# 주요 범죄사실 개요 분석을 통한 범죄사실 타임라인 자동 작성 시스템 연구

정종진*, 박종빈**, 박성주*
*,**,*** 한국전자기술연구원 정보디이어연구센터
mozzalt@keti.re.kr, jpark@keti.re.kr, bpark@keti.re.kr

# A Study On An Automatic Crime Timeline Creation System Through An Overview Analysis Of Major Crime Facts

Jong-Jin, Jung*, Jong-Bin Park**, Sung-Ju Park***
*,**,*** Korea Electronics Technology Institute

요 약

본 논문에서 제안한 주요 범죄사실 개요 분석을 통한 범죄사실 타임라인 자동 작성 시스템은 범죄를 수사하는 현장 수사관들이 수사과정에서 취득한 수사문건 중 범죄사실 파트에 기재된 문장들을 자연어분석을 통해 주요 범죄사실을 파악하는데 필요한 요소들을 자동으로 알아내고, 이 요소들을 타임 라인 상에 시간 순으로 표시함으로써 주요범죄사실 개용을 쉽게 이해 할수 있도록 타임라인을 자동 만들어 내는 시스템에 관한 연구이다. 이를 위해 수사문건에 포함된 범죄사실 개요 범위를 자동으로 알아내고, 그 범위내 포함된 문장들 속에서 주요행위자, 관심을 둬야 하는 시각, 그 시각에 벌어진 주요 행위들을 자동으로 분석하는 연구를 소개한다.

## 1. 서론

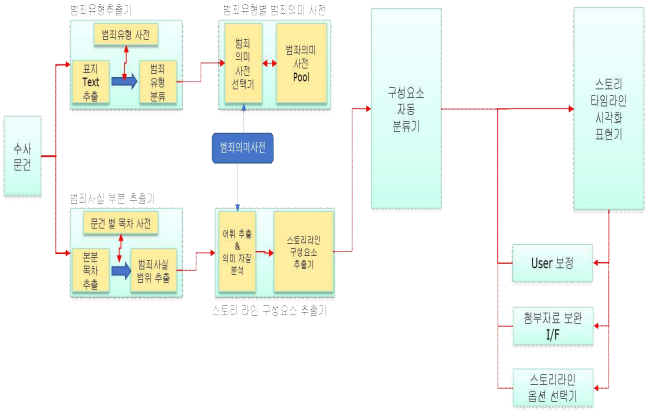
많은 현직 수사관들은 접수된 범죄사건의 개략적 개요를 파악하기 위해 확보된 자료를 토대로 직접 빈종이에 수기로 타임라인을 그려 주요 사건관계자, 주요 행위가 일어난 시점, 시점간 비교, 수사쟁점 등등을 최우선 파악한다고 한다. 이 경우 많은 시간이 걸릴뿐더러 현직경찰관 경력에 따라 많은 중요한 정보들이 포함되지 않을 수도 있어 이를 효과적으로 개선되었으면 하는 요구들이 적지 않다. 이런 현장 수요를 해결하기 위해 수사문건내 표함된 주요 범죄사실을 타임라인에 맞춰 생성하는 프로그램은 상용으로 소수 존재 하지만 프로그램을 이용한 타임라인 작성 자체를 직접 사람이 GUI를 통해 직접 그리다 시피 해야 하는 번거로움이 있어 정작 이를 사용하는 사람은 매우 드물고 기존에 하던대로 직접 빈종이에 그리는 사람이 많다고 한다. 따라서 분석해야 하는 사건문서를 대상으로 수사 대상인 범죄 내용에 대한 개요를 텍스트 분석을 통해 자동으로 시각화된 결과를 제시하여 사건을 종합적이고 입체적으로 파악함으로써 올바른 수사 방향 수립에 도움을 줄 필요가 있다.

## 2. 제안한 시스템의 동작 순서 및 구성

제안한 시스템이 동작하는 순서 및 구성은 아래의 설명 및 그림 1과 같다.

