ISSN 2383-630X(Print) / ISSN 2383-6296(Online)
Journal of KIISE, Vol. 50, No. 1, pp. 47-56, 2023. 1
https://doi.org/10.5626/JOK.2023.50.1.47

범죄 수사를 위한 상호작용이 가능한 다중 뷰 시각적 분석 시스템

(Interactive Visual Analytics System for Criminal Intelligence Analysts with Multiple Coordinated Views)

정 석 원 † (Seokweon Jung)
신 동 화 †† (Donghwa Shin)
복 진 욱 † (Jinwook Bok)
박 석 현 † (Seokhyeon Park)
전 현 † (Hyeon Jeon)
서 진 욱 ††† (Jinwook Seo)
이 인 수 †††† (Insoo Lee)
박 수 영 †††† (Sooyoung Park)

요 약 수사해야 하는 데이터는 날이 갈수록 많아지며 복잡해지고 있지만, 아직 수사 환경이나 방법은 그 변화를 따라가지 못하고 있다. 본 연구에서는 수사관들의 수사 환경과 수사를 위해 사용하는 소프트웨어에 대해 분석하고, 수사 과정에서의 센스메이킹 측면에 주목하여 기존의 센스메이킹을 위한 시각화 분석 기법을 수사에 적용함으로써 수사를 효율화할 수 있는 방법을 모색하였다. 분석 결과에 기반하여 과업과 디자인 요구사항을 도출하고, 이를 만족시키는 수사를 위한 다중 뷰 시각적 분석 시스템을 디자인하였다. 최종적으로, 제작한 프로토타입의 사례연구를 통하여 시각화 시스템의 활용 방법을 모색하였다.

키워드: 정보 시각화, 시각적 분석, 센스메이킹, 범죄 수사, 범죄 정보 분석

Abstract Data that criminal intelligence analysts have to analyze have become much larger and more complex in recent decades. However, the environment and methods of investigation have not yet kept up with those changes. In this study, we examined current investigation practices in Korean Government Agency. We focused on the sensemaking process of investigation and tried to adopt visual analytics approaches for sensemaking into the investigation. We derived tasks and design requirements and designed a multi-view visual analytics system that could satisfy them. We validated our design with a high-fidelity prototype through a case study to show realistic use cases.

Keywords: information visualization, visual analytics, sensemaking, crime investigation, criminal intelligence analysis

· This work was supported by Supreme Prosecutors' Office of the Republic of Korea grant funded by Ministry of Science and ICT (SPO2022A1202digitalB).
· 본 연구는 2022년도 정부(과학기술정보통신부)의 재원으로 대검찰청의 지원을 받아 수행된 연구결과임(SPO2022A1202digitalB).

† 비 회 원 : 서울대학교 컴퓨터공학부 학생
swjung@hcil.snu.ac.kr
bok@hcil.snu.ac.kr
shpark@hcil.snu.ac.kr
hj@hcil.snu.ac.kr
†† 비 회 원 : 서울대학교 지능형컴퓨팅사업단 박사후연구원
(Seoul Nat'l Univ.)
dhshin@hcil.snu.ac.kr
(Corresponding author임)
††† 종신회원 : 서울대학교 컴퓨터공학부 교수
jseo@snu.ac.kr
†††† 비 회 원 : 대검찰청 디지털수사과
insoo21@spo.go.kr
hilda01@spo.go.kr

논문접수 : 2022년 6월 2일
(Received 2 June 2022)
논문수정 : 2022년 8월 30일
(Revised 30 August 2022)
심사완료 : 2022년 11월 4일
(Accepted 4 November 2022)

정보과학회논문지 제50권 제1호(2023. 1)

1. 서 론

수사를 진행할 때는 사건과 관련된 다양한 데이터들을 탐색함으로써 증거를 발견해 내거나 혹은 가설을 검증하는 작업이 필요하다. 그러나 최근에는 인간의 다양한 기록들이 모두 디지털화되는 추세에 따라 탐색해야 하는 데이터의 양이 과거에 비해 급증하고 있다. 그러나 수사기관의 인력과 수사 기간은 한정되어 있어서 짧은 기간 동안 모든 데이터를 확인하는 것은 현실적인 접근 방식으로 보기 힘들다. 데이터의 급증과 더불어, 수사 대상들의 은닉행위 또한 수사를 어렵게 하는 걸림돌이다. 수사 대상들은 다양한 방식으로 본인들의 행위를 은폐하려 시도하기 때문에 데이터에는 누락, 변조 혹은 중복이 존재할 수 있다. 이런 문제들로 인해 수사기관은 데이터를 효과적으로 탐색하고 분석을 하는 데 어려움을 겪고 있다.

시각적 분석(visual analytics)은 복잡한 문제의 해결을 도우려는 방법의 하나로 다양한 분야에 활용되고 있다[1,2]. 상호 연결된 다중 뷰(multiple coordinated views)로 구성된 시각적 분석 시스템을 활용하여 사용자는 쉽게 데이터를 분석하고 이해할 수 있다[3,4]. 최근에는 수사 분야에서도 시각적 분석을 통해 수사를 효율화하는 연구가 활발히 진행 중이며, 실제 수사 기관들과의 협력을 통해 시각적 분석 시스템을 설계한 연구들도 존재한다[5-10]. 수사 중 혐의에 대한 가설을 수립하고 검증하는 과정은 일종의 센스메이킹(sensemaking)과정이라 할 수 있는데, 시각적 분석은 이 과정에서의 인지 부하를 경감하여 분석자에게 필요한 분석 시간과 노력을 줄일 수 있다[11,12].

본 연구에서는 시각적 분석 방법론을 활용하여 수사, 그 중에서도 다수의 인물들 사이의 연관 관계를 분석할 필요가 있는 사건의 수사에 있어서 발생하는 어려움들을 해결하려 시도하였다. 해당 분야의 전문가들의 수사 방법을 조사 및 분석하여 수사의 과업(task)을 정리하고 과업들의 수행을 효과적으로 지원하기 위한 시스템의 디자인 요구사항(design requirements)을 수립하였다. 최종적으로 이러한 요구사항에 부합하는 효과적인 프로토타입을 설계하였으며, 이를 활용한 분석 시나리오를 제시하였다.

본 논문의 구성은 다음과 같다. 1장의 서론에 이어 2장에서는 센스메이킹, 그리고 범죄 정보 분석을 위한 시각적 분석과 관련된 관련 연구를 다룬다. 3장에서는 규모가 크고 복잡한 사건에서 시각적 분석 방법론이 활용되는 방법 및 사건에서 다루는 데이터의 특성에 관해 설명하며, 효과적인 분석 도구를 설계하기 위한 과업 및 디자인 요구사항을 정리한다. 4장에서는 시스템 디자인과 각 시각화에 대한 설명, 그리고 5장에서는 도구에 대한 평가로서 시스템의 수사에서의 활용 시나리오를 제시한다. 마지막으로 6장에서는 결론과 향후 연구 방향을 기술한다.

