Ontology population framework of MAGNETO for instantiating heterogeneous forensic data modalities

Ernst-Josef Behmer[1], Krishna Chandramouli[2], Victor Garrido[3], Dirk Mühlenberg[1], Dennis Müller[1], Wilmuth Müller[1], Dirk Pallmer[1], Francisco J. Pérez[3], Tomas Piatrik[4], Camilo Vargas[4]

[1] Fraunhofer Institute for Optronics, System Technologies and Image Exploitation IOSB, Karlsruhe, Germany, {ernst-josef.behmer, dirk.muehlenberg, wilmuth.mueller, dirk.pallmer}@iosb.fraunhofer.de,
[2] Venaka Media Limited, London, UK, k.chandramouli@venaka.co.uk
[3]Universitat Politècnica de València, Spain, frapecar@upvnet.upv.es
[4]Queen Mary University of London, UK, t.piatrik@qmul.ac.uk

Abstract. The growth in digital technologies has influenced three characteristics of information namely the volume, the modality and the frequency. As the amount of information generated by individuals increases, there is a critical need for the Law Enforcement Agencies to exploit all available resources to effectively carry out criminal investigation. Addressing the increasing challenges in handling the large amount of diversified media modalities generated at high-frequency, the paper outlines a systematic approach adopted for the processing and extraction of semantic concepts formalized to assist criminal investigations. The novelty of the proposed framework relies on the semantic processing of heterogeneous data sources including audio-visual footage, speech-to-text, text mining, suspect tracking and identification using distinctive region or pattern. Information extraction from textual data, machine-translated into English from various European languages, uses semantic role labeling. All extracted information is stored in one unifying system based on an ontology developed specifically for this task. The described technologies will be implemented in the Multimedia Analysis and correlation enGine for orgaNised crime prEvention and invesTigatiOn (MAGNETO).

Keywords: Information Extraction, Ontology Population, Law Enforcement.

1 Introduction

The Multimedia Analysis and correlation enGine for orgaNised crime prEvention and invesTigatiOn (MAGNETO) formalizes a structural framework to enhance the operational capability of Law Enforcement Agencies (LEAs) in their fight against organized crime and terrorist organizations. Following the recent terrorist attacks reported across Europe in London, Paris, Berlin, Brussels, the critical challenges faced by respective investigative agencies include the processing of large volumes, the heterogeneity and the fragmentation of the data that officers have to analyze for the prevention, investigation and prosecution of criminal offences. This paper contains the first stage of MAGNETO, co-funded by the European Commission within the "Horizon 2020" program. The overall integrated framework is presented in **Fig. 1**. As the overall investigative framework brings together several research components, the paper highlights the research carried out in "heterogeneous data mining and evidence collection" which is interfaced with the high-level semantic reasoning component.

