KI - Künstliche Intelligenz (2022) 36:163–169
https://doi.org/10.1007/s13218-022-00762-w

SYSTEMS DESCRIPTION

MoNA: A Forensic Analysis Platform for Mobile Communication

Michael Spranger[1] · Jian Xi[1] · Lukas Jaeckel[1] · Jenny Felser[1] · Dirk Labudde[1]

Received: 27 September 2021 / Accepted: 26 April 2022 / Published online: 24 May 2022

Abstract
Mobile communication devices are a popular means of planning, commissioning and carrying out criminal offenses. In particular, data from messengers such as WhatsApp or Telegram often contain conclusive information. Organized crime also usually involves many devices, but not all of them contain the full history of communication. Rather, it is heavily fragmented due to individual deletions of messages or different joining times to groups. A singular evaluation of individual devices is therefore often not expedient, since important relationships cannot be recognized. Furthermore, communication is often distributed across different channels and modalities and can only be fully and correctly understood through a joint semantic analysis. The linking of related communications of different devices enables an almost complete reconstruction of the communication with a simultaneous reduction in reading effort by merging identical messages. Grouping coherent messages into conversations enables efficient comparison with a knowledge model. Building such a model is complex, but can be supported by a term recommender system. In this paper, MoNA is presented as a platform that implements these approaches and enables an assisted analysis of mobile communications.

Keywords Digital forensics · Mobile communication · Expert system

1 Introduction

Mobile communication devices have become an integral and indispensable part of everyday communication. For example, the number of smartphone users in Germany has risen since its invention to currently over 60 million [19] which corresponds to around 72% of Germany's current population. This is also accompanied by a steady increase in the use of these devices for planning, commissioning and carrying out criminal activities, and thus in the number of devices to be analyzed in the course of criminal investigations.The challenge here is to search through the huge amount of communication data on a device for the often little case-relevant information. In the case of organized crime, even entire networks of mobile communication devices usually have to be examined.

The forensic investigation of mobile communication devices includes, on the one hand, the physical and logical backup and reconstruction of data on mobile devices, such as smartphones or tablets, and, on the other hand, the content analysis of text, image, audio and video data. Together with available metadata, such as timestamps, log files, geo and contact data, apps used, etc., a variety of forensic questions can be answered and a kind of digital profile of the user can be generated.

While much of the work addresses the provisioning and recovery of data on mobile devices [10, 12, 15] or the exploration of database structures [1, 2, 5, 20], little work is dedicated to their content analysis.

In text-based communication, such as SMS or various messenger services, understanding the content of the conversation depends heavily on the existence of a chat history that is as complete as possible. However, this is often not the case, and there are many reasons for this. However,

✉ Michael Spranger
spranger@hs-mittweida.de

Jian Xi
xi@hs-mittweida.de

Lukas Jaeckel
jaeckel1@hs-mittweida.de

Jenny Felser
felser@hs-mittweida.de

Dirk Labudde
labudde@hs-mittweida.de

[1] University of Applied Sciences Mittweida, Faculty Applied Computer Sciences & Biosciences, Technikumplatz 17, 09648 Mittweida, Germany