① 입력된 판결문 문서내에서 표지와 범죄사실이 기재된 부분을 발췌한다.
② 자연어 분석 (형태소 분석)을 통해 분석 대상인 text만을 다시 발췌한다.
③ 표지에서 발췌된 text 단어 들로부터 어떤 범죄유형인지를 우선 알아낸다.
④ 범죄유형이 자동 분류되면 범죄사실 구간에서 발췌된 많은 text들이 해당 죄종에서 어떤 의미를 갖는지를 알기 위해 죄종별 어휘사전을 바탕으로 의미를 알아낸다.
⑤ 이 과정에서 범죄사실과 관계가 깊은 주요행위자, 행위시간, 행위사실을 자연어 분석을 통해 도출한다.
⑥ 추출된 분석 내용중 세로의 타임축에 주요 행위시각을 의미하는 타임노드가 기록되고, 해당 타임노드에 누가 어떤 행위를 했었는지에 대해 분석한 결과를 시각적으로 표시한다.
⑦ 기본적으로는 행위자 수만큰 타임축을 구성하

여 동시에 스토리라인을 표출하여 전체 사실 흐름을 담은 시각화 결과 제시
⑧ 선택한 옵션에 따라 다양한 각도에서 사실 파악이 가능하도록 별도의 스토리라인 시각화 표현한다
⑨ 특정 타임노드별 행위자의 행위사실을 뒷받침할수 있는 증거자료, 사신, 수사관 메모들을 추가 기재 할 수 있게 하여 객관성을 담보 할 수 있도록 자료 보관 관리 한다.




(그림 1) 범죄사실 개요 및 스토리라인 시스템 동작 도

그림 1에 구성된 세부 모듈들은 차례로 알아본다.

## 2-1. 범죄유형 추출기

수사문건에 대한 개요가 작성된 text 문서로부터 추출한 핵심어 들 대상으로 국내법이 정의하고 있는 범죄유형에 맞춰 각 범죄유형별 특징들이 정의되어 있는 범죄유형 사전과 K-Means 값을 도출하여 가장 많이 분류되는 항목을 해당 문건의 범죄유형으로 도출한다. K-Mean 분류로 도출하는 과정은 아래와 식 (1)같다.

$$argmin \sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk} \| x_k - c_k \|^2 \quad \text{.....} \quad (식\ 1)$$

범죄유형 분류를 위해 수사문건 표지로부터 추출한 후보 단어들 N개 데이터들 X = ($x_1, x_2, \ldots, x_n$)에 대해 데이터가 속할 K개의 cluster(범죄유형) 중심점($c_1, c_2, \ldots, c_k$)간 차이가 최소화 되도록 한다.

이때 어느 특정 클러스터(특정 범죄유형)에 가장 분포되어 있는 범죄유형을 찾아 해당 범죄유형으로 정한다.

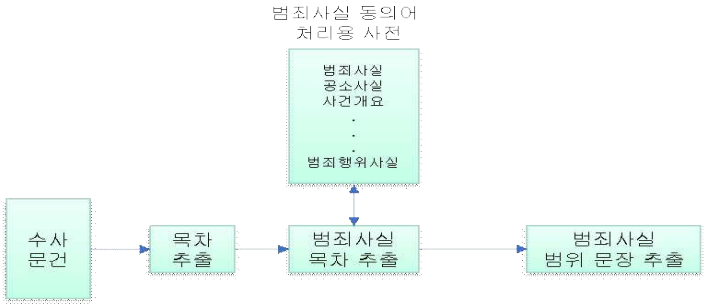
## 2-2. 범죄유형별 범죄의미 사전 선택기

범죄유형이 결정되면, 본문에서 추출한 여러단어들이 해당 범죄에서 어떤 의미인지를 해석하기 위해 해당 범죄의미 사전이 활용된다. 이를 위해 여러범죄 유형을 대응하기 위해 시스템이 확보중인 범죄의미사전 Pool에서 해당 범죄에 해당하는 범죄의미 사전 하나를 선택하는 과정이다.

## 2-3. 범죄사실 부분(범위) 추출기

수사문건은 내용에 따라 목차를 기준으로 별도 기록되어 있다. 또한 해당 범죄사실, 즉 사건의 개요를 나타내는 부분은 범죄사실, 공소사실, 사건개요 등등의 목차로 사람마다 조금씩 다르게 기록을 하고 있지만 몇가지 형식으로 거의 정해져 있다.

따라서 사람들이 주로 범죄사실을 기술하기 위해 주로 사용하는 목차명들을 사전 조사 및 정의하여 맵핑테이블로 구성한되 새롭게 입력된 문서에서 도출한 목차명을 동의어 처리하여 범죄사실 범위를 아래 그림2와 같이 추출한다.