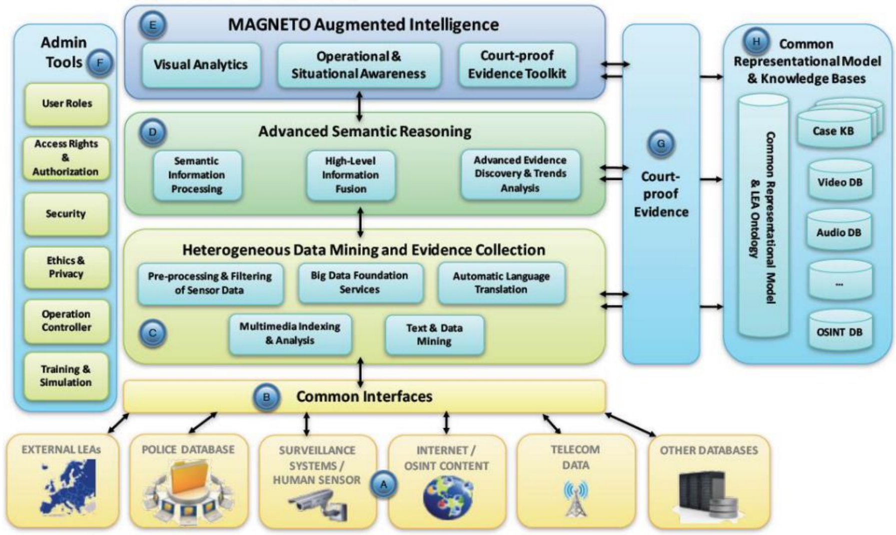


Fig. 1. Proposed framework of MAGNETO

The first step to make use of the available data is the extraction of the contained information and its representation in a machine-readable way. Due to the diversity in the origin of the different evidences, e.g. audio, video, pictures and textual data, different methods for the extraction of information are necessary. For multimedia content indexing, background/foreground subtraction provides the separation of a video stream into foreground, which is characterized by its appearance at unique moments in time, and into omnipresent background. For indexing audio streams, a speech to text module will be supplied for transforming the speech contained in audio files into text. Furthermore, the Distinctive Region or Pattern (DROP) detection module allows the retrieval of videos in large databases containing a predefined pattern of distinctive visual effect. The text mining module extracts information contained in English text data composed in natural language. Text data which is not in English language requires a translation engine to translate selected European languages to English. All available information will be stored using an ontology designed for the specific needs of the MAGNETO project.

In Section 2, we describe approaches and implementations of the different information extraction methods whereas Section 3 is devoted to the underlying ontology and its population.

2 Information Extraction Methods

2.1 Multimedia Content Indexing

In this section, we give an overview of multimedia content indexing solutions implemented in MAGNETO for analyzing heterogeneous data in the process of evidence collection.

Foreground/background subtraction

Foreground/background subtraction can be defined as a segmentation of a video stream into foreground, which appears at unique moments in time, and the background, which is always present. It is an important video processing task that contributes to higher precision and accuracy when integrated with other video analytics solutions for forensic evidence collection.

Problem Definition

Vast range of video analysis tasks often begins with background subtraction, which consists of creating a background model that allows distinguishing foreground pixels. Processing per-pixel basis from the background is not only time-consuming but can also dramatically affect foreground region detection, if region cohesion and contiguity is not considered in the model. Decomposition of a video scene into background and foreground is a very challenging but crucial process that can significantly improve performance of many other video analytics solutions. The robust subspace approach based on a low-rank plus sparse matrix decomposition has shown a great ability to identify static parts from moving objects in video sequences. However, those models are still insufficient in realistic environments. The separation of locally moving or deforming image areas from static or globally moving background is a focused video-processing task with manifold applications.

Approach and configuration used for MAGNETO

In MAGNETO we implement a new method in which we regard the image sequence to be made up of the sum of a low-rank background matrix and a dynamic tree-structured sparse matrix, and solve the decomposition using an approximated Robust Principal Component Analysis (RPCA) method extended to handle camera motion, see [1]. Furthermore, to reduce the curse of dimensionality and scale, a low-rank background modelling is integrated via Column Subset Selection [2] that reduces the order of complexity, decreases computation time, and eliminates the huge storage need for large videos. The proposed approach builds on the findings accomplished by Candès et al. [3], where the authors provided a practical solution for the long-standing problem of recovering the low-rank and sparse parts of a large matrix made up of the sum of these two components. The background part of the video sequence is modelled by the low-rank matrix, while the locally deforming parts constitute the sparse matrix component. Although the approach described by Candès et al. [3] leads to a computationally feasible solution, the complexity is still high involving the calculation of many Singular Value Decompositions (SVD) for a very large matrix.