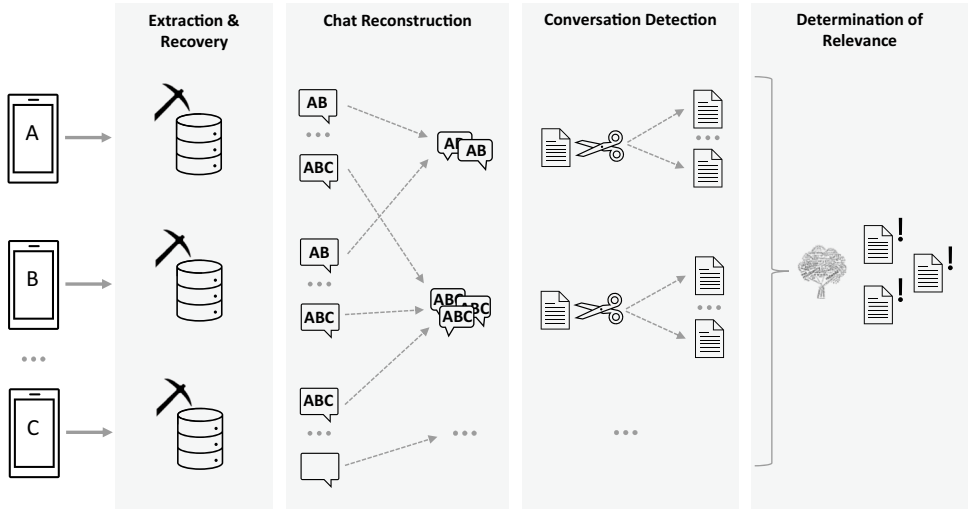

Fig. 1 Interactive process for intelligent analysis of communication data. After the extraction and recovery of chat histories, identical chats from different devices are merged. This is followed by the detection of conversations for each chat, each of which contains contiguous messages. Finally, a knowledge model is used to classify conversations that are relevant to the process and return them to the user for evaluation

reconstruction is possible, especially in group chats, by linking different devices. A positive side effect is the drastic reduction of the analysis effort, since it is now sufficient to analyze the chat history of a single conversation participant. The subsequent detection of coherent conversations enables a more error-tolerant search, while preserving the context. This is particularly necessary for the later assessment of the meaning and significance by an investigator.

When assessing the case-specific relevance of parts of the communication, classic machine learning models fail due to the lack of availability of annotated training data and the special characteristics of mobile communication in the forensic context. This can be remedied by a knowledge model that incorporates the investigator's experience and case-specific knowledge. However, the typically challenging and time-consuming creation of such a model can be supported by incorporating a term recommendation system.

In this paper, we present MoNA, a prototype application that implements the aforementioned problem-solving approaches.

2 MoNA's Analysis Process

The focus of the analysis process integrated in MoNA is on linking communication data and reducing it to case-relevant data. As can be seen in Fig. 1, the entire process can be divided into several steps. Most of these steps are performed automatically, but some of them actively involve the user.

After the initial extraction of chat histories from the respective databases on each device, deleted content is recovered using the Forensic SQLite Data Recovery Tool [14] to fill in gaps. By comparing all entries in the backups with the current main database, deleted messages or entire chat histories can be recognized and restored. If every backup has a time stamp, the period in which the respective data was deleted can also be determined.

Subsequently, identical chats from different devices are each merged into a single common chat, which represents the maximum amount of available information.

This is followed by conversation detection, which divides a chat into several coherent conversations. A conversation consists of a set of related messages, which are in a common temporal context. It is precisely these conversations, and not the individual messages, that are ultimately used to determine relevance [16–18].

In order to correctly classify conversations as case relevant, a knowledge model is required that enables arbitrarily complex search queries that go far beyond classic text searches. In MoNA, this model is called a term tree. When creating the term tree, the user's knowledge can be included, e.g., specific case knowledge. After applying the term tree, the system automatically returns all conversations that contain at least one message classified as case relevant. As mentioned at the beginning, the user only has to assess a usually significantly reduced number of conversations compared to the total amount of data.

2.1 Reconstruction of Chat Histories

It is entirely possible that certain data cannot be restored on a mobile device. Nevertheless, incomplete or deleted chats can be reconstructed, provided that not one but several devices are present in the context of the forensic examination. For this purpose, MoNA searches for identical chats on all of these devices.