2. 관련 연구

2.1 센스메이킹을 위한 시각적 분석

센스메이킹은 많은 의사 결정에 포함되는 필수적인 과정으로, 시각적 분석을 통하여 이를 보조하기 위한 연구들이 많이 진행되었다. Shrinivasan은 센스메이킹에서 시각화의 활용 가능성에 관해 탐구하고, 효율적인 시각화 활용을 위하여 데이터 뷰, 지식 뷰, 네비게이션 뷰로 구성되는 시각화가 필요함을 보였다[13]. Stasko는 대량의 문서들에 내재된 데이터들을 분석하기 위하여 상호 연결된 다중 뷰들로 구성된 시각적 분석 시스템인 직쏘(Jigsaw)를 제시하였다. 이를 통해 문서의 객체들을 시각화하고, 객체들 사이의 연결 관계를 탐색할 수 있도록 하였다[3]. Kang은 직쏘를 여러 분야의 전문가들에게 배포하여, 해당 분야에서 실제로 시각적 분석 시스템을 통한 데이터의 탐색이 효율적으로 센스메이킹을 보조할 수 있음을 보였다[14]. 본 연구는 위의 연구들을 통해 센스메이킹 과정에서의 효과가 입증된 상호 연결된 다중 뷰의 개념을 범죄 정보 분석 분야에 효과적으로 접목하기 위한 시도이다.

2.2 수사를 위한 시각적 분석

범죄 정보 분석(criminal intelligence analysis, CIA)은 범죄에 대해 수집한 정보와 정보를 이용하여 수사에 활용할 여러 접근 방법들을 제시하는 방법론이다. 범죄 정보 분석을 위한 시각적 분석의 필요성은 다양한 논문들에서 제기되어왔다[14,15]. 유럽에서는 여러 대학과 수사기관이 연합하여 수사를 위한 시각적 분석 도구를 연구하는 프로젝트를 진행하였다[5-8]. 그중 수사관들의 추론 과정을 이해하기 위한 시도로서 시각화 시스템을 통해 분석 기원(analytic provenance)을 추적하고자 시도한 연구가 있다[8]. 분석 기원이란 수사관들이 결론을 도출하는 과정에서 어떤 데이터를, 어떤 분석 과정을 통해, 그리고 어떤 추론 과정으로 수행하였는지를 뜻하는 개념이다[16]. 이를 위해서 연구진은 범죄 정보 분석을 위한 시각화를 제공한 뒤, 분석가들이 시각화들을 활용하여 분석을 진행한 과정을 기록하고 분석하였다. 이러한 분석 기원은 범죄 정보 수사를 이해하는 데에 큰 의미가 있지만, 어떻게 데이터를 탐색하는 것이 효과적인지에 대해서는 더 연구가 필요하다. 본 연구는 시각적 분석을 통해 범죄 정보 분석에서 어떻게 데이터를 탐색하는 것이 효과적인지에 더욱 초점을 맞추었다.

Visilant는 경찰과의 협력 연구 끝에 설계된 협력 수사를 위한 시각적 분석 시스템이다[9]. 이는 데이터에 대한 오버뷰(overview), 객체 간의 관계, 수사의 진행도,

그리고 검색 결과에 대한 자세한 내용을 보여주는 등의 기능을 지원한다. 이렇게 수사 과정에 대한 전반적인 지원이 가능하지만, Visilant는 매우 다양한 형태로 존재하는 범죄 정보 분석 데이터의 속성들을 그 특징에 맞는 다양한 관점으로 탐색할 수 있는 시각화를 지원하는 데에 취약하다. 본 연구에서는 데이터의 다양한 속성들 각각을 효과적으로 시각화할 수 있는 다중 뷰를 설계하여 수사관이 분석에 활용할 수 있도록 하였다.

3. 대상 분야 현황 분석

3.1 시각적 분석 방법론을 활용한 수사

경제범죄 등과 같이 규모가 크고 복잡한 사건일수록 자료가 방대하기에 통합분석 및 시각화 등 기능을 지원하는 소프트웨어의 사용이 필수불가결하다[10]. 수사관들은 정보를 기반으로 사실관계를 파악하고 추가 여죄 및 공범관계를 인지하기 위해 계좌 거래내역, 통화내역 등 대량의 텍스트 데이터를 직접 검토해야만 한다. 이 과정에서 데이터를 사람이 직접 분석하여 의미 있는 정보를 인지하는 것은 지나치게 노동 집약적이고 많은 시간을 요구한다. 이러한 어려움들을 극복하기 위해 시각화 소프트웨어들이 활용되고 있다.

IBM i2 Analysts's Notebook(i2)은 2015년부터 경찰청에서 지능형 수사자료 분석을 위해 도입한 소프트웨어이다[10]. i2 복잡한 사건에 존재하는 다양한 데이터를 조합하여 사용자에게 시각화를 제공한다. 사용자는 시각화를 활용하여 데이터에 대한 깊이 있는 분석 및 데이터들 사이에 존재하는 연관성을 파악하는 것이 가능하다[17]. 이는 수사에서 발생하는 인지능력 부하를 경감시켜 수사관이 수사를 효율적으로 수행할 수 있도록 돕는다. 그러나 i2는 5만 건이 넘는 대용량의 데이터는 일반적인 방법으로는 한 번에 시각화 할 수 없는 문제가 존재하는데, 이는 대용량의 데이터를 다뤄야 하는 경제범죄 사건에서 큰 걸림돌이 된다. 또한, i2에서는 상용 소프트웨어로 다양한 상황에서의 활용을 목표로 하기 때문에, 데이터에 최적화된 시각화를 제공해주지 못한다.

본 연구에서는 기존의 시각화 소프트웨어에 존재하는 대용량 데이터 처리의 한계를 극복하고 수사의 효율성을 증진시키기 위한 새로운 데이터 전처리 방법을 모색하였다. 또한, 데이터에 최적화된 다양한 시각화를 제공, 사용자의 유기적인 데이터 탐색에 대한 요구를 만족시키기 위하여 상호 연계된 다중 시각화 뷰로 구성된 시각적 분석 시스템을 설계하였다.