The proposed solution implemented in MAGNETO addresses a number of critical issues and limitations of RPCA which are: embedding global motion parameters in the model, i.e. estimation of global motion parameters simultaneously with the foreground/background separation task; considering matrix block-sparsity rather than generic matrix sparsity as natural feature in video processing applications; and more critically providing an extremely efficient algorithm to solve the low-rank/sparse decomposition task. The first model aims at video sequences captured by a moving camera, by estimating the global motion parameters while performing the targeted background/foreground separation task. The second model exploits the fact that in video processing applications the sparse matrix has a very special structure. In other words, the non-zero matrix entries are not randomly distributed, but they build small blocks within the sparse matrix. Finally, the last solution targets the fact that RPCA approaches are computationally expensive. The proposed model introduces an extremely efficient "SVD-free" technique that enables real-time foreground separation.

Fig. 2. Results of foreground/background subtraction (pictures in second row represent extracted foreground objects)

The implemented method can handle camera movement, various foreground object sizes, and slow-moving foreground pixels as well as sudden and gradual illumination changes in a scene. The foreground/background subtraction (see **Fig. 2**) serves as an important pre-processing step for other video analytics modules within the MAGNETO system.

DROP detection and tracking

As the volume of information to be processed and analyzed increases exponentially, the suspect identification becomes an ever-challenging activity for the LEAs. The problem is compounded by the fact that potential suspects tend to mask their face during the act of violence. In addition, the limited resolution of the largely deployed CCTV network across several locations limits the ability of officers to successfully identify suspects. Addressing these challenges, MAGENTO platform integrates, the Distinctive Region or Pattern (DROP) [4] module that relies on the visually distinctive region as identified by the investigators to guide the search across large datasets of CCTV footage. This query region is then used as a query by example for the search of similar patterns across the large databases of CCTV footages.

Problem Definition

To illustrate a case of DROP based orchestration, a large corpus of CCTV and mobile data was used to identify an offender in a large city. The four images in **Fig. 3** outline a common case of subject identification performed by security forces. In this case the distinctive "orange zipper" (DROP) in the black jacket of the perpetrator (attacking another individual in the first image), can be used for person identification in subsequent images.

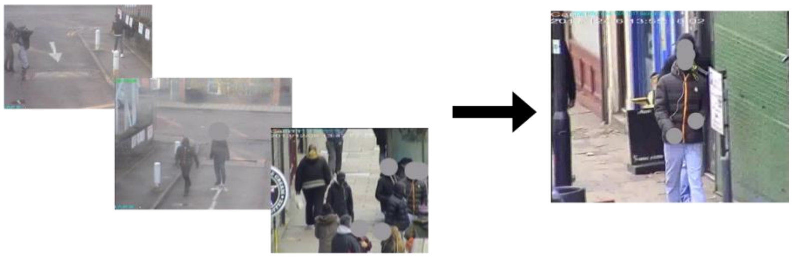


Fig. 3. Identifying offenders by DROP based orchestration. Scene of the incident (image at the left), images of strong DROP correlation (middle), final image with the same drop and clear face of the perpetrator.

Therefore, the first image illustrating the criminal act is used as starting point for the investigation with the investigator identifying the DROP to be used to filter the repository content. The next two images depict the results of the content retrieval where standardized DROP metadata is used to search the media repository. Finally, the image at the right displays a picture containing the same DROP in which the face of the perpetrator can be clearly identified. It is expected that the DROP based media retrieval contributes to solve similar and more complex cases by performing automated searches of standardized metadata in all images and videos available for case solving.

Approach and configuration used for MAGNETO

The DROP module combines a machine learning approach with a tailored cascade classification technique. The machine learning part is based on a random forest of decision trees, which is trained with the content of the video to be searched. Traditional methods train a forest with exemplars of the pattern or object class to be detected. By training the forest with the haystack, rather than the needle, it can be reused to detect any pattern that the user may choose. The cascade classification approach involves two stages: definition and search. Firstly, the target DROP is described using color and Haar-like features [5] in order to define the conditions to be matched in the search stage. Subsequently, a sliding-window based search stage is performed in the target images. The search space in the image is initially reduced by defining an attentional grid using a dominant color matching approach. Then, for each possible sliding-window in the attentional grid a set of conditions is assessed in order to compare the current region against the query pattern. If every condition is satisfied for a specific sliding-window, the region covered by the window is marked as a positive detection. The steps involved are highlighted in **Fig. 4**.