Subsequently, all messages of the same chats are compared. The basic idea is that messages that were deleted on one mobile device can still be present in the same chat on other devices. Therefore, if a chat exists on two devices, but certain messages of the chat are only on one device, these

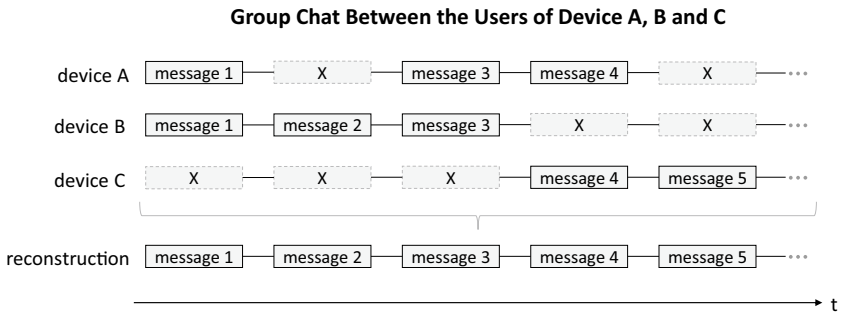

Fig. 2 Complete reconstruction of a group chat using three devices. Devices A, B and C each contain their own versions of the common group chat, which have gaps at different points in time. A gap marked as dashed and with "X" represents a deleted or not received message. Merging the different versions leads to the reconstruction of the complete chat history

messages were certainly deleted or not received on the other device. To check whether two messages are identical, a comparison of the message ID or the message content in combination with the transmission time is suitable, depending on the circumstances. If deleted or not received messages were found, MoNA automatically inserts them into the respective gapped chats. The chance of being able to completely reconstruct as many chats as possible is higher the more devices are included in the analysis process. Since the content of identical chats is guaranteed to be the same at the end of the process step, both the user and MoNA only need to analyze these chats in a single, complete version. For demonstration purposes, Fig. 2 shows an example of reconstructing a group chat using three devices, each of which contains parts of the entire chat history.

2.2 Conversation Detection

After the recovery and reconstruction of individual chat messages $m \in M$ described in the last sections, their grouping into individual conversations is done as shown in Equation 1.

$$c = (m_1, ...m_n | t_i^m - t_{i+1}^m <= \epsilon, \forall i = 1...n) \quad (1)$$

Each conversation thus consists of the set of messages that were exchanged consecutively at times t_i and t_{i+1}, without exceeding an individually determined maximum response time ϵ as detailed in [16]. This grouping makes the use of conservative word matching algorithms more promising, since the larger number of words in a conversation naturally increases the probability of a search hit compared to individual messages.

2.3 Term Tree

After determining the conversations $C = c_1, ..., c_n$, the next step is to find out which of these conversations contain crime-related relevant messages. Determining relevant conversations regarding a specific case, requires the inclusion of investigator knowledge for several reasons, as detailed in [16].

Taking into account the fact that a relevance decision is based to a not insignificant extent on empirical knowledge, MoNA relies on a rule-based approach that allows complex systems of syntagms to be described in a tree structure. This knowledge structure, called a term tree, is represented in Fig. 3. A syntagma is a set of linguistic elements (here terms) that occur together in a local context (here message). A term is represented not only by a word, but by a vector $\mathbf{t} = (w_0, ..., w_n, p_0, ..., p_k)$, where w_i denotes a set of linguistic variations (word variants, synonyms, group-specific expressions, etc.) and p_i denotes a set of pattern definitions (here regular expressions).

A syntagma *syn* is then the obligatory combination of different terms t_i in a message m_j in the sense of a conjunction, i.e. $syn = t_0 \wedge t_1 \wedge ....$ A term tree $\xi = syn_0 \vee syn_1 \vee ...$ is then the disjunctive conjunction of different syntagms. Of course, this principle can be applied recursively, i.e. $\xi_{total} = \xi_0 \vee \xi_1 \vee ...$, allowing the reuse of cross-case or case-independent knowledge encoded in this way, which successively limits the generation effort to term trees for case-specific knowledge.

If at least one syntagma matches a conversation, it is classified as case-relevant and highlighted accordingly.

