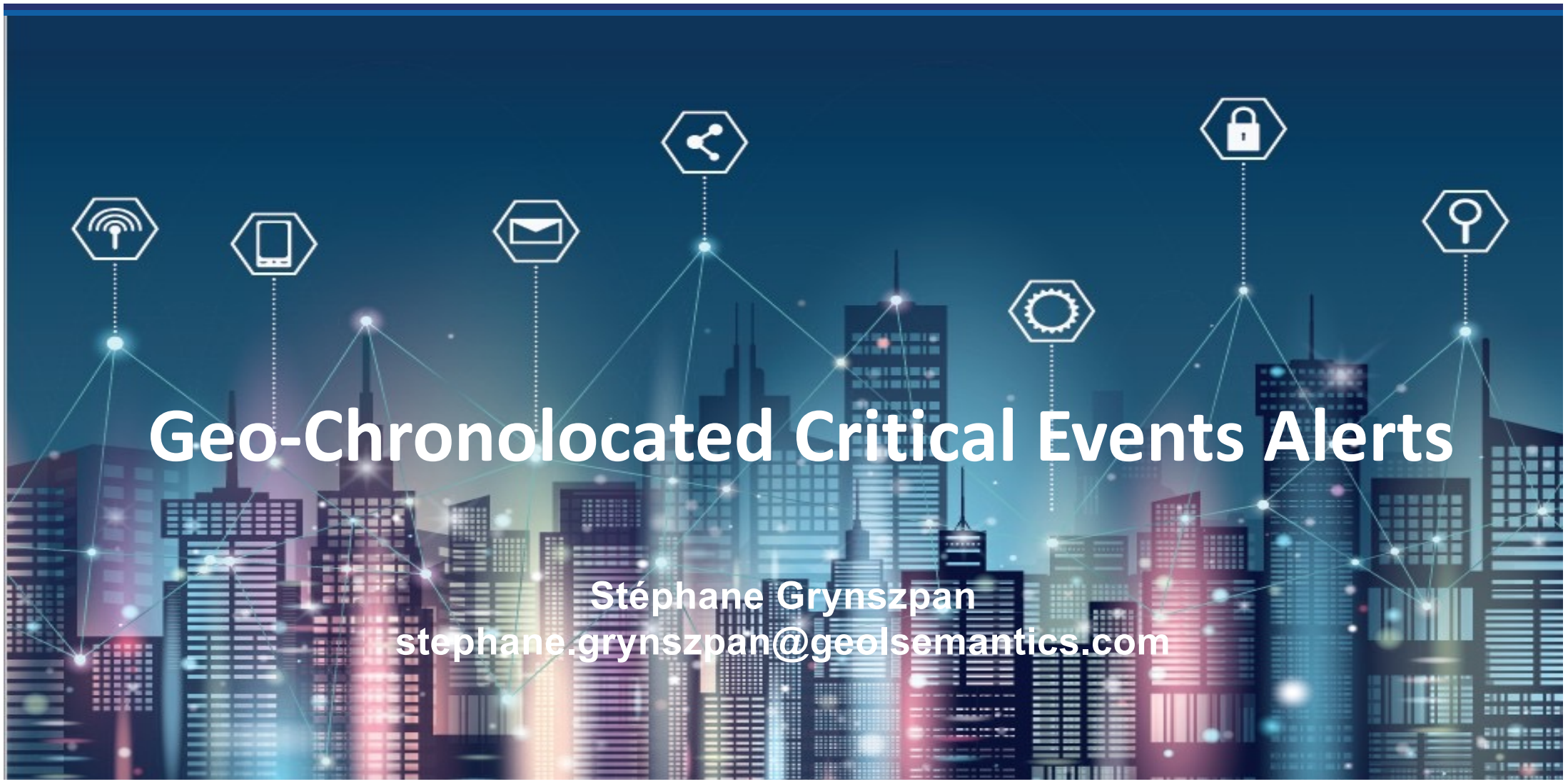


Geo-Chronolocated Critical Events Alerts

Stéphane Grynszpan
stephane.grynszpan@geolsemantics.com

GEOL Semantics INTELLIGENCE

RESEARCH CONTEXT

SafeCity

- French national security collaborative research & development project during 3 years funded by the BPI (public bank of investment)
- Enhance the security of smart cities with innovating products & services for road & school safety, connected patrol, command posts and communication
- 2 pilot sites: Nice (French Riviera) and Paris-La Défense (business district area)
- GEOLSemantics role in this project is to analyze social networks to generate alerts in case of events requesting first responders action
- As a semantics technological solutions publisher, GEOLSemantics offers various tools to understand text including social media messages and geo-chronolocated events

GEOL Semantics Intelligence

HYPOTHESIS FROM SCRATCH

