REDALERT

D4.3 Probabilistic Model Implementation (Publishable Version)

Report: Probabilistic Model Implementation

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:	**NA**
Document Reference:	**D4.3**
Dissemination Level:	**CO**
Version:	**V1.0 (PUBLISHABLE VERSION)**
Date:	**27.08.2018**

1. Executive Summary

Task 4.3 “Implement the inference mechanism” deals with the implementation of the inference mechanism into the CEP engine. Specific mini-CEP applications are being implemented for the selected scenarios/use cases. Event patterns will be developed by two methods: 1) by domain experts and 2) by machine learning techniques. This implies that domain experts will provide the definition of event patterns and machine learning methods will be developed in order to discover these patterns. The domain experts will be able to define their event patterns directly at the configuration screen of the CEP prior to deploy their “own mini-CEP” while the ML methods will improve the already defined event patterns.

The implemented mini-CEPs will be able to query past events and to efficiently handle these querying results, so it’s able to compare current and historical states and to reason over time and space (two current limitations of existing semantic CEP tools).

To understand the task in the context of RED-Alert project, as its objective is to create real-time alerts of suspicious content on social networks, one of the main features is to trigger alarms based on the content of events. The figure below is the schema of the whole process:

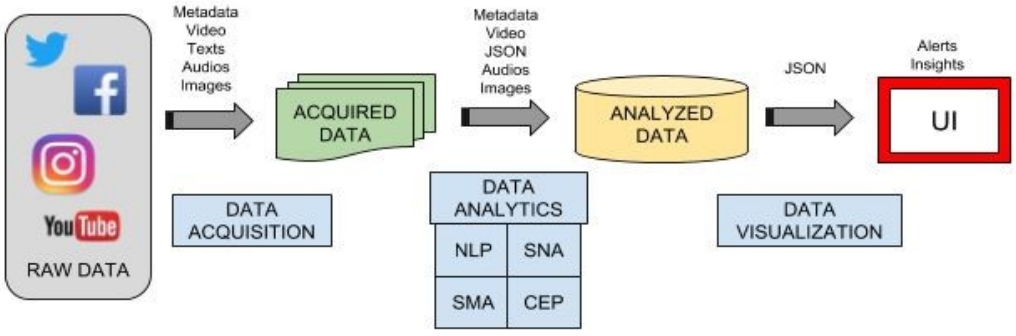

Figure 1: Data transformation from sources to alerts

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 4.3 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 T4.2 results will be the generation of the probabilistic models for such alerts generation, the result of T4.3 will be the generation of the alerts themselves.

The inference mechanism will be executed on top of the user-defined patterns and rules to effectively trigger alarms. The mechanism will be observing the input data and continuously comparing it to the user-defined patterns and rules and, in combination with the probabilistic and uncertainty model from T4.2, content will be identified and communicated to the LEAs in the form of alarms so they can take appropriate action.

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 T4.3 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 deliverable is to set up the environment where the event probabilistic model from the previous task (T4.2) would be executed and suspicious activities would be detected to trigger alarms, being these communicated to the final user of the RED-Alert tool, i.e. the LEAs.

The CEP is, in a nutshell, based on the following modules:

- Messaging component: a component used for feeding the different mini-CEP apps to be deployed depending on the particular demands of each scenario
- Configuration WebApp: a web-based form developed to configure the mini-CEP apps according to the needs of each LEA
- Alerts Dashboard: a web-based dashboard specially designed to help the end user in identifying the different items investigated and where alarms could be configured for their trigger.