3.2 수사 데이터의 수준 및 전처리

본 연구에서 다루고자 하는 규모가 크고 복잡한 사건의 데이터는 일반적으로 계좌 거래내역과 통화 내역으로 구성된다. 본 연구에서는 데이터가 가지는 특성에 따라 원시 데이터를 전처리하여 3가지의 수준: '엔트리(entry)', '노드(node)', '관계(relation)'로 나누었다(그림 1)

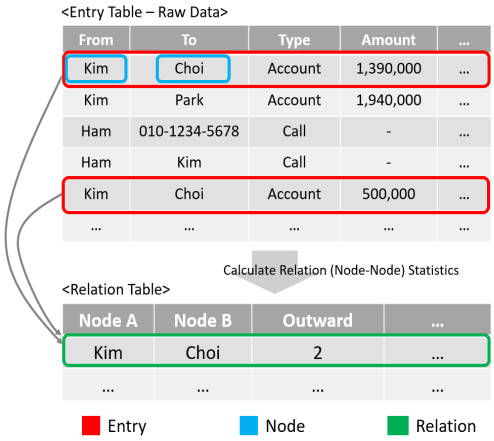

그림 1 디지털 증거 데이터의 수준: 엔트리, 노드, 관계
Fig. 1 Major levels of digital evidence data: entries, nodes, and relationships

- 엔트리: 원시 데이터에서 하나의 행(row)은 하나의 기록을 담고 있다. 본 연구에서 이를 엔트리라고 칭한다. 엔트리는 항상 두 대상 간의 상호작용으로 구성된다. 예를 들면, 통화기록 문서에 존재하는 하나의 행, 즉 엔트리는 어떤 인물이 다른 인물에게 발신한 통화에 대한 정보이다.
- 노드: 엔트리 상에서 대상으로 취급되는 것들을 본 연구에서는 노드라고 칭한다. 노드가 될 수 있는 것들의 예로는 인물, 계좌번호, 또는 전화번호가 있다.
- 관계: 여러 엔트리를 살펴보면 특정 노드 간에 특별히 높은 통화 횟수나, 거래 건수 등이 발견된다. 이렇게 노드와 노드 간의 상호작용들에 대한 통계량을 본 연구에서는 관계라고 칭한다.

3.3 과업과 디자인 요구사항

앞서 조사한 시각적 분석 방법론을 활용한 수사 방식과 환경을 바탕으로데이터의 탐색 및 분석을 수행하는 단계에서 시각적 분석 시스템을 통해 수행되어야 하는 주요 과업을 아래와 같이 정리하였다.

- T1: 특정 행위의 규명을 위해 해당 행위의 존재를 뒷받침할 증거를 탐색
- T2: 특정 행위에 대한 공범을 알아내기 위해 피의자 주변의 인물 중 의심스러운 인물을 탐색

그리고 위의 두 과업을 사용자가 효과적으로 수행할 수 있게 하기 위한 시각적 분석 시스템의 디자인 요구사항을 아래와 같이 정의하였다.

- R1: 수사 대상을 지정할 수 있는 수단 제공
- R2: 다양한 종류의 데이터 패턴을 나타낼 수 있는 다중 뷰 인터페이스 제공
  - R2-1: 데이터의 수준별(엔트리, 노드, 관계) 패턴 시각화
  - R2-2: 데이터의 공간적 패턴 시각화
  - R2-3: 데이터의 시간적 패턴 시각화
- R3: 데이터의 각 패턴을 조건으로 설정할 수 있는 필터링을 제공
- R4: 수사 대상과 관련된 다른 의심스러운 수사 대상을 나타낼 수 있는 시각화

4. 인터페이스 디자인

위에서 정의한 과업과 디자인 요구사항에 기반하여 수사 업무를 효율적으로 수행할 수 있도록 돕는 시각적 분석 시스템을 디자인하고, 프로토타입을 개발하였다.

4.1 시스템 개괄

시각적 분석 방법론을 활용한 수사 방법 및 환경에 대한 조사에 따르면, 수사관들은 여러 데이터를 넘나들며 인과 관계를 탐색하는 수사 방법을 활용한다. 따라서 여러 개의 다중 뷰로 구성된 하나의 시각적 분석 시스템을 디자인하고, 각 뷰들이 사용자와의 상호작용에 따라 서로 유기적으로 연결되어 있도록 하였다(R2). 또한, 각 뷰 들에서 나타내는 데이터의 패턴을 조건으로 하는 필터링을 지원하여 사용자가 탐색해야 하는 데이터의 범위를 효과적으로 줄일 수 있도록 하였다(R3). 과업 T2의 효과적인 수행을 위해 인터페이스 상에서 사용자에게 주요 수사 대상으로 지정된 대상들과 밀접한 관련이 있을 것으로 의심되는 수사 대상을 다중 뷰들 내에서 다양한 방식의 시각화로 제공하였다(R4).

결과적으로, 본 논문에서 제안하는 시각적 분석 시스템은 총 5개의 서로 다른 시각화 뷰: 데이터 테이블 뷰(그림 2-A), 대상 노드 뷰(그림 2-B), 위치 분석 뷰(그