(그림 2) 범죄사실 범위 추출 과정 설명도

## 2-4. 스토리라인 구성 요소 추출기

타임축 포함된 스토리라인에 구성할 요소는 사건 관련 주요 행위자, 주요사건시점(타임노드), 해당 시점에 벌어진 사건행위등은 표1와 같이 구성된다.

< 표 1 스토리라인 구성 요소 및 그 의미 설명 >

| 요소 | 설명 |
|---|---|
| time node | ● 본문에서 추출한 시간관련된 데이터<br>○ 예시) 2018.10.04. 6:50,<br>● 특정 시간이 표시된 뒤 바로 뒷문장에서 시간임을 암시하는 단어<br>○ 예시) 당일 저녁, 얼마뒤, 하루경과 뒤 등<br>● 위 타임노드를 식별 할 수 있는 주요 어휘들을 사전으로 구축하여 입력된 문서내 범죄사실에서 추출된 단어들을 비교해 가며 사전에 정의된 단어와 유사도가 높으면 이를 타임노드 정보로 추출한다. |

| | |
|---|---|
| person | ● 해당 문장의 내용의 사건 주요 행위자<br>● 피고인, 피해자, 목격자 등<br>○ 예시) 피고인1, 피고인2, 참조인 |
| activity | ● phrase 에서 뽑을 수 있는 사실관계를 담고 있는 핵심 구절 및 단어<br>○ 예시1) 피해자와 말다툼을 벌였으며,<br>○ 예시2) 경찰관 PC 방에 도착, 15 분 머무르고 A 와 피해자로부터 사건의 경위를 들었다.<br>● 마침표, 쉼표, 접속사 단위로 문장이 분리되며, 타임노드와 행위자를 제외한 문장을 대상으로 행위로 판단되는 부분만 범죄행위로 정보 추출한다.<br>● 해당하는 문장중 범죄행위라고 분류될만한 특징표현을 사전 학습하여 학습된 특징들과 가까운 값을 나타내면 이를 스토리라인의 범죄행위로 판단한다.<br>(이때 범죄행위를 분류할 수 있는 알고리즘은 다양하게 선택하여 구축할 수 있다) |

## 2-5. 스토리라인 시각화 표현

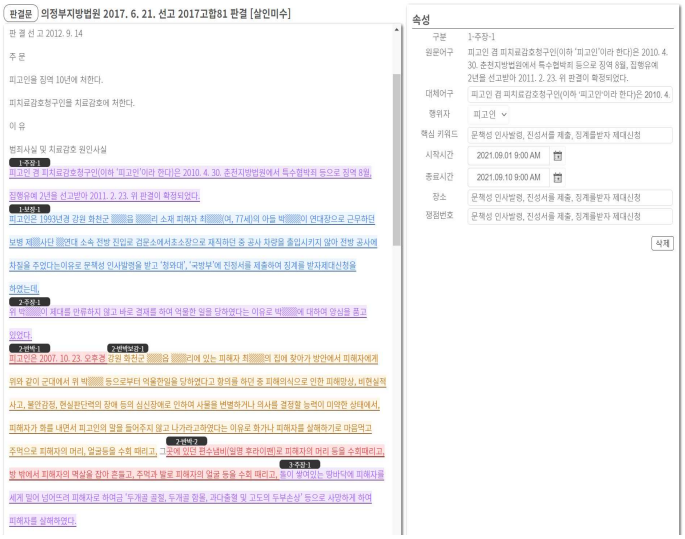
위 행위자 분석을 통해 발견된 행위자 수 만큼 타임시간축을 구성하고 이를 동시에 표현함으로써 관계된 사람들간 같은 시간대에 어떤 행위들이 있었는지를 입체적으로 표현할 수 있게 한다. 사건 행위자가 여러명인 경우 관계된 사람들 조합으로만을 볼 수 있도록 옵션으로 선택 표시하고 이들 대상으로만 볼 수 있게 하여 양자간 관계를 면힐히 볼 수 있도록 기능 지원한다. 타임라인에 표시된 특정 시간대 기록된 범죄행위 사실을 뒷 받침하는 사진자료, 사건기록, 감정기록, 객관적 감정자료 등을 첨부 할 수 있도록 기능 지원하여 향후 사건기록을 객관적으로 볼 수 있도록 지원한다.