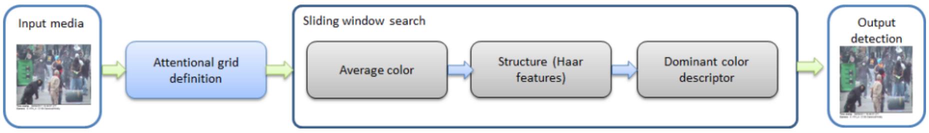


Fig. 4. - DROP search process

The extracted DROPs are indexed as bounding box coordinates to be subsequently used by the investigators for the recognition of the potential perpetrators. Each pattern generated by the investigators is treated as an image region query for which a set of video segments from the CCTV dataset is retrieved. The spatio-temporal reasoning algorithms implemented on the MAGNETO knowledge repository will facilitate the filtering of irrelevant suspects and thus simplify the process of suspect identification.

Speech to text transformation

LEA investigators deal with enormous and complex amounts of data from calls or recorded audio files; it is impossible to listen all of them to discover hidden or unsuspected connections within big datasets. This provokes expensive investigations with enormous delays solving crimes and hard time to prevent them. Multilingualism conversations are hard to decrypt and usually data sources are not contextualized. Since MAGNETO's platform provides new semantic technologies and augmented intelligence tools for indexed textual contents, a speech to text module brings to MAGNETO a solution that combined with other modules improves investigation capabilities.

Problem Definition

Nowadays telephone tapping, recorded audio files from social networks, questionings or interrogations, among others, are suitable evidences that can be used in real use cases. The problem appears when LEAs have to listen and process all these audio files. Most of the records contain irrelevant information which implies a waste of time and resources from other conducted tasks or investigations.

Actual LEA's computer infrastructures are not able to manage the different algorithms based on machine learning used in audio processing techniques yet, high performance computing is needed.

Approach and configuration used for MAGNETO

The speech to text module provides to MAGNETO a powerful tool able to combine online conversions through the APIs exposed by the main providers (e.g. GOOGLE, MICROSOFT, IBM, WIT, among others) in case that internet access at LEA premises is allowed, or offline conversions based on local dictionaries in case of security policies or privacy restrictions, as shown in **Fig. 5**

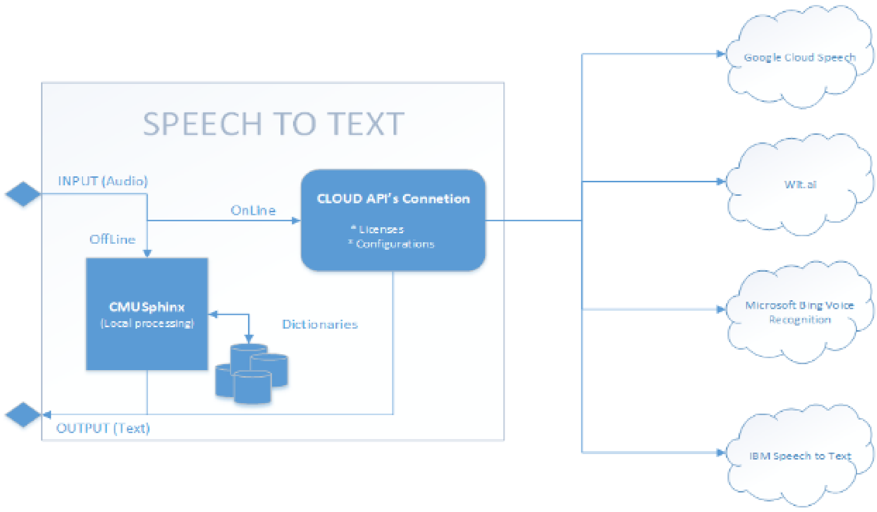


Fig. 5. Speech to text transformation flow diagram using CMUSphinx [6]