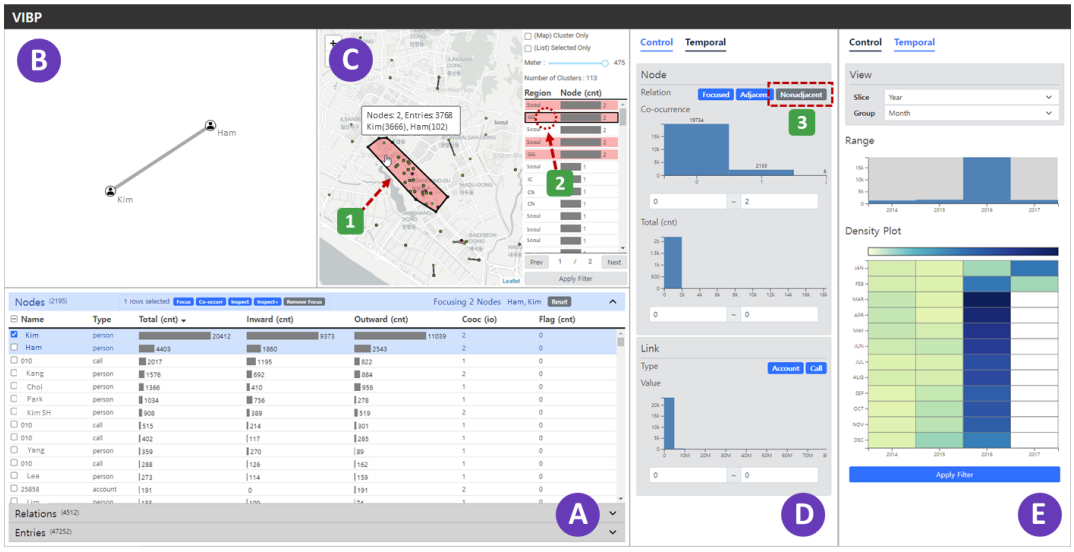

그림 2 시각적 분석 시스템 개괄. (A) 3가지 데이터 수준에 따라 전처리된 데이터를 아코디언 형식의 테이블 시각화로 제공한 데이터 테이블 뷰. (B) 선택한 대상 노드들을 노드-링크 다이어그램 시각화로 제공하는 대상 노드 뷰. (C) 선택한 대상 노드들이 포함하는 엔트리들의 발생 위치를 지도 위에 시각화하고, DBSCAN 알고리즘에 따른 군집분석 결과를 보여주는 위치 분석 뷰. (D) 선택한 대상 노드들과 데이터 테이블 뷰에서 탐색 중인 다른 노드들의 관계성 정보를 제공하는 관계 분석 뷰. (E) 데이터 테이블 뷰에서 탐색 중인 노드들의 시간에 따른 엔트리 발생 정보를 제공하는 시간 분석 뷰

Fig. 2 Interface of the system. (A) Data table view presenting data in terms of the three major data structures through an accordion-style table visualization. (B) Focus node view showing user-designated focus nodes with a node-link diagram. (C) Spatial analysis view showing geographic location of entries on the map and also DBSCAN clustering results of entries. (D) Relationship analysis view providing various information about the relationship between focus nodes and other nodes. (E) Temporal analysis view showing temporal information of entries at various time levels

림 2-C), 관계 분석 뷰(그림 2-D), 시간 분석 뷰(그림 2-E)로 구성되었다.

4.2 데이터 테이블 뷰

데이터 테이블 뷰(그림 2-A)는 데이터를 3가지 수준(엔트리, 노드, 관계)에 따라 탐색할 수 있도록(R2-1) 각각의 수준을 하나의 테이블 형태로 나타낸다. 엔트리, 노드, 관계에 해당하는 각각의 증거 데이터들은 모두 다차원 정보를 포함하고 있다(표 1). 그 때문에 한정된 공간상에 다차원 정보들을 나타낼 수 있으면서, 사용자가 익숙하게 받아들일 수 있는 테이블 형태의 시각화를 활용하였다. 또한, 수준이 3가지이므로 테이블도 3가지가 존재하며, 이는 아코디언(accordion) 기반의 탭 디자인을 활용하여 표현하였다. 사용자는 원하는 데이터 수준을 나타내는 탭을 클릭하는 방식으로 실시간으로 3가지 테이블을 자유롭게 전환하며 활용할 수 있다.

표 1 데이터 테이블 뷰의 각 탭에서 제공하는 항목
Table 1 Data attributes displayed in data table view

Attribute	Description	Tab
Name	Name of the node	All
Inspect Node	Base node in the relation and entry tables	Relation
Inspect+ Node	Opponent node in the entry table	Entry
Type	Type of the node	Node
Total (cnt)	Total number of entries containing the node or the relationship	Node, Relation
Inward (cnt)	Total count of the node as 'to' node	Node, Relation
Outward (cnt)	Total count of the node as 'from' node	Node, Relation
Cooc (io)	Number of relations with focus nodes	Node, Relation
Value	Representative value of the entry	Entry
Direction	Direction of the entry	Entry
Time	Occurrence time of the entry	Entry
Location	Occurrence location of the entry	Entry

데이터 테이블 뷰는 크게 3가지의 주요 상호작용을 제공함으로써 사용자들이 다양한 여러 테이블과 다른 뷰들을 효과적으로 활용할 수 있도록 한다(그림 3). 첫째로, (1) '대상 노드 설정'은 대상 노드를 지정하기 위해서 노드 테이블에서는 사용자가 미리 관심을 두고 있던 대상 노드가 있는 행을 클릭하고, 'Focus' 버튼을 통해 지정한다(그림 5-1, R1). 이렇게 대상 노드로 설정된

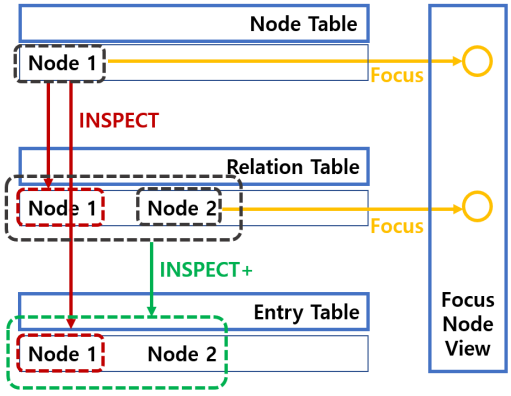

그림 3 데이터 테이블 뷰에서 지원하는 주요 상호작용 기법들: 인스펙트(Inspect), 인스펙트+(Inspect+), 대상 노드 설정(Focus)

Fig. 3 Major interaction techniques supported by data table view: Inspect, Inspect+ and Focus

노드는 후술할 대상 노드 뷰에 시각적으로 표현된다. 둘째로, (2) '인스펙트'는 특정한 노드가 포함된 관계 혹은 엔트리를 알아보기 위해 노드 테이블에서 특정한 노드 행을 선택 후 'Inspect' 버튼을 통해 지정한다. 셋째로, (3) '인스펙트+'는 인스펙트를 통해 관계 테이블을 탐색하다가 특정한 관계에 대한 엔트리들을 모두 찾아보고 싶을 때 해당하는 행을 선택 후 'Inspect+' 버튼을 통해 지정한다.

이외에도, 각 테이블 상의 다차원 정보 중 특정한 정보를 기준으로 데이터를 정렬할 수 있게 하여, 이를 통해 선택한 수사 대상과의 관계가 특이하게 밀접한, 어떠한 관계가 있을 것으로 추정되는 의심스러운 수사 대상들을 새로이 발견할 수 있다(R4).

4.3 대상 노드 뷰

대상 노드란 노드 테이블 상에서 'Focus' 버튼을 통해 수사 대상으로 지정된 노드를 일컫는다. 대상 노드 뷰(그림 2-B)는 사용자로부터 지정된 대상 노드들을 노드-링크(node-link) 다이어그램으로 시각화하여, 사용자가 각 대상 노드들 사이의 관계를 직관적으로 관찰할 수 있도록 돕는다. 노드-링크 다이어그램을 활용한 그래프 형식의 시각화는 이해가 쉽고 직관적이기 때문에 관련 연구들[3,13]에서도 많이 활용되었다. 대상 노드는 종류(인물/계좌번호/전화번호)에 따라서 알맞은 원형의 아이콘으로 표시되고(그림 4), 두 노드 사이의 관계는 횟수에 비례하는 두께의 선으로 나타내었다. 사용자는 이 시각화를 통해 선택 노드들 사이의 관계에 대해서 파악할 수 있기에, 대상 노드 주변의 의심스러운 수사 대상을 확정하는 데에 활용할 수 있다(R4).