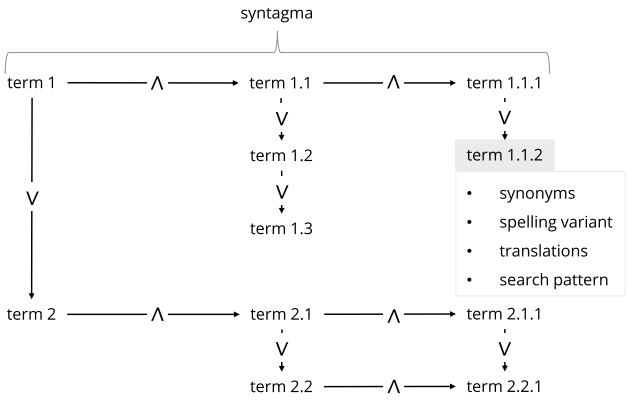

Fig. 3 Term tree as a central classification element for deciding relevance of individual conversations

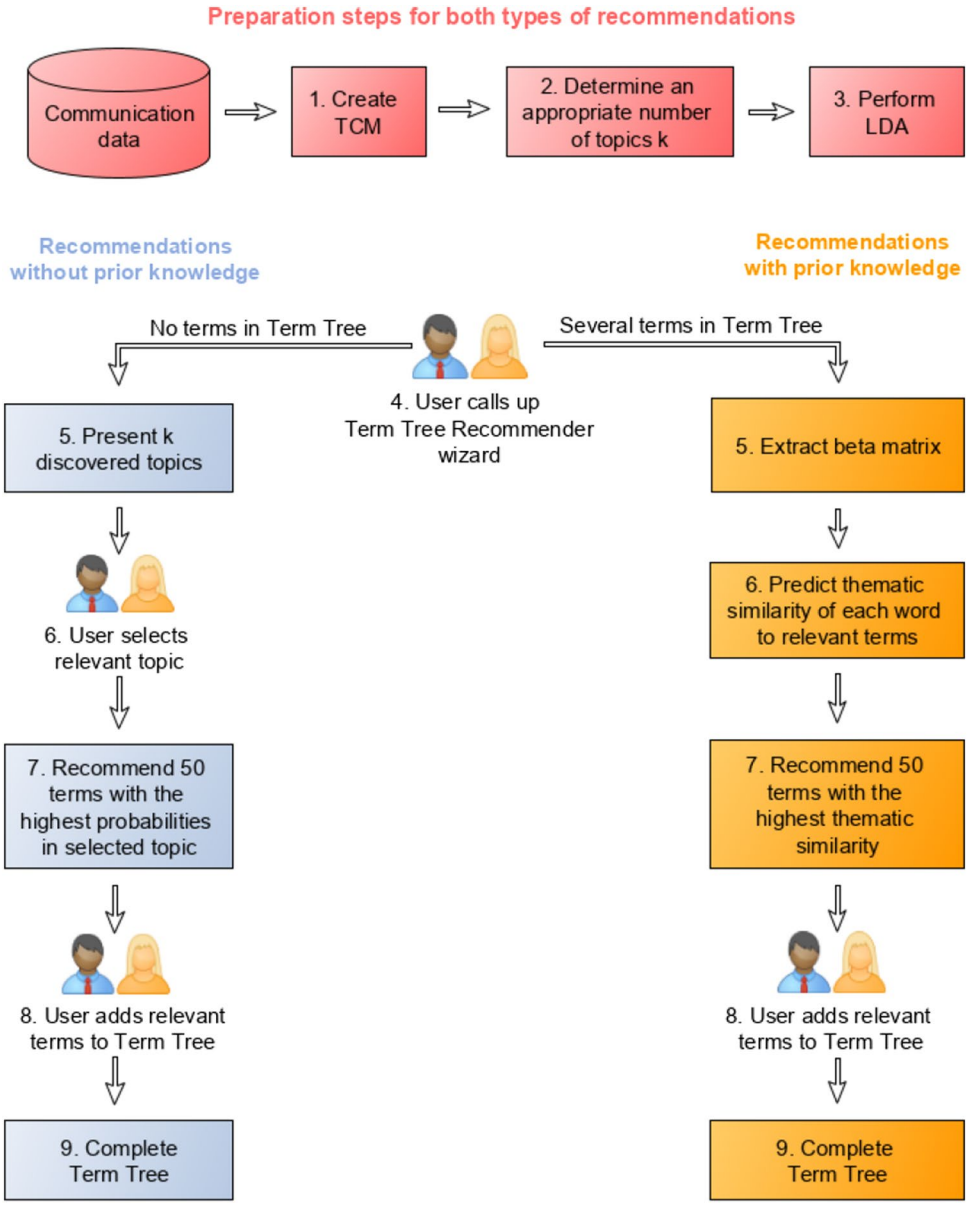

Fig. 4 Structure of the Term Tree Recommender. The Recommender can provide recommendations with or without prior knowledge. The algorithm is based on an LDA over the entire communications network

3 Term Recommendations

Probably the biggest challenge in creating the knowledge base is finding appropriate terms to generate the syntagms. The approach of a term recommendation engine as implemented in MoNA is shown schematically in Fig. 4.

Here, a topic modeling based on a term co-occurrence matrix with the help of the Latent Dirichlet Allocation (LDA) [4] forms the basis for the recommendation of terms. The optimal number of topics can be set arbitrarily or determined by probabilistic coherence as shown in [8]. If the term tree does not contain any terms, i.e., there is no prior knowledge, the investigator can then choose the topic that is likely to best answer the current forensic question. Subsequently, the 50 terms that are most likely to be represented in the chosen topic are suggested. The term tree can then be supplemented with case-relevant terms from this selection.

Much more interesting is the case that some words or syntagms are already contained in the term tree. These entries can, for example, be based on knowledge from the interrogations of witnesses or suspects, but also on recommendations without prior knowledge, as just discussed. Griffiths et al. [7] explain how the LDA can be used to predict whether w_{n+1} is the next word in a sequence $s = w_1, w_2 ... w_n$ of words. However, they point out that the words in s do not necessarily have to follow each other in the text, but that it can also be an unordered set of terms [6, 7].