- Observation of social networks to detect present or future incidents earlier than other types of alerts

- Detection of relevant incidents such as accidents, traffic jams, fires, floods, falling trees, landslides, forgotten luggages, illegal garbage dumping, acts of violence, spontaneous parties, fight between groups

- Selection of Twitter as the most appropriate platform to communicate breaking news, spotting events

342 million monthly active users
350 million tweets a day
500 million tweets are sent each day
6th-ranked mobile application
Image captions / Photo-sharing / Streaming video / Geo-location
Easy categorization hashtags, retweets and replies
Demographic Age : 63 % of Twitter users are between 35-65 years old (mature audience)

GEOL Semantics Intelligence

EVENTS GEO-CHRONOLOCATION

Extraction of: Event type, town, precise location with its GPS coordinates and date/time

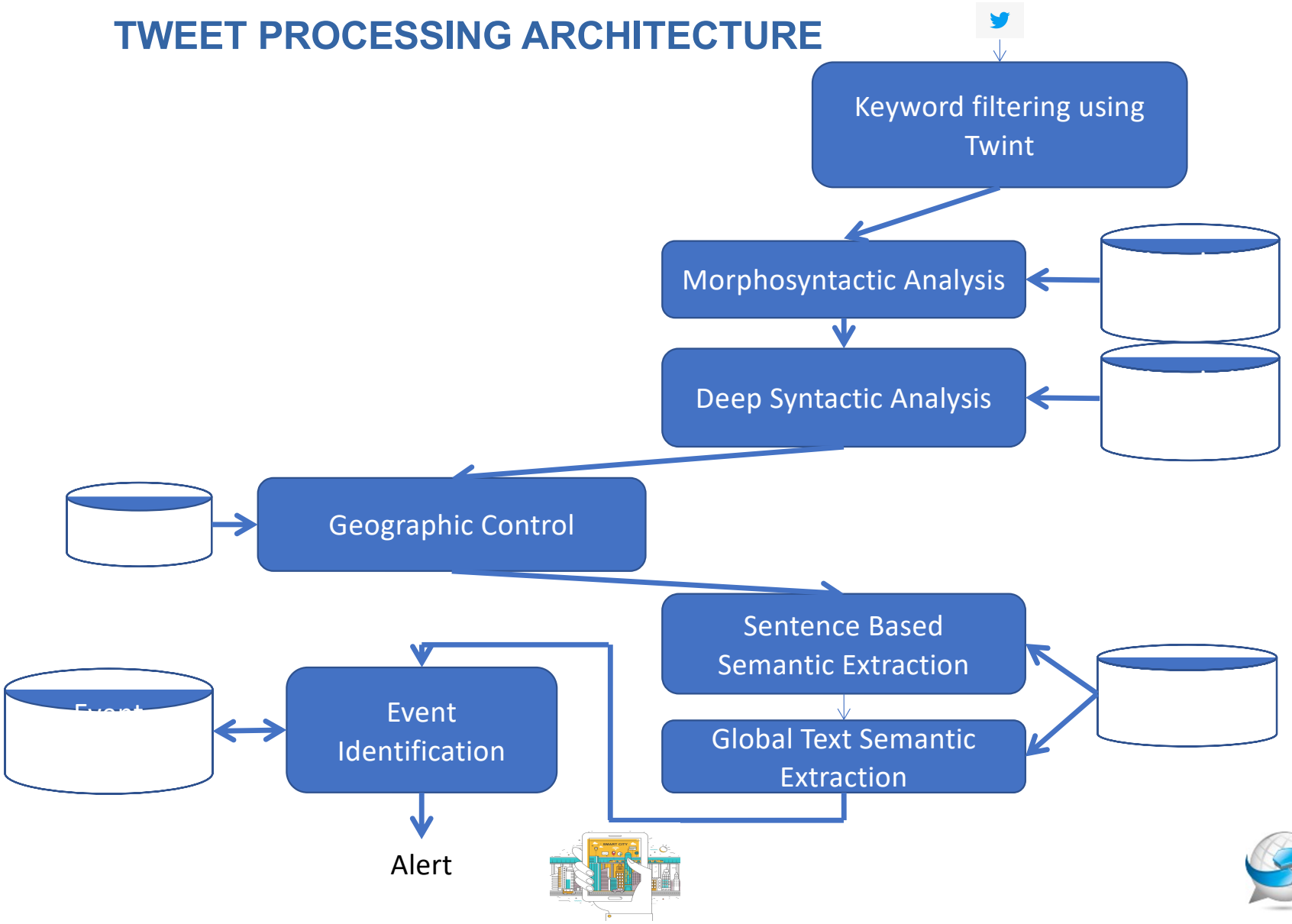
Example: Tweet sent on July 23rd, 2019 at 4:04 pm

