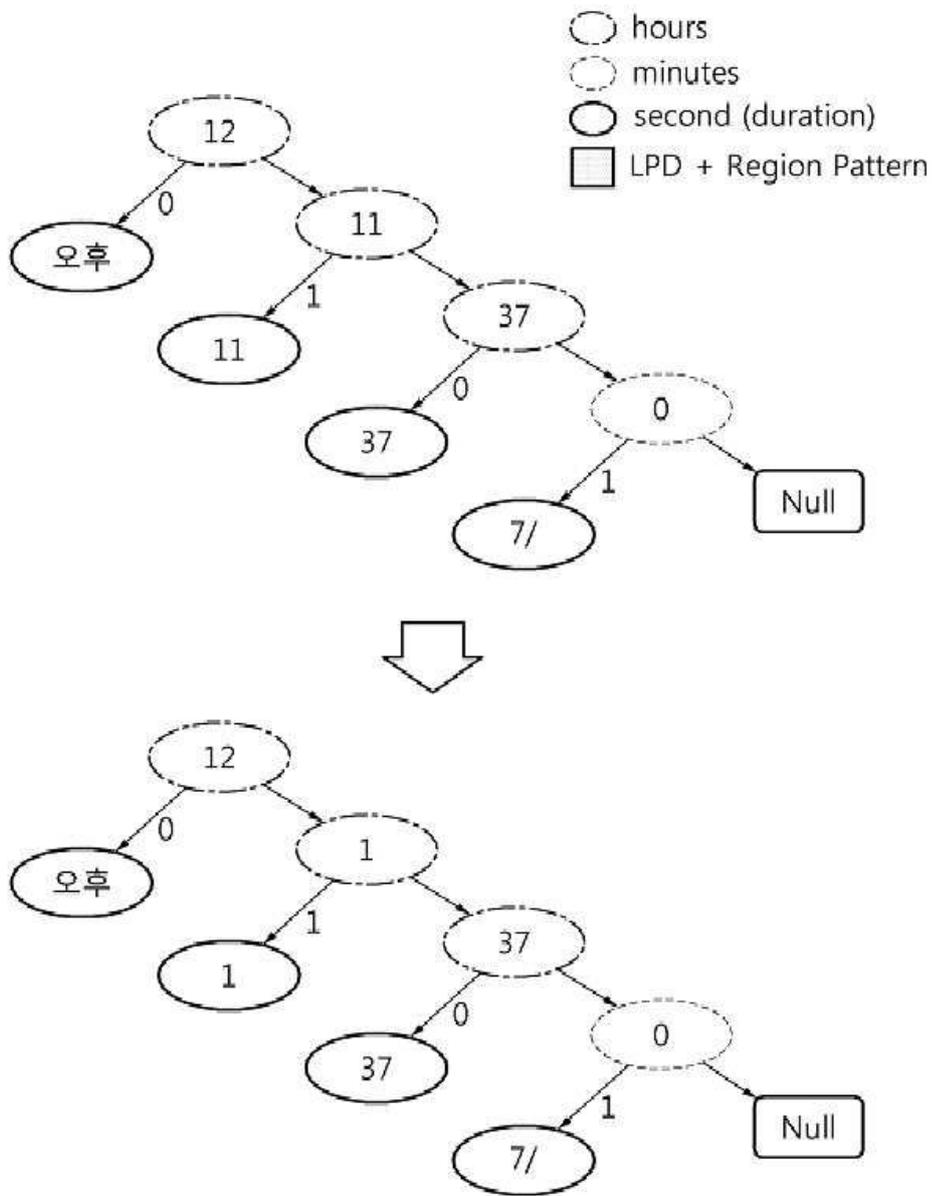


도면5a



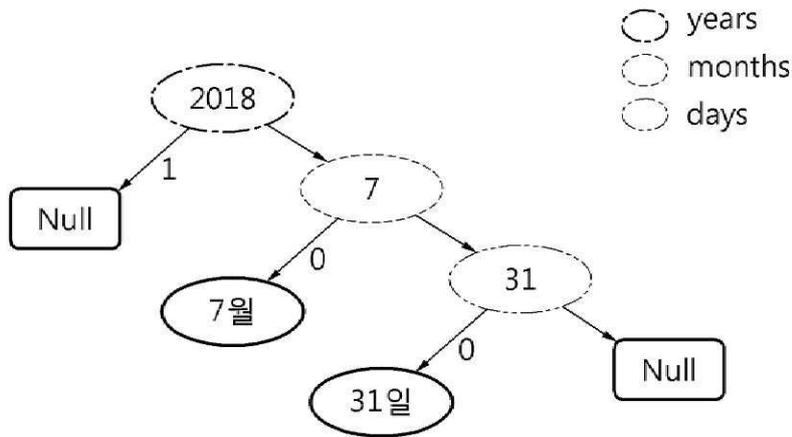
도면5b

520



도면5c

530

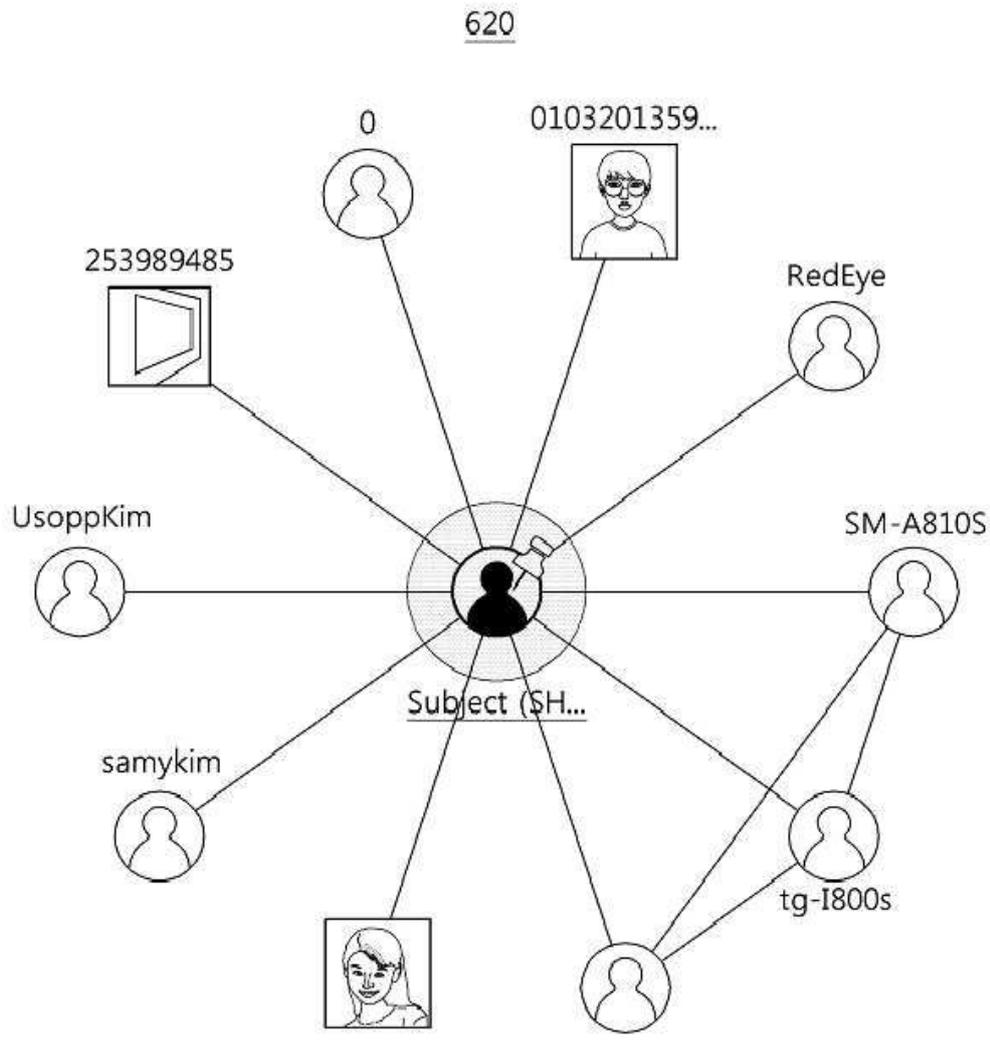


도면6a

610

Index	Category	Type	DateTime	Content	Partner	Call Duration
0	Message	Receive	2018-07-31 13:34:00	Hello	Parth	
1	Message	Send	2018-07-31 13:34:00	이용하세요	Parth	
2	Message	Receive	2018-07-31 13:35:00	안녕하십니까	Parth	
3	Message	Receive	2018-07-31 13:35:00	그는 지난 시즌에 대한 각오를 드러냈다	Parth	
4	Message	Send	2018-07-31 13:36:00	네	Parth	
5	Message	Send	2018-07-31 13:37:00	오늘 오후에 다시 한번 생각하면 정말 잘랐습니다	Parth	
6	Call History	Send	2018-07-31 13:39:00		Parth	4s
7	Call History	Receive	2018-07-31 13:40:00		Parth	0s
8	Call History	Receive	2018-07-31 13:40:00		Parth	5s

도면6b



도면7