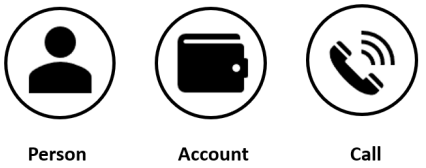

그림 4 대상 노드 뷰 상의 노드 종류별 아이콘

Fig. 4 Icons for each node type on the focus node view

4.4 위치 분석 뷰

위치 분석 뷰(그림 2-C)는 다양한 타입의 엔트리가 갖는 공통 속성 중 하나인 위치 정보를 시각화한다(R2-2). 위치 정보를 다루는 다른 시각적 분석 시스템[3,4]과 유사하게, 대상 노드들이 갖는 모든 데이터 엔트리를 지도상에 하나의 점으로 표현하여 공간상에서 엔트리의 분포를 한눈에 확인할 수 있게 하였다. 또한, DBSCAN 알고리즘[18]을 이용한 엔트리들의 군집분석 결과를 함께 보여줌으로써 수사 대상들의 활동에 대한 공간적 패턴을 좀 더 효과적으로 파악할 수 있도록 하였다. 지도상에서 군집들은 그것들이 포함하고 있는 엔트리 모두를 포함하는 볼록 껍질(convex hull)로 시각화된다(그림 2-1). 사용자는 이 다각형들의 분포를 확인함으로써 대상 노드들의 주 거주지를 파악할 수 있다. 만약 군집 내에 다수의 대상 노드들이 포함되어 있다면, 이는 해당 대상 노드들 사이에 공간적인 접점이 존재한다고 판단할 수 있다(R4).

지도 우측의 탭에서는 DBSCAN 알고리즘의 주요 매개변수 중 하나인 기준거리(epsilon) 값을 조절할 수 있다. 그리고 군집들을 리스트의 형태로 표현하였는데, 이는 화면 공간의 한계로 인해 현재 지도상에 나타나고 있지 않은 군집들의 정보도 확인할 수 있도록 하기 위함이다. 리스트에서는 군집들을 각 군집에 포함된 대상 노드의 수를 기준으로 정렬하여 보여준다. 따라서 지도상에서 직접 군집을 탐색하지 않더라도, 사용자는 대상 노드를 가장 많이 포함하고 있는 군집을 곧바로 알 수 있다. 또한, 돋보기 버튼(그림 2-2)을 클릭하면, 해당 군집의 위치로 지도의 시점이 자동으로 이동하기 때문에, 군집의 실제 공간상의 위치를 효율적으로 파악할 수 있다.

4.5 관계 분석 뷰

관계 분석 뷰(그림 2-D)는 노드들 사이의 관계를 시각화하고, 이를 기반으로 필터링 기능을 제공하는 뷰이다.

관계 분석 뷰는 전체 노드 사이의 관계 정보에 기반을 둔 4가지 서로 다른 필터로 구성되어 있다. 사용자는 필터와의 상호작용을 통해서 데이터 테이블 뷰에 표시되는 데이터의 범위를 좁힐 수 있다(R3).

- 관계 필터: 노드들을 대상 노드, 대상 노드로부터 거리가 1인 인접 노드, 그리고 그 외 비인접 노드의 세 그룹으로 구분하여, 각 그룹을 데이터 탐색 범위에 포함할지 선택할 수 있다.
- 동시출현 필터: 동시출현도(표 1)는 인접 노드들이 몇 개의 대상 노드들과 관계를 갖는지 알려주는 지표이다. 동시출현 필터는 데이터에 존재하는 노드들의 동시출현도를 히스토그램을 통하여 사용자에게 제공한다. 사용자는 탐색을 원하는 동시 출현도 그룹을 선택하여 해당 그룹에 속하는 노드들을 필터링할 수 있다.
- 연결도 필터: 각 노드는 다른 노드들과 연결 관계를 바탕으로 계산한 Total(종합 연결도), Inward(내심 연결도), Outward(외심 연결도)를 가진다(표 1). 이를 히스토그램을 통해서 사용자에게 제공하고, 직접 구간을 선택하여 필터링할 수 있도록 하였다.
- 값 필터: 노드들 사이 관계에 속한 엔트리들의 값이 지나치게 작은 경우는 유효하지 않은 데이터일 가능성이 있다. 따라서 연결 관계가 갖는 값을 바탕으로 필터링을 진행하여 데이터의 탐색 범위를 좁힐 수 있도록 하였다.

4.6 시간 분석 뷰

시간 분석 뷰(그림 2-E)는 데이터의 엔트리에 존재하는 발생 시간에 대한 정보를 시각화하고, 이를 기준으로 하는 필터링을 지원하는 뷰이다(R2-3). 시간은 순차적인 속성과 주기적인 속성에 따른 패턴을 모두 가지고 있고, 데이터에서는 이 두 가지 속성을 모두 고려하여 어떤 관계가 수상한 관계인지 구분해야 한다[4]. 예를 들어, 매달 비슷한 시기 반복적으로 발생하는 주기적인 엔트리를 가지는 관계보다는 어느 순간을 기준으로 급격하게 빈도가 늘어난 관계가 수상하다 할 수 있을 것이다.

따라서 순차적 시간 단위(년, 월, 일)로 데이터를 나누어 스냅샷(snapshot)을 생성하고, 이를 다시 주기적 단위(1년 중 12월, 1개월 중 28~31일, 요일, 1일 중 24시간)로 나누어 데이터의 순차적 패턴과 주기적 패턴을 모두 탐색할 수 있도록 하였다. 사용자는 상단의 히스토그램을 통해서 순차적 단위에 따라 생성된 스냅샷에 포함된 데이터 엔트리의 분포를 확인할 수 있다. 또한, 사용자는 브러싱(brushing) 상호작용을 통하여 주기적 패턴 탐색을 진행하려는 영역을 선택할 수 있다.

하단의 밀도 그림(density plot)에서는 위의 순차적 스냅샷 히스토그램에서 선택된 시간 영역의 데이터를 다시 주기적 단위로 나눈 분포를 시각화한다. 이를 통하여 사용자는 시간의 흐름에 따라 각 주기적 단위 시간의 데이터 분포의 변화를 관찰하여 패턴을 찾아낼 수

있다. 또한, 위와 마찬가지로 브러싱을 통하여 더 자세히 관찰하려는 스냅샷의 범위를 선택하여 해당 시간 구간의 데이터만 필터링할 수 있다(R4).

