

Research on methods for identifying and recognizing same person in messenger chat rooms

Jong Jin Jung
 Korea Electronics Technology Institute
 Seoul, Korea
 mozzalt@keti.re.kr

Kyung Won Kim
 Korea Electronics Technology Institute
 Seoul, Korea
 kwkim@keti.re.kr

Jong Bin Park
 Korea Electronics Technology Institute
 Seoul, Korea
 jpark@keti.re.kr

Abstract—This paper describes a technology that extracts the necessary information and infers these information so that words referring to the same person, such as a person's name, position, nickname, and title described in a messenger chat room, can be identified as a single person among the data extracted through mobile phone forensics from a suspect's digital device during a criminal investigation.

Keywords—Same person, Identification, Digital Forensic, Knowledge graph (key words)

I. INTRODUCTION

As the Internet and digital devices have become popular, most people's information activities are conducted through smartphones, so the analysis of data extracted through mobile phone forensics in criminal investigations has become a very important factor in identifying the circumstances of a crime and proving the charges. Analysis of the contents of messenger chat rooms, emails, text messages, etc. obtained through mobile phone forensics extraction is very important in determining whether there is a criminal suspicion, motive for the crime, circumstances of prior conspiracy, joint crime and relationship, specific means of crime, etc., and investigators spend a lot of time and effort analyzing this. In the case of chat room analysis, which takes up the most time in analyzing digital evidence, if a single chat room is maintained for a long time, there are too many conversations, making it difficult for a person to read the conversation content one by one to detect criminal suspicion [1]. To improve this, in the conventional technology, search results are retrieved by entering a search term at the keyword level, or the time of suspected crime is directly specified to look at only the relevant conversation. In addition, in order to examine the relationship between major criminal suspects, the level is to directly select people as a node and directly specify their relationship to designate a relationship line. The most regrettable aspect of existing technology is that in the process of finding out the relationship between the main suspects, if the analyst fails to recognize not only the participants in the chat room (participants in one-on-one chat rooms, group chat rooms), but also the people mentioned in the conversation, their titles, etc., even though they are the same person, there are many cases where they are recognized as different people, or they are not recognized as people at all and are missed in the analysis. Therefore, improvement is needed in this area.

II. USE IDENTIFICATION OF THE SAME PERSON

A. Identification process

This is the overall process of analyzing data that is extracted from data obtained by digital forensics of a suspect's digital device (mobile phone) during a criminal investigation, and data that contains conversations with others in chat rooms, stored in text format. This is the process of analyzing the similarity of surrounding words/sentence expressions to various terms that refer to a specific person, to identify the same person. The process of identifying the same person is largely composed of a) the process of clustering by identifying the frequency of the knowledge set extracted in the SPO structure and the similarity of surrounding words, and b) the process of connecting the clustered and basic contact information for a representative specific person as “person-name” or “person-title-nickname”, as shown in Fig. 1 below.[2]

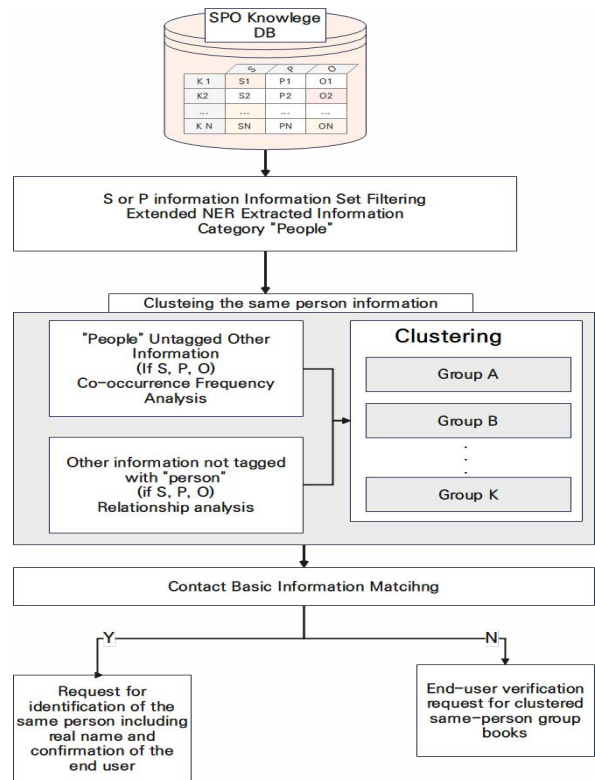


Fig 1. The same person identification process

knowledge in the form of SPO extracted after analyzing the chat room.[6]

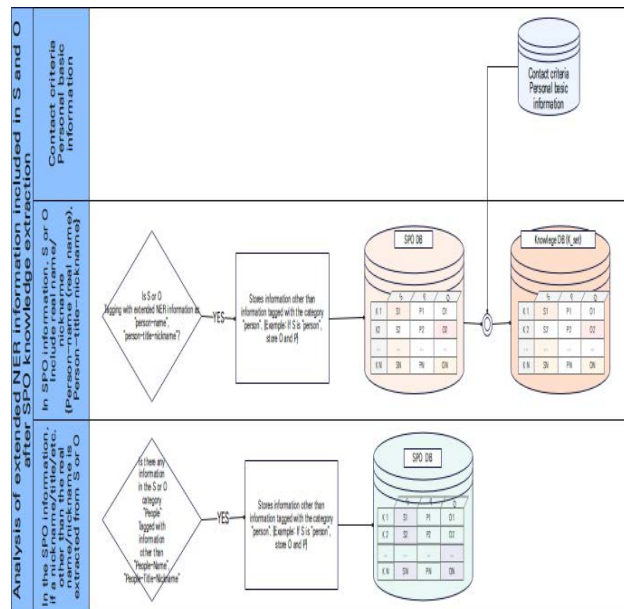


Fig. 54. How to create an extended SPO database by combining the extended NER criteria classification Person-X-X information with basic contact information among SPO knowledge

First of all, the process that is performed first is, among the knowledge extracted in the form of SPO, for the information extracted as “person-name” or “person-alias-position” in S or O, the expanded NER extracted information is matched with the contact information for the corresponding person recorded in the mobile phone, and if the company name is extracted, the position information that

fits the company type is matched and key-bound. If the company name of a specific person included in the contact information among the information included in the basic information is written, it is key-bound by matching it with the position information defined in the book according to the company type. The bound position information is used as a major factor in the weighted identification of “person-alias-position” in S or O. The second step performs clustering on information that has not yet been identified as a specific person, that is, information that has been extended NER-processed to “person-X-X”, excluding “person-name” and “person-title-nickname” in the S or O of the SPO structure.

ACKNOWLEDGMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2023-00225661, Development of causal reasoning and expression technology to enhance the proof of digital evidence)

REFERENCES

- [1] Hee-Dou Kim, Heuseok Lim, A Named Entity Recognition Model in Criminal Investigation Domain using
- [2] K. R. Rahem & N. Omar. (2014). Drug-related crime information extraction and analysis. Proceedings of the
- [3] A. Alkaff & M. Mohd. (2013). Extraction of nationality from crime news. Journal of Theoretical and Applied
- [4] S. Sathyadevan, M. S. Devan & S. S. Gangadharan (2014). Crime analysis and prediction using data mining.
- [5] M. Asharef, N. Omar & M. Albared. (2012). Arabic named entity recognition in crime documents. Journal of
- [6] P Gohel. (2016) Crime information extraction from news articles. M Tech Dissertations. Dhirubhai Ambani