In this work, s consists of the already known, relevant words in the term tree. For another word w_{n+1}, a prediction is made whether it is associated with the existing important

words in s or whether it is thematically similar to them. The only requirement imposed on the set of words in s is that the contained terms have a high probability in the same topic [7]. This applies to the existing words in the term tree at least if they were determined from one of the case-relevant topics. The probability that w_{n+1} is associated with s can be formulated as a conditional probability $p(w_{n+1}|s)$ [6] and calculated using Equation 2.

$$P(w_{n+1|s;\beta}) = \frac{\sum_t \prod_{i=1}^{n+1} \beta_{w_i}^{(t)}}{\sum_t \prod_{i=1}^{n} \beta_{w_i}^{(t)}} \tag{2}$$

The β-matrix computed during LDA is now needed for prediction because it contains the $\beta_{w_i}^{(t)}$, i.e., the probability of the word w in topic t [7]. If w_{n+1} is thematically similar to the words s, it has a high probability in the same topics as the words in s, which leads to a higher value for $P(w_{n+1|s;\beta})$ [7]. Similar to the recommendations without consideration of prior knowledge, the 50 terms with the highest thematic similarity to prior knowledge are suggested to the investigator. Both approaches generally lead to similar recommendations, but incorporating prior knowledge provides more specific terms, at least in part.