5. 평가 - 대포통장 수사에서의 활용 사례

본 연구에서 개발한 시각적 분석 시스템의 효용을 증명하기 위해 빈번하게 발생하고 있는 대포통장을 이용한 자금세탁 사례[19,20]를 가정하여 가상의 데이터를 생성하고, 이를 바탕으로 한 시나리오를 제시한다. 시나리오에 대한 사전 가정은 다음과 같다.

- H1: “Kim”과 “Ham”은 공범 관계에 있으며, 자금세탁을 위해 대포통장을 활용하고 있다.
- H2: 대포통장을 거치는 자금의 흐름은 일반적이지 않은 패턴을 띄고 있다.
- H3: “Kim”과 “Ham”은 추적을 회피하기 위해 주기적으로 대포통장의 명의 등을 교체한다.

본 시나리오에서는 “Kim”과 “Ham”을 중심으로 시각적 분석 시스템을 활용한 데이터 탐색을 통해서 아래의 두 가지 목표를 달성하려 한다. 각 목표는 3.3장에서 상술한 전문가들의 과업(T1, T2)과 각각 대응한다.

- O1: 두 인물 “Kim”과 “Ham” 사이의 공모 관계를 증명할 추가 증거 탐색(T1)
- O2: 두 인물 “Kim”과 “Ham” 주변의 숨겨진 대포통장탐색(T2)

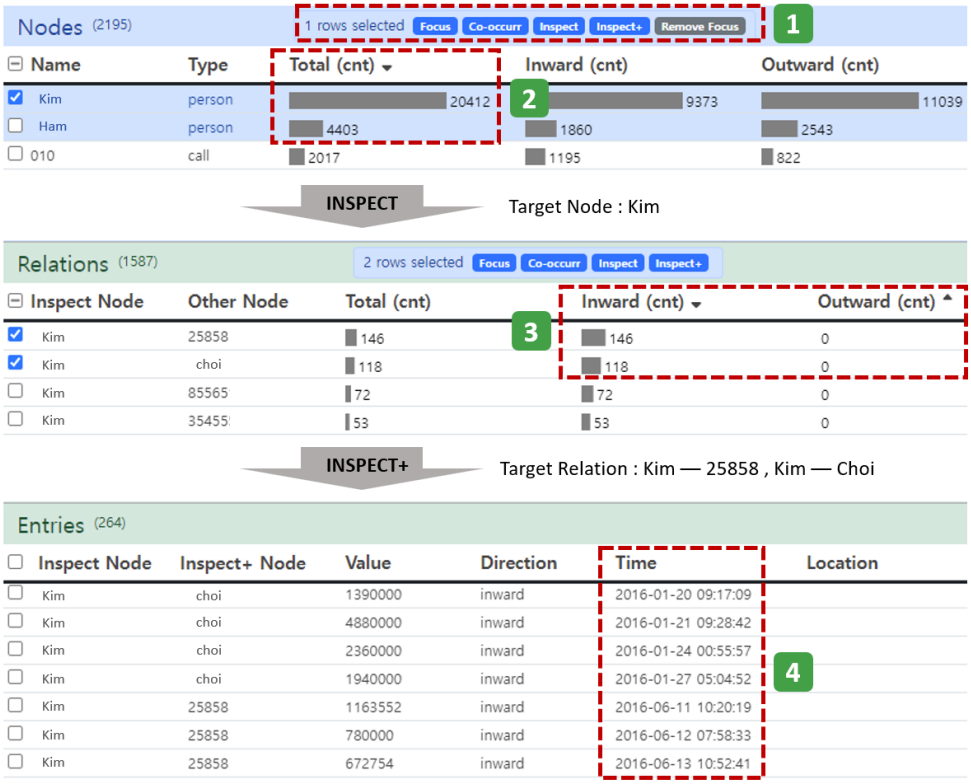

그림 5 데이터 테이블 뷰에서 Inspect와 Inspect+ 상호작용을 통해 데이터를 분석하는 과정. 상단의 노드 테이블에서 “Kim”을 인스펙트하여 중단의 관계 테이블을 형성한다. 그렇게 형성된 관계 테이블에서 “25858”, “Choi”와 “Kim” 사이에 불균형한 관계가 있음을 확인하였다. 마지막으로, “Choi”, “25858”을 인스펙트+ 하여 하단의 엔트리 테이블을 형성한다. 테이블의 각 항목에 대해서는 (표 1)에서 확인이 가능하다.

Fig. 5 Process of analyzing data through Inspect and Inspect+ interactions in the data table view. Users generate items of the relationship table by applying Inspect to ″Kim″ in the node table. In the relationship table, users confirm that there are imbalanced relationships between “25858”, “Choi” and “Kim.” Finally, users generate items of the entry table by applying Inspect+ to “Choi,” and “25858” in the relationship table. Explanations about attributes of tables are shown in (Table 1).

“Kim”과 “Ham”이 공범 관계가 있으며, 자금 세탁을 위해 대포통장을 활용한다는 시나리오의 가정(H1)에 따라 테이블에서 “Kim”, “Ham” 두 노드를 선택한 뒤 ‘Focus’ 버튼을 통해서 대상 노드로 추가하였다(그림 5-1). 대상 노드 뷰에 두 노드가 추가되고, 데이터가 이 두 노드를 기준으로 다시 계산된다. 관계 필터에서 ‘Nonadjacent’ 버튼을 눌러서 비인접 노드의 선택을 해제함으로써 탐색의 범위를 대상 노드와 그 인접 노드까지로 한정하도록 한다(그림 2-3).

우선 현재 두 노드만이 대상 노드로 지정되어있는 상태에서 두 노드 사이의 관계에 대해서 살펴보도록 한다(O1). 대상 노드 뷰에서는 두 노드 사이에 링크가 존재하여 직접적인 연결 관계를 확인할 수 있다(그림 2-B). 그러나 이러한 직접적인 관계가 존재하는 것만으로 둘 사이가 밀접하다는 것을 증명할 수는 없다. 다양한 관점에서 분석하기 위하여 위치 분석 뷰를 활용하였다(그림 2-C). 지도에서 군집분석 결과를 확인해본 결과, “Kim”과 “Ham”을 동시에 포함하고 있는 군집을 총 5개 발견하였다. 그중 한 군집(그림 2-1)의 경우, “Kim”의 출현 횟수가 3000건 이상이고, “Ham”의 출현 횟수도 102건을 기록하고 있었다. 이를 통해 해당 군집이 존재하는 지역이 “Kim”의 주 거주 지역이면서, “Ham”이 해당 지역을 방문하여 “Kim”와 접촉을 했을 것으로 추측할 수 있다. 물리적인 접촉이 존재했다는 것은 단순히 전화 연결이나 계좌 거래가 있는 것 보다 둘 사이에 더 밀접한 관계가 존재하고, 이는 공모 관계가 있음을 뒷받침하는 증거로 활용할 수 있다.