The result obtained from MAGNETO's Speech to Text module is the audio file transcription. The file can be analyzed by other MAGNETO modules responsible for pattern identification, sentimental analysis or texts classification. This automated approach is clearly more feasible than extracting information by listening directly to the audio files. LEAs will be provided with a fast and robust solution also capable to interoperate among other platform functionalities.

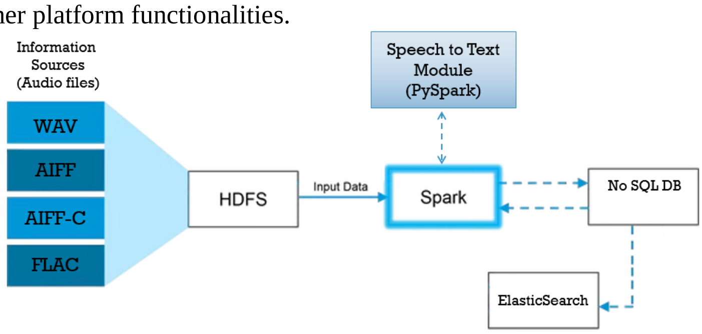


Fig. 6. Speech to text module integration

To improve the performance of the speech to text module, it has been designed taking in to account the integration into MAGNETO's Big Data Services. The solution proposed in **Fig. 6**, uses the Hadoop Distributed File System (HDFS) as a repository for all the audio files to be transcribed. Once a file is uploaded to the repository, an Apache Spark job starts the conversion, and the result of this transformation is associated as metadata to the audio file and is stored in a NoSQL database for future analysis by other modules.

2.2 Text Mining

An ongoing investigation produces vast amounts of textual data, some of which is composed in natural language with all its vagueness and ambiguity. The purpose of the text mining module is to provide the methods and algorithms for retrieving high-quality information from these textual data, which then can be incorporated into the MAGNETO knowledge database.

Problem Definition

A core problem in deriving information from textual data is the identification of information fragments. Information fragments in this context are basic concepts as names, locations, organizations, time specification, but also relations between the former. In order to extract these fragments from the data, linguistic analysis of the textual data is necessary, which then clarifies the meaning of a given word in the context of the textual data.

Approach and configuration used for MAGNETO

Text mining, i.e. the retrieval of information from textual data, has possible applications in the mining of confiscated emails and documents confiscated during house-searches, transcribed telecom/audio data, and other documents used in an investigation.

The linguistic analysis includes e.g. part-of-speech tagging, semantic role labelling (shallow parsing), named entity recognition, and capturing the meaning of a sentence in natural language. All these rely on theoretical models used either in rule-based natural language processing (NLP) or in statistical NLP, which is based on algorithms from machine learning or on a new approach using deep neural networks.

The result of the processing is an annotated document from which the relevant information can be extracted in form of a conceptual graph of pre-processed forensic information. At last, the information contained in the conceptual graph needs to be incorporated into the ontology.

For extracting information from textual data, we employ the technique of Semantic Role Labelling using Pikes [7] which tries to capture the meaning of a sentence in natural language. The goal is to represent the meaning with a rooted, directed, acyclic graph with labels on edges (relations) and leaves (concepts). Pikes performs traditional tasks such as named entity recognition, semantic role labelling and word sense disambiguation by using techniques like Probabilistic Graphical Models (PGM), including Bayesian Networks, Hidden Markov Models, or Conditional Random Fields.

Different frameworks are available which implement the various techniques. In the case of Pikes, among others the product called Stanford CoreNLP [8], which is an object-oriented framework implemented in Java, is used.