## 3. 실험결과

수사문건중 범죄사실 부분을 인식하여 주요시각화 요소를 자동으로 추출하고, 이 요소들을 시각화된 형태로 제공하여 사건수사에 도움을 주는 실험결과를 제시한다 실험을 위해 수사문건은 법원 1심 판결문을 활용한다.

### 3-1. 스토리라인 구성요소 자동 추출 결과

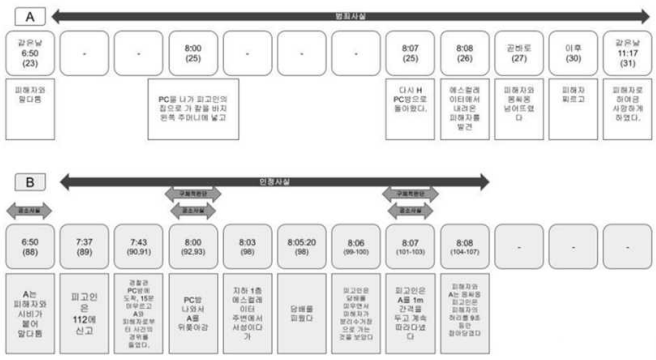
제안한 시스템이 판결문내 범죄사실 범위를 파악하고 이 범위에 포함된 각 문장들에 포함된 각 단어들을 대상으로 스토리라인 표현에 필요한 <표 1>에 정의된 요소들을 자동으로 추출하고 이를 수사관에게 제공하여 확인 뒤 오류 및 누락 사항에 대해 보완을 할 수 있도록 지원한 스토리라인 구성요소 자동 추출 결과 제시 화면이다.




(그림 3) 스토리라인 구성요소 자동 추출 결과 예

### 3-2. 스토리라인 시각화 표현 결과

그림 3에서 살펴보듯, 범죄사실 범위에 포함된 문장들 대상으로 주요행위자, 주요시각, 주요 범죄행위에 대해 행위자별로 타임축상에 주요행위를 함께 구성함으로써 벌어진 행위를 입체적으로 확인 할 수 있는 시각화 결과 이다.




(그림 4) 주요행위자별 주요 범죄사실 스토리라인 시각화 결과

## 4. 결론

본 논문에서는 현직수사관들이 접수된 범죄사건에 대해 올바른 수사방향을 잡기 위해 통산 초기에 하는 직접 수기로 타임라인을 작성하는 행위를 텍스트 기반 스토리라인 자동 시각화 시스템에 대해 살펴보았다. 본 시스템으로 인해 많은 사건을 살펴봐야 하는 수사관들에게 손쉽고 빠르게 범죄사실 스토리라인을 자동으로 볼 수 있게 하고, 주요행위자별, 관심

이슈등 다양한 옵션별 스토리라인 재표현등이 가능하여 다양한 각도로 수사방향을 예측할 수 있게 하여 과학수사에 도움을 주는 연구이다.

Acknowledgement

이 논문은 2021년도 정부(경찰청)의 재원으로 지원받아 수행된 연구결과임 [내역사업명: AI 기반 범죄수사 지원 / 연구개발과제번호: PR10-02-000-21]

## 참고문헌

[1] 임준태, “강력범죄에서의 범죄자 유형 분석기법(Criminal Profiling)의 도입에 관한 연구”, 한국공안행정학회한국공안행정학회보한국공안행정학회보 제17호, 2004, pp. 351-388

[2] 홍동숙, 김정준, 강홍구, 이기영, 서종수, “시공간 분석 기반 연쇄 범죄 거점 위치 예측 알고리즘”, 한국공간정보시스템학회 논문지 10(2), 2008.6, 63-

[3] 홍승표,임선영,고으뜸,홍성초,김정운,임중연, “판결문 분석을 통한 범죄 빅 데이터의 활용 가능성에 관한 소고”, 한국경찰학회보, 22권 6호, 2020, pp.209-230

[4] 조석영, “디지털 정보의 수사방법과 규제원칙”, 한국형사정책학회, 형사정책, 22권 1호, pp.75-98

[5] 안동욱, 임춘성, “소셜데이터 분석 및 인공지능 알고리즘 기반 범죄 수사 기법 연구”, 한국빅데이터학회지, 4권 2호, 2019, pp.23-34