4 Joint Semantic Analysis

Up to this point, the communication of a single communication channel, e.g., a service like WhatsApp or Telegram, across all participating devices, was considered. But communication in real life is not limited to one channel and especially not to one modality. The term modality is used to refer to a communication medium, such as text, image audio or video. We have to take into account that people do not communicate by just writing texts. Rather, texts are interspersed with images, videos and voice messages. These different modalities can add important new or repeat existing information. Especially in the first case, a common understanding of all the modalities used is necessary in order to correctly understand the communication and subsequently transmitted information. Details of the approach described below have already been discussed by the authors in [21].

Aiming at explaining the coherent semantic content and hidden connections in a mobile communication consistently, we formally formulate the joint semantic analysis as follows:

$$\tilde{e} = \text{argmax}_{\theta} \tilde{P}(e|d_{cm};\theta) \tag{3}$$

where e is the semantic context in the conversation data D, which is mostly represented by a topic and possibly connected to a concrete crime, $d_{cm} \in D$ denotes a single message spread via the communication channel $c \in D_c = \{WhatsApp, Telegram, email...\}$ and represented in the modality $m \in D_m = \{Text, Image, Audio, Video ...\}$. D is temporally and semantically coherent and chronologically structured. We use θ to denote the set of parameters inferred during topic modeling that captures the latent semantics in the data.

The crucial task is to find an intermodal relationship that implies a semantic concept between different modalities and channels. Therefore, at first a textual representations for all non-textual modalities has to be determined. At first, we need to map the content of all multimedia data into a textual semantic space in order to extract topics. In this way, the entire communication space becomes searchable. Subsequently, the semantic linking (intermodal correspondence) can be determined by considering the entire context in communication.

For image data, the traditional classification approach [23] or image captioning [11] can be used, where the former delivers only discrete labels like people or car, etc., while the latter describes the coherent information of image as a whole scene with a natural sentence, e.g., a man is holding a gun in a bank. Instead of focusing on describing semantic content of an image, the semantic interpretations and the relations between image and text can be determined as shown in [13]. In future research, a scene graph will be extracted to determine how it contributes to the understanding of a conversation [22]. Similar, a video can also be translated to a textual representation, i.e., a natural sentence with respect to the content [9]. The audio data can be transcribed into text form by means of Automatic Speech Recognition (ASR) [3]. Once the semantic textual representation of the multimedia data is available, the coherent semantic topics of the data can be extracted by using LDA.

5 Conclusion and Future Work

Analyzing communication data from mobile devices is a time-consuming and error-prone task. Current analysis applications can support this process so far, but do little to reduce the effort. In this work, therefore, a process chain was presented that significantly reduces the analysis effort after extraction and recovery of the communication data in three steps.

In a first step, the messenger data from different devices is linked together, allowing the communication history to be reconstructed almost completely. As a positive side effect, deduplication significantly reduces the reading effort. The next step is to divide the chat history messages into temporally related conversations, which preserves the context of a message and reduces the need to match each relevant message. At the same time, the associated context preservation simplifies the interpretation of the results by investigators. In a final step, process-relevant conversations are filtered with the help of a semantic rule-based knowledge base,