다음으로, “Kim”과 “Ham” 주변의 노드 중 대포통장 역할을 하는 노드를 찾아보려 한다(O2). 데이터 테이블 뷰에서 살펴본 결과 “Kim”가 “Ham”보다 종합연결도가 높고(그림 5-2), 이는 “Kim”와 관계된 노드가 더욱 많을 것이라는 의미로 파악할 수 있다. 따라서 “Kim”의 주변을 탐색하는 것이 주변의 대포통장 탐색에 더 유리할 것이라는 판단을 내렸다. 따라서 데이터 테이블 뷰에서 “Kim”를 선택 후, ‘인스펙트(Inspect)’ 버튼을 클릭하여 데이터의 탐색을 진행하였다.

“Kim”를 인스펙트 대상으로 추가한 뒤, 데이터 테이블 뷰에서 관계 테이블을 보면, “Kim”를 기준으로 다른 노드들과의 관계를 관찰할 수 있다. 내심연결도와 외심연결도의 불균형함은 일반적인 관계에서는 나오기 힘든 일방적인 관계를 나타낸다고 볼 수 있고, 이는 해당 노드들이 대포통장임을 추측할 수 있다(H2), 이와 비슷한 경우를 탐색하기 위해 테이블의 정렬 기능을 통해서 외심연결도는 오름차순으로, 내심연결도는 내림차순으로 정렬을 하였다. 그 결과, 일부 노드들은 “Kim”와의 내심연결도는 100 이상이지만 외심연결도는 0, 즉 100건이 넘는 거래를 했지만 “Kim”에게 주기만 했을 뿐 받은 적은 한 번도 없다는 사실을 발견할 수 있었다(그림 5-3). 이러한 일반적이지 않은 패턴을 보이는 두 노드가 “Choi”(인물)와 “25858”(계좌)임을 알아냈고, 두 노드를 대상 노드로 추가하여 해당 노드들에 대한 추가적인 탐색을 진행하였다.

대상 노드 뷰에서 새로 추가된 대상 노드들의 연결 관계를 관찰한 결과, “Choi”와 “Kim”, “25858”과 “Kim”, “25858”과 “Ham” 사이의 간선이 새로 추가된 것을 발견하였다. 이를 통해 “Choi”는 “Kim” 외에 다른 어떤 노드와도 거래한 적이 없고, “25858”는 “Kim”, “Ham” 모두와 거래 관계가 있다는 것을 알 수 있다. “Kim”과 새로 추가한 두 노드 “Choi”, “25858” 사이의 관계를 더 자세히 살펴보기 위해서, 데이터 테이블 뷰에서 “Choi”와 “25858” 두 노드를 선택 후 ‘인스펙트+(inspect+)’를 클릭하여 엔트리의 범위를 “Kim”와 두 노드 사이의 엔트리로 한정시켰다. 그 후 데이터 테이블 뷰의 엔트리 탭으로 이동하여 엔트리에 대한 탐색을 진행하였다(그림 5). “Kim”와 두 노드 사이에는 총 264건의 거래가 존재하고, 특히 Choi의 경우 2015년 12월 14일부터 2016년 1월 27일까지, “25858”는 2016년 6월 11일부터 2017년 1월 23일까지 거래를 했음을 확인할 수 있다(그림 5-4). 각 노드의 거래 패턴을 관찰한 결과, 해당 노드들은 현금으로 돈을 수집하여 “Kim” 또는 “Ham”에게 송금하는 대포통장의 역할을 한 것으로 보였다.

“Choi”와 “25858” 노드의 활동 사이에는 2016년 3월부터 5월까지 약 2달의 공백이 존재한다. 수상한 관계가 휴식기 없이 지속되었다는 가정 하에 해당 시간대에도 “25858”이나 “Choi”와 같이 비슷한 역할을 하는 다른 노드가 존재한다고 추측할 수 있다(H3). 또한, 시간 분석 뷰에서 확인해본 결과 해당 기간은 전체 데이터에서 가장 엔트리의 개수가 많은 시간대이다. 따라서 비슷한 역할을 하는 노드가 해당 시간대에 여럿 존재할 수 있다고 추측할 수 있다. 이를 직접 확인하기 위해 시간 필터를 활용하여 데이터의 탐색 범위를 2016년 3월부터 5월까지로 한정해보았다. 그 후 지금까지 수행한 방식과 동일하게 다시 한 번 “Kim”를 기준으로 ‘인스펙트’를 한 결과, “85565”, “85564” 두 노드가 해당 기간에 “25858”와 비슷한 패턴을 보인 것을 확인했다. 결국 “Choi”, “25858”, “85565”, “85564” 이렇게 네 노드를 “Kim”과 “Ham” 사이의 자금세탁에 활용된 대포통장으로 파악할 수 있었다.

6. 결론 및 향후 연구

본 연구에서는 수사 과정에서 발생하는 비효율성을 시각적 분석을 통하여 해소하고자 수사 방법과 환경에 부합하는 과업과 디자인 요구사항을 도출한 뒤, 시각적

분석 시스템을 디자인하고, 프로토타입을 구현하였다. 해당 프로토타입을 활용한 수사 활용 시나리오를 제시함으로써, 본 연구에서 제시하는 시각적 분석 시스템이 실제 효과적으로 활용될 수 있는 방향을 제시하였다.

그러나 몇 가지 한계점 또한 존재한다. 첫 번째는 적용 가능한 데이터 종류의 한계이다. 실제 수사 환경에서 분석 대상이 되는 데이터의 종류는 본 연구에서 활용한 계좌 거래, 통화 데이터 외에도 다양하므로, 적용 가능한 데이터의 범위를 확장해 나갈 것이다. 두 번째는 지원 가능한 수사 단계의 한계이다. 마지막으로, 추후 실제 현업의 전문가들을 대상으로 한 사용자 실험을 진행함으로써 연구의 타당성을 강화할 예정이다.