Le prochain Nissapero sera donc demain soir au Bateleur dans le @VieuxNice
(Next Nissapero meeting will be tomorrow evening at the « Bateleur » restaurant in the old city of Nice)

Event	Party Organization
Town	Nice
Location	Le Bateleur (restaurant)
Date/Time	July 24, 2019 17:00 – 23:59

GEOL Semantics Intelligence

TWEET PROCESSING ARCHITECTURE



GEOL Semantics Intelligence

TWEETS LANGUAGE DIFFICULTIES

- **Tweets language is different as users are writing as they speak and according to social media habits**
- **The limited number of characters motivate users to summarize the events clearly**
- **Text is often associated with images, video or URL containing additional information**
- **A common spelling mistake is the substitution of a word by another one with the same pronunciation**

Original spelling; Ont pourra trouvé a Paris des monuments rénover

Corrected spelling: On pourra trouver à Paris des monuments rénovés.

ont (verb to have) ≠ on (one pronoun)

a (verb to have) ≠ à (« in » preposition)

trouvé (found past participle of trouver) ≠ trouver (to find infinitive form of trouver)

GEOL Semantics Intelligence

TWEETS LANGUAGE DIFFICULTIES

Phonetic Spelling:

JPP abbreviation « J'en Peux Plus » *(I can't bare)*

V1 phonetic spelling « vient » *(about to do)*

poubel phonetic spelling « poubelle » *(garbage)*

Translation:

I can't bare, a girl has just set fire in a garbage can and runned away discreetly

Correct spelling:

GEOL Semantics Intelligence

TWEETS LANGUAGE DIFFICULTIES

Syntactic Problems : Limitation in terms of characters number push users :

- **to erase words which are not necessary for the comprehension such as article, preposition, obliging to have a more semantic approach of the analysis and to adapt disambiguation rules;**

- **to concatenate words without spaces which forces to split them into several tokens.**

Hash tags and arobase tags must be also splitted into parts to be integrated into the natural language understanding ex: #VilledeParis → Ville de Paris (City of Paris)

Generally only a part of Twitter messages is written in natural language. A first part is a list of arobase or hash tags and often the name of a city. A second part contains the text in a kind of natural language, and a third part with a list of sources (arobase tags), hashtags and/or URLs at the end of tweets.