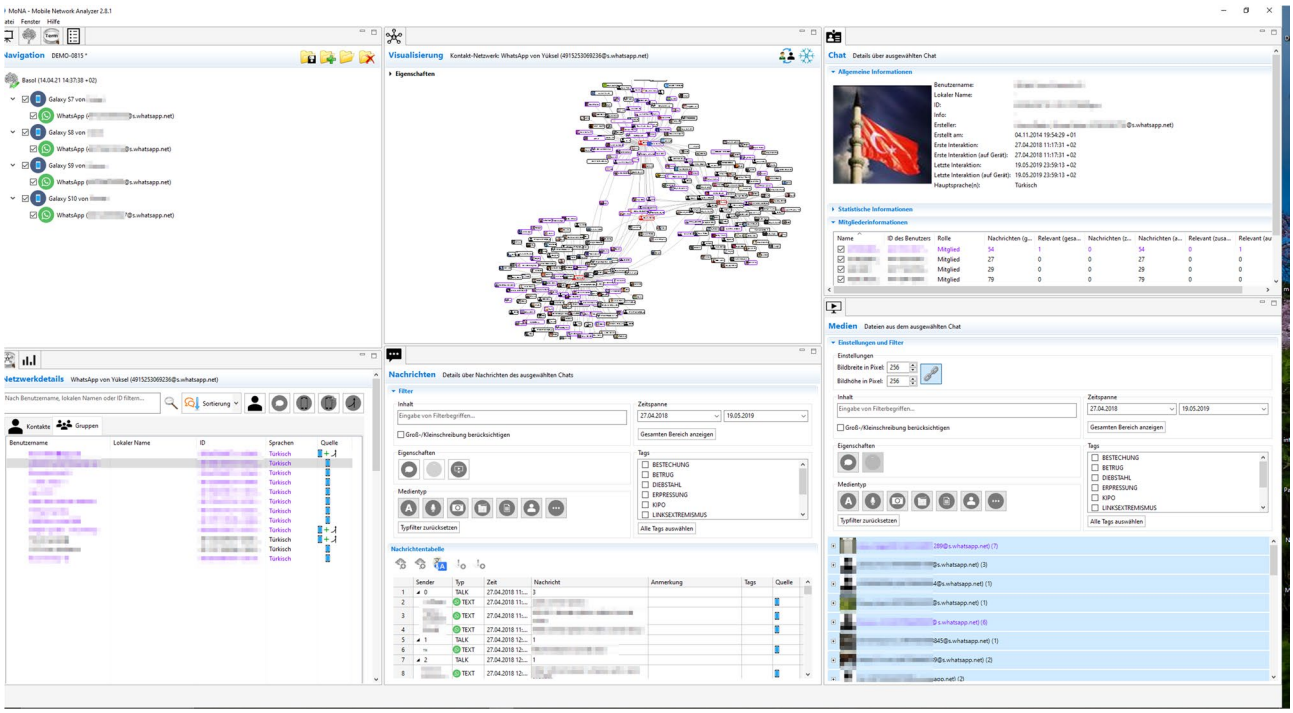

Fig. 5 Screenshot of the current MoNA version in which the presented concepts are implemented

the so-called term tree, which again drastically reduces the search effort.

To reduce the effort of knowledge base creation, a two-stage recommendation system based on extracted topics and semantic coherence can be used. In particular, semantically and thematically similar terms are suggested for already existing case-relevant knowledge of the investigator. Since real communication usually spans different channels and modalities , all media data are mapped into a common text-based semantic space and in this way jointly included in the determination of case-relevant communication.

The proposed process chain has been implemented in a prototype application, the Mobile Network Analyzer (MoNA), and is currently being evaluated by various investigative agencies (see Fig. 5). A time-limited trial version is available for download[1], along with instructions for obtaining unrestricted usage rights. Nevertheless, the current implementation is limited to the messengers WhatsApp, Facebook and Telegram. Future research will address the reverse engineering of further communication services. The results will be integrated into upcoming versions of MoNA at irregular intervals. Furthermore, the joint semantic analysis of different modalities currently places high demands on the hardware and is therefore not included in the trial version.

Future work needs to address empirical evaluation to highlight the superiority of the presented approach compared to existing solutions. Opportunities for further development exist primarily in the formation of user profiles through stylistic analyses across different communication channels.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Anglano C (2014) Forensic analysis of whatsapp messenger on android smartphones. Digit Investig 11(3):201–213. https://doi.org/10.1016/j.diin.2014.04.003
2. Anglano C, Canonico M, Guazzone M (2017) Forensic analysis of Telegram Messenger on Android smartphones. Digit Investig 23:31–49. https://doi.org/10.1016/j.diin.2017.09.002

[1] https://www.hs-mittweida.de/spranger/.