References

[1] B. K. Park, H. E. Kwon, H. S. Son, Y. S. Kim, S.-E. Lee, and Y. C. R. Kim, “A Case Study on Improving SW Quality through Software Visualization,” *Journal of KIISE*, Vol. 41, No. 11, pp. 935-942, Nov. 2014. (in Korean)

[2] M. Pi, H. Yeon, H. S. Son, and Y. Jang, “A Visual Analytics Technique for Analyzing the Cause and Influence of Traffic Congestion,” *Journal of KIISE*, Vol. 47, No. 2, pp. 195-206, Feb. 2020. (in Korean)

[3] J. Stasko, G. Carsten, and L. Zhicheng, "Jigsaw: Supporting Investigative Analysis through Interactive Visualization," *Information Visualization*, Vol. 7, No. 2, pp. 118-132, Jan. 2008.

[4] Hanbyul Yeon, Yun Jang, "Visual Analytics for Abnormal Event detection using Seasonal-Trend Decomposition and Serial-Correlation" *Journal of KIISE*, Vol. 41, No. 12, pp. 1066-1074, 2014.

[5] B. L. Wong, L. Zhang, and I. D. H. Shepherd, “VALCRI: Addressing European Needs for Information Exploitation of Large Complex Data in Criminal Intelligence Analysis,” *Proc. of European Data Forum*, 2014.

[6] B. L. W. Wong, and N. Kodagoda, "How Analysts Think: Inference Making Strategies," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 59, No. 1, pp. 269-273, Sep. 2015.

[7] J. D. Haider, P. Seidler, M. Pohl, N. Kodagoda, R. Adderley, and B. L. W. Wong, "How Analysts Think: Sense-making Strategies in the Analysis of Temporal Evolution and Criminal Network Structures and Activities," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61, No. 1, pp. 193-197, Sep. 2017.

[8] J. Islam, C. Anslow, K. Xu, W. Wong, and L. Zhang, "Towards Analytical Provenance Visualization for Criminal Intelligence Analysis," *Proc. of Computer Graphics & Visual Computing*, pp. 17-24, 2016.

[9] K. Zákopčanová, Ř. Marko, B. Jozef, P. Daniel, S. Sergej, and K. Barbora, "Visilant: Visual Support for the Exploration and Analytical Process Tracking in Criminal Investigations," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 27, No. 2, pp. 881-890, Feb. 2021.

[10] Kiho Seo, "A Study of Identifying a Group of Users Based on Visualization Analysis of Users' Events", *Journal of Digital Forensics*, Vol. 13, No. 2, pp. 111-126, June. 2019.

[11] K. Xu, S. Attfield, T. J. Jankun-Kelly, A. Wheat, P. H. Nguyen, and N. Selvaraj, "Analytic Provenance for Sensemaking: A Research Agenda," *IEEE Computer Graphics and Applications*, Vol. 35, No. 3, pp. 56-64, May-June 2015.

[12] N. Goyal, G. Leshed, and S. R. Fussell, "Effects of Visualization and Note-taking on Sensemaking and Analysis," *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2721-2724, 2013.

[13] Y. B. Shrinivasan, and J. J. V. Wijk, "Supporting the Analytical Reasoning Process in Information Visualization," *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1237-1246, 2008.

[14] Y.-A. Kang, and J. Stasko, "Examining the Use of a Visual Analytics System for Sensemaking Tasks: Case Studies with Domain Experts," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18. No. 12, pp. 2869-2878, Dec. 2012.

[15] J. Xu, and H. Chen, "Criminal Network Analysis and Visualization," *Communications of the ACM*, Vol. 48, No. 6, pp.100-107, June 2005.

[16] C. North, R. Chang, A. Endert, W. Dou, R. May, B. Pike, and G. Fink, "Analytic Provenance: Process+ Interaction+ Insight," *Proc. of CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pp. 33-36, 2011.

[17] IBM Security. (2017, Mar). IBM i2 Analysts's Notebook [Online]. Available:https://www.ibm.com/downloads/cas/QNGO6RNA (downloaded 2022, Aug. 16)

[18] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN," ACM Transactions on Database Systems, Vol. 42, No. 3, pp. 1-21, Sep. 2017.

[19] Sunghyo Yoon. (2022, Mar 24), Open 300 fake bank accounts to launder 400 billion won in criminal funds [Online]. Avaliable:https://news.nate.com/view/20220324n26408 (downloaded 2022, Aug. 16)

[20] Sangguk Kwon. (2013, Jun 26). 16.7 billion wons in profits from operating gambling sites overseas [Online]. Available: http://www.busan.com/view/busan/view.php?code=20130626000160 (downloaded 2022, Aug. 16)

정 석 원

2019년 서울대학교 컴퓨터공학 학사 2019년~현재 서울대학교 컴퓨터공학부 석박사통합과정. 관심분야는 인간-컴퓨터 상호작용, 정보 시각화

신 동 화

2013년 한국과학기술원 전산학 학사 2021년 서울대학교 전기·컴퓨터공학부 박사. 2022년~현재 서울대학교 지능형컴퓨팅사업단(박사후연구원). 관심분야는 인간-컴퓨터 상호작용, 정보 시각화

복 진 욱

2015년 서울대학교 컴퓨터공학 학사 2015년~현재 서울대학교 컴퓨터공학부 석박사통합과정. 관심분야는 인간-컴퓨터 상호작용, 정보 시각화

박 석 현

2020년 서울대학교 컴퓨터공학, 정보문화학 학사. 2020년~현재 서울대학교 컴퓨터공학부 석박사통합과정. 관심분야는 인간-컴퓨터 상호작용, 데이터시각화

전 현

2020년 포항공과대학교 컴퓨터공학 학사 2019년~현재 서울대학교 컴퓨터공학부 석박사통합과정. 관심분야는 정보 시각화, 차원 축소.

서 진 욱

1995년 서울대학교 계산통계학과 전산과학 학사. 1997년 서울대학교 컴퓨터공학 석사. 2005년 메릴랜드 대학교 컴퓨터공학 박사. 2005년~2006년 Children's Research Institute 연구원. 2006년~2008년 Children's Research Institute & George Washington University School of Medicine 조교수. 2009년~현재 서울대학교 컴퓨터공학부 전임교수. 관심분야는 인간-컴퓨터 상호작용, 정보 시각화

이 인 수

1992년 연세대학교 수학 학사. 1994년 연세대학교 수학 석사. 2005년 고려대학교 수학 박사. 1996년~2000년 한국정보보호센터 연구원. 2000년~2007년 (주)비씨큐어 암호기술연구소장. 2007년~2020년 대검찰청 검찰사무관. 2020년~현재 대검찰청 검찰수사서기관. 관심분야는 암호학, 디지털포렌식, 정보보호

박 수 영

2014년 고려대학교 정보보호학 석사 2014년~2017년 (주)안랩 연구원. 2018년~현재 대검찰청 검찰수사관. 관심분야는 디지털포렌식, 데이터 분석