@MathieuNobles Trignac. Deux morts dans l'incendie d'un appartement https://t.co/TO5qMU5obv

Trignac (town) 2 deads in an apparti

GEOL Semantics Intelligence

GEO-CHRONOLOCATED EVENTS PROBLEMS

Geolocation problems

- Twitter geolocation was previously easy with the use of Twitter user GPS position, but only 5 % of French users have given their consent.

- Natural language processing is used to extract locations from message text and relations between them

- GIS is used to solve some ambiguities for dictionary unknown location, in order to get GPS positions of each place. Then, we combine them to get the most precise one.

Determining Current and Ongoing Events

- The purpose of the system is to alert authorities if the related event is current or future

- The analysis has to take into consideration not only verbs but also adverbs (suddenly, currently, tomorrow, …), use of conditional tense, hypothetical words (whether), ...

GEOL Semantics Intelligence

GENERAL APPROACH

- **This approach is rule-based**

Dictionary Search

- **Systematic Words Confusion Spelling : « a » and « à » will both have preposition and verb interpretations**
- **Dictionary entries used for re-spelling are phonetic representations of each word linked to their correct spellings**

For tokens which are not in the dictionary:

- **Token with hyphen are splitted ex: brûle-il → brûle** burns (verb) **il** it (personal pronoun)

- **A splitting of unknown token into valid tokens is performed with validation rules**

ex: jviens → j viens → je viens (I come)

And with hash or arobase codes

ex: #SeineMaritime → Seine Maritime (a West region of France)

- **Remaining unknown token get a default grammatical tag value like for numbers or uppercase first character tokens**

ex: 150 (number) **Chomsky (proper name)**

GEOL Semantics Intelligence

TAGGING

- Tagging is processed by machine learning based on semi-automatically tagged text. A tweet corpus has been added for tweets processing.
- Rules are applied in a Boolean manner
- In case of remaining ambiguities, rules using broader context are applied. To resolve some semantic ambiguities, the system uses particular rules based on specific words

ex : sommes is a present tense 1st plural form (nous sommes), lemma to be
sommes is a present tense 2nd singular form (tu sommes), lemma to add up
Other sommes is a plural noun, lemma sum

GEOL Semantics Intelligence

SYNTACTIC ANALYSIS AND ENTITIES RECOGNITION

- **Named entities recognition and categorization are performed from a syntactic perspective by rules**

Saint-Jean-de-Luz **Named Entity, Location, Town**

Société nationale des chemins de fer français **Named Entity, Company**

- **Dependencies inside noun phrases and relationships between noun and verb phrases are also issued from rules.**

- **This phase concerns also relationships between pronouns and their referents.**
- **It analyzes negative structures to determine what is denied. The other modalities are also taken into consideration (doubts, obligations, permission).**

GEOL Semantics Intelligence

GEOGRAPHICAL INFORMATION SYSTEM

GIS has several roles :

- **Confirm or even correct locations found by the language processing.**
- **Removing ambiguities : the wider the area to be monitored, the greater the risk that ambiguities will arise**

Concerning locations known from a few people, Twitter users give usually additional information like the region or a well known place nearby
ex: Paris Texas , Gentilly near Paris

Location can be given as a place « between » two positions or at the intersection of two ways. The geographical information system is able to compute these GPS positions.