An example of a graph generated by Pikes for the sentence “Lawyers for the survivors have filed a complaint against Sharon in Belgium” is given in **Fig. 7**.

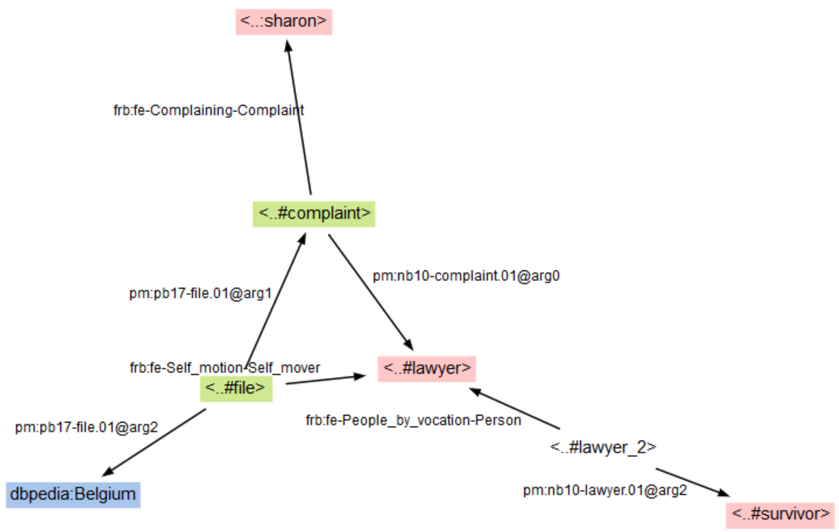


Fig. 7. An example Pikes graph. Source: http://pikes.fbk.eu/

2.3 Automatic Language Translation

This module translates text composed in predefined European languages to English.

Problem Definition

MAGNETO's algorithms are configured for English language, while the information comes in different languages. The lack of an English version of the texts significantly slows down the investigation process.

A typical human translation is an expensive and long process. Nowadays, the amount of information increases constantly, and the use of human translation creates a bottleneck in the effectiveness of organizations. If the text volume ramps up to hundreds or thousands of words, it becomes nearly impossible to use human translation.

The majority of machine translation solutions available on the market is cloud based or have licensing limitations. In both cases, they are impossible for use in MAGNETO since it requires no severe licensing limitations and on-the-premise installation with no external connection to the internet.

Approach and configuration used for MAGNETO

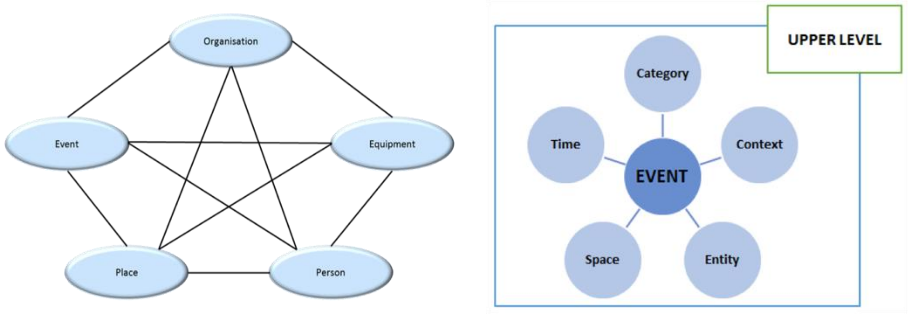
The pursued approach is to train a proprietary engine based on available open-source code technology (e.g. OpenNMT [9]) for statistical machine translations. The idea behind statistical machine translation comes from information theory. A document is translated according to the probability distribution that a string e in the target language (for example, English) is the translation of a string f in the source language (for example, French). Therefore, for the training of the engines large corpora of domain specific translations are needed. In order to ensure a high quality translation, the engines for MAGNETO will be trained with 1,000,000 bilingual segments.