3. Baevski A, Zhou Y, Mohamed A, Auli M (2020) wav2vec 2.0 A framework for self-supervised learning of speech representations. In: Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H (eds) Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS, pp 6–12
4. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022
5. Chang M, Yen CP (2020) Evidence gathering of facebook messenger on android. Int J Netw Secur 22:828–837
6. Griffiths T, Steyvers M (2002) A probabilistic approach to semantic representation. In: Proceedings of the 24th Annual Conference of the Cognitive Science Society, pp 381–386
7. Griffiths TL, Steyvers M, Tenenbaum JB (2007) Topics in semantic representation. Psychol Rev. https://doi.org/10.1037/0033-295X.114.2.211
8. Hay Mele B, Russo L, D'Alelio D (2019) Combining marine ecology and economy to roadmap the integrated coastal management: a systematic literature review. Sustainability. https://doi.org/10.3390/su11164393
9. Iashin V, Rahtu E (2020) Multi-modal dense video captioning. CoRR abs/2003.07758
10. Jeon S, Bang J, Byun K, Lee S (2012) A recovery method of deleted record for sqlite database. Pers Ubiquitous Comput 16:1–9. https://doi.org/10.1007/s00779-011-0428-7
11. Karpathy A, Fei-Fei L (2015) Deep visual-semantic alignments for generating image descriptions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp 3128–3137
12. Liu Y, Xu M, Xu J, Zheng N, Lin X (2017) Sqlite forensic analysis based on wal. In: Deng R, Weng J, Ren K, Yegneswaran V (eds) Security and privacy in communication networks. Springer International Publishing, Cham, pp 557–574
13. Otto C, Springstein M, Anand A, Ewerth R (2019) Understanding, categorizing and predicting semantic image-text relations. In: Proceedings of the 2019 International Conference on Multimedia Retrieval, pp 168–176
14. Pawlaszczyk D (2021) Forensic SQLite data recovery tool. https://www.staff.hs-mittweida.de/~pawlaszc/fqlite/. Accessed 11 May 2022
15. Pawlaszczyk D, Hummert C (2021) Making the invisible visible - techniques for recovering deleted sqlite data records. Int J Cyber Forensics Adv Threat Investig 5:5
16. Spranger M, Heinke F, Appelt L, Puder M, Labudde D (2016) MoNA: automated identification of evidence in forensic short messages. Int J Adv Secur 9(1&2):14–24
17. Spranger M, Labudde D (2014) Semantic tools for forensics: towards finding evidence in short messages. In: Schmidt A, Yarali A (eds) Proceedings International Conference on Advances in Information Mining and Management (IMMM). IARIA, Paris, France, pp 1–4
18. Spranger M, Labudde D (2017) Textforensik. In: Labudde D, Spranger M (eds) Forensik in der digitalen Welt. Springer Spektrum Akademischer Verlag, Amsterdam, pp 167–198. https://doi.org/10.1007/978-3-662-53801-2_6
19. Tenzer F (2021) Anzahl der Smartphone-Nutzer in Deutschland bis 2020. https://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonenutzer-in-deutschland-seit-2010/. Accessed 11 May 2022
20. Thebaity MA, Mishra S, Shukla MK (2020) Forensic analysis of third-party mobile application. HELIX 10:32–38. https://doi.org/10.29042/2020-10-4-32-38
21. Xi J, Spranger M, Labudde D (2021) A concept for a comprehensive understanding of communication in mobile forensics. In: Bhulai S, Semanjski I, Sztandera I (eds) Proceedings international conference on data analytics. IARIA, Barcelona, Spain, pp 74–76
22. Xu D, Zhu Y, Choy CB, Fei-Fei L (2017) Scene graph generation by iterative message passing. IEEE Conf Comput Vis Pattern Recogn (CVPR). https://doi.org/10.1109/CVPR.2017.330
23. Zhai X, Kolesnikov A, Houlsby N, Beyer L (2021) Scaling vision transformers. CoRR abs/2106.04560