GEOL Semantics Intelligence

SEMANTIC EXTRACTION

- Semantic extraction is carried out in two phases in our solution

  - A first phase analyzes text sentence by sentence to extract events

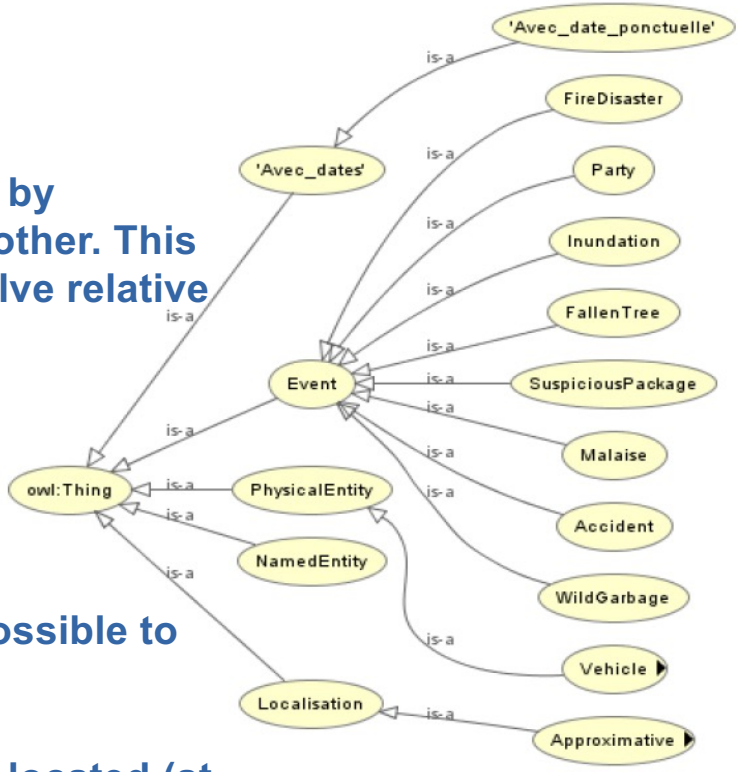
  - A second step collects information between sentences in particular by identifying entities who are named differently from one position to another. This phase takes also into account metadata like issue date and time to solve relative dates in the text

ex : D-3 for the next CEIS show

- Events set to be extracted is defined in an ontology

- Semantic extraction is also carried out here by rules that make it possible to relate texts to ontology concepts

- If an event is recognized as ongoing or future and if it is accurately located (at least the city) then an alert is triggered



GEOL Semantics Intelligence

INFORMATION ENRICHMENT RELATED TO THE SAME EVENT

- Our approach triggers alert from the first message. However, not all important information is necessarily in the first tweet
- It is therefore necessary to identify that a new tweet corresponds to the same event already described in previous tweets. This is primarily due to the geographical and temporal proximity and then to the events types compatibility
- Furthermore, ontological information for an event found in a tweet is stored on an Elasticsearch base. It is then completed thanks to tweets about the same event

For example, during the attack in the Strasbourg Christmas market (December 11th, 2018), one tweet allowed the terrorist to be located in a specific place at a given time and another one described his clothing. Both information would have been very useful if it would have been received by the police immediately.

GEOL Semantics Intelligence

EVALUATION

Evaluation has been carried out on Twitter observation in French language during 4 days from Saturday 00:00 to Tuesday 11:59 pm

Triggered Alerts Analysis (Precision)
Number of alerts: 347

Fire	37 %
Accident	40 %
Flooding	1,9 %
Violence Act	4,3 %
False Type	5,3 %

False Alerts	10,6 %
Partial Accurate Locations	4,7 %
Full Recognition of Locations	84,7 %
Alerts Sent to the Right Town	89,4 %

Analysis of good events that have not triggered an alert (Recall)

Number of Received Messages	8919
Number of Alerts not Produced in a Day	25

Precision	**89,4 %**
Recall	**79,1 %**

GEOL Semantics Intelligence

CONCLUSION

- Even if French Twitter users would rather use official channels to warn rescue services (US 911, France 18), the system remains very useful to enrich events data or detect events which does not request rescue intervention such as unauthorized garbage dump

- For disasters or terrorist events where tweets number is much more important, rescue services are interested to get information on the situation from different points of view. Generally, tweets are associated with images or video that can give usefull information to organise the response

- Detection of future open organized events are usefull for rescue services to organize the event security

- Geo-chronolocated events detection has many other applications which are not restricted to future and ongoing events such as debriefing of disasters for resilience improvement, suicidal intentions detection, dialogs analysis between jihadist influencers and influencees on social networks, preparing military actions using past activities in the same place, or for police investigation reports management

GEOL Semantics Intelligence