3 MAGNETO Knowledge Base

The knowledge base consists of several databases for storing the raw datasets and an ontology [10], which summarizes the higher-level information, retrieved from the raw datasets. The latter is described in detail in this section.

3.1 The underlying Ontology

In MAGNETO, the knowledge modelling process is crucial to ensure the goal of developing a platform that represents all relevant investigational elements, integrates different and heterogeneous information, performs advanced computational analytics, generates recommendations and inferences, and assists LEAs with their criminal investigation. In order to effectively define knowledge representation, it is necessary to understand what an investigative process is. An investigation is a complex process that involves event, person, circumstance, various types of resources, data extracted from resources, scenarios and relations that link all these elements. The objective of an investigation is understand to who, why, how, by what means, when and where a crime was perpetrated in order to identify each role of all involved persons.



(a) – Pentagram model of main concepts (b) – Event specification of MAGNETO

Fig. 8. MAGNETO ontology conceptualisation for criminal investigation

In [11], the author describes a core ontology based on NATO standards to improve military intelligence analysis. The main concepts of this core ontology have been selected as a basis of the MAGNETO ontology: Organization, Equipment, Person, Place, and Event as depicted in **Fig. 8(a)**. A more detailed outline of the Event category is presented in **Fig. 8(b)**. Besides these concepts, further top-level concepts are required. The general concept PersonAttributes, for example, serves the purpose to describe the characteristics of a suspect or perpetrator. (Sub-)Sub-concepts for describing details are, e.g., FingerPrint, Clothing, or Tattoos. Lower levels of the ontology represent domain-specific concepts, e.g. Crimes, Arson, Thief, Victim, Middleman, Witness, on the one hand, and scenario-specific concepts on the other hand, e.g. HighSecurityLock, ArmouredCar, RealEstateInvestor, HeartAttack.

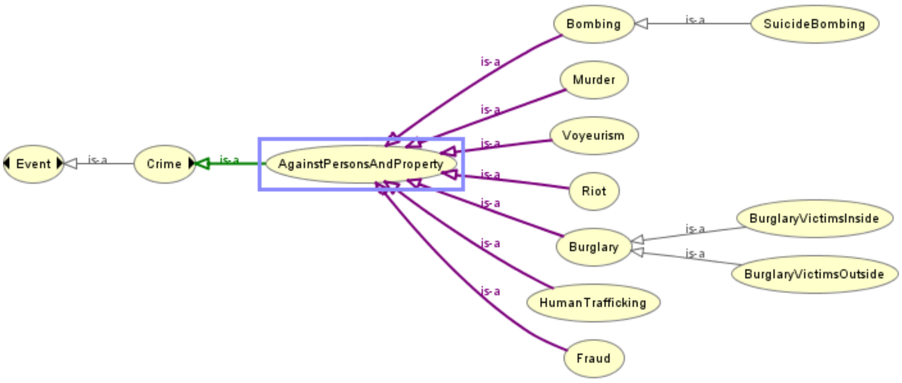


Fig. 9. Event hierarchy of criminal investigation

Exemplary parts of the further elaboration of the top level concept Event are depicted in **Fig. 9**, which is specifically tailored to address broadly the various criminal investigation processes and procedures. Similarly, each concept in the pentagram model is further elaborated. These concepts and their sub-concepts are associated among each other through object relationships. The involved LEAs will evaluate the MAGNETO platform and ontology using real investigation cases. This experience will be used to refine the ontology, which will then be published and submitted for standardization.

3.2 Population of the MAGNETO Ontology

The instantiation of the proposed ontology through metadata extracted from analytics components described in Sect. 2, presents a unique challenge. The technical implementation of the above approach is done as in the following: The instances representing the high-level information in a given investigation are stored in a RDF triple-store provided by the Apache Jena [12] framework. The information extracted from heterogeneous data sources is communicated to the triple-store via the interface implemented by an Apache Fuseki server. Ideally, the entire triple-store is held in the memory to guarantee computationally efficient runtime. However, nowadays real world cases have to process a vast amount of information preventing to keep everything in memory at the same time. A solution to this is the TDB2 storage concept of Apache Jena, which has theoretically no size limitations and allows 100's of millions of triples.

