



D4.2 Event driven probabilistic model (Publishable Version)

Report: Event driven probabilistic model

Project acronym:	RED-Alert
Project full title:	Real-time Early Detection and Alert System for Online Terrorist Content based on Natural Language Processing, Social Network Analysis, Artificial Intelligence and Complex Event Processing
Grant agreement no.:	740688
Responsible:	ICE
Contributors:	MAV, ELTE, INSKT, BCU, CITY, SPP, MOPS-INP, SO15, GUCI, SPPS
Document Reference:	D4.2
Dissemination Level:	CO
Version:	FINAL (PUBLISHABLE VERSION)
Date:	10.05.2018

1. Executive Summary

Task 4.2 “Develop an event driven inference mechanism to cope with uncertainty” deals with the development of this mechanism able to propagate uncertainty from the input events (social media) to output events (alerts). This inference mechanism will consider all types of uncertainty (event content, event occurrence, and event rules).

To understand the task in the context of the Project, as the objective of the Project is to create real-time alerts of suspicious content on social networks, one of the main features is to analysis text content. The figure below is the schema of the whole process:

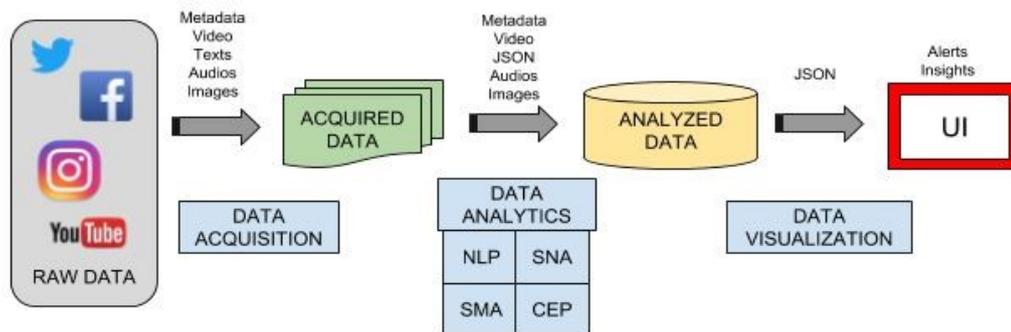


Figure 1. Data transformation from sources to alerts

The main steps of the data flow are:

1. **Data Acquisition:** Data is directly extracted from social networks or imported by the user Law Enforcement Agencies (LEAs) in a pre-defined format.
2. **Data Analytics:** Data collected is analysed using several methods, such as:
 - a. Texts of messages using Natural Language Processing (NLP) technologies (e.g. extraction of concepts, sentiment, topics, etc.) of the content.
 - b. Relations between users through Social Network Analysis (SNA) to obtain communities, the most influential users, etc.
 - c. The videos, audios and images (Semantic Multimedia Analysis, SMA) by means of analysis technologies for these types of contents with the aim to recognize symbols, images, actions, etc. classified as suspicious and convert audio to text.
 - d. Alerts, which are created by means of Complex Event Processing (CEP) technology. A set of patterns (rules) will be defined to identify suspicious messages, from the point of view of the content or the author.
3. **Data Visualization:** Data analysed is amalgamated into a visual UI for the LEAs to get insights and leads to the corresponding actions.

The result of Task 4.2 is a component part of the Data Analytics core. In this core, the CEP is based on the generation of the Event probabilistic model (T4.2) and the generation of the alerts and alarms (T4.3). Whilst the result of T4.3 will be the generation of the alerts themselves, the T4.2 results will be the generation of the probabilistic models for such alerts generation.

The probabilistic models are based on the conditions and needs from the LEAs forming patterns and rules to trigger alarms. These models will consider the uncertainty in the input events, combine these probabilities within the different event patterns and derive a probabilistic output event. The uncertainty will be propagated through the event processing network (EPN), therefore more abstract/complex events will be derived as well.

This task is supported by a software prototype that has been developed to validate the probabilistic model by the LEAs. This prototype, though, is based on the results of the previous components as the CEP is the last component on the Data Analytics chain (see Figure 1). Thus, it has to be noted that the scope of this prototype is limited due to the lack of processed data that has arrived to the CEP.

This document, therefore, provides a short report on the T4.2 software prototype. As stated in the Description of Action (DoA), this deliverable is a prototype (software) deliverable. The software developed is accessible through the instructions laid out in the body of the document.

2. Conclusions

The purpose of this document is to draw the grounds, development and execution of the event probabilistic model with the goal to pave the way for triggering alerts and alarms of suspicious activities under the context of RED-Alert.

The event probabilistic model has been implemented after the analysis of the responses offered by the LEAs to the “Patterns and Suspicious Activities” questionnaire. Since the LEAs want to track different kind of social data, the model has been developed under a high configurability model. In the questionnaire, the LEAs were asked to respond about what kind of information they would like to track and what metadata, in the form of e.g. location, classification of threats, of suspects, etc. so they would be able to configure the model by themselves.

To show the results of the probabilistic model a software prototype has been developed. This can be deployed and tested by following the steps laid out in this report. However, since the CEP input is factually the output of the previous Data Analytics components, i.e. SNA, NLP, SNA and Anonymisation Tool, not many conclusions can be extracted due to the lack of processed data. In any case, LEAs feedback will be collected and documented together with the strategy adopted in order to answer to their feedback. The resulting report will be integrated as part of the next WP4 deliverable, D4.3 - Probabilistic model implementation, and its different iterations.