The task at hand in the population of the concepts is to decide which concepts of the ontology need to be instantiated to store the information extracted by the different information extraction modules. As a start, this may be done with lookup tables, which decode which instances are created as a result of incoming information.

In the context of natural language processing, the lookup tables are supplied by databases like DBpedia [13] and SUMO [14]. The use of DBpedia allows refining the type of the main concepts (see Section 3.1) Events are refined based on SUMO database. The found types need to be mapped to concepts modelled in the MAGNETO ontology. The mapping is implemented by determining the semantic similarity. For the similarity, a statistical probability measure based on word embedding [15] is used.

4 Conclusions and future work

In this paper, an outline of various information extraction and processing components is presented to aid the LEAs in their criminal investigation. The output of the information extraction components is used to populate an ontology which has been developed specifically for MAGNETO's purposes. In detail, the MAGNETO ontology concepts are instantiated according to the extracted information and stored by using the Apache Jena framework completed by an Apache Fuseki server for communicating with the knowledge repository. The future work will focus on the usage of the instantiated information, e.g. temporal sequences, to support semantic reasoning of event occurrences leading to perpetrator identification.

References

1. Erfanian Ebadi, S., Guerra Ones, V., Izquierdo, E., "Efficient background subtraction with low-rank and sparse matrix decomposition", in Image Processing (ICIP), 2015 IEEE International Conference on, 2015.
2. Boutsidis, C., Mahoney, M.W., Drineas, P., "An Improved Approximation Algorithm for the Column Subset Selection Problem", Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, 2009.
3. Candes, E.J., Li, X., Ma, Y., Wright, J., "Robust principal component analysis?" J. ACM, vol. 58, no. 3, pp. 11:1–11:37, 2011.
4. Mitrea, C.A., Piatrik, T., Ionescu, B., Neville, M., "Retrieval of Distinctive Regions of Interest from Video Surveillance Footage: A Real Use Case Study", 6th International Conference on Imaging for Crime Prevention and Detection (ICDP-15), 2015
5. Viola, P., Jones, M., "Rapid object detection using a boosted cascade of simple features", Computer Vision and Pattern Recognition, 2001.
6. CMUSphinx Homepage, https://cmusphinx.github.io/wiki/, last accessed 2019/01/31.
7. Corcoglioniti, F., Rospocher, M., Palmero Aprosio, A., "A 2-phase frame-based knowledge extraction framework", In: Proc. of ACM Symposium on Applied Computing, 2016.
8. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D., "The Stanford CoreNLP Natural Language Processing Toolkit", In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60. 2014.
9. OpenNMT Homepage, http://opennmt.net/, last accessed 2019/02/04.
10. Borst, W., "Construction of Engineering Ontologies", PhD thesis, Institute for Telematica and Information Technology, University of Twente, Enschede, The Netherlands, 1997.
11. Dragos, V., "Developing a core ontology to improve military intelligence analysis", International Journal of Knowledge-Based and Intelligent Engineering Systems, IOS Press, 2013.
12. Apache Jena Homepage, https://jena.apache.org/, last accessed 2019/02/06.
13. Mendes, P.N., Jakob, M., Bizer, C., "DBpedia: A Multilingual Cross-domain Knowledge Base", International Conference on Language Resources and Evaluation, 2012
14. Pease, A., Niles, I., Li, J., "The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications", AAAI Technical Report WS-02-11, 2002.
15. Mikolov, T., Chen, K., Corrado, G., Dean, J., "Efficient estimation of word representations in vector space," in Proceedings of Workshop at ICLR, 2013.