MondeΦIA

Détection d’événements Euro 2016

Team

- ➔ Ghislain (Mondeca)
- ➔ Hacene (Mondeca)
- ➔ Anh (Mondeca)
- ➔ Arnaud (Mondeca)
- ➔ Julien (Eurecom)
- ➔ Edouard (INRIA)

Our Approach

- ➔ Temporal Analysis
- ➔ Topic Detection relevant to the Euro 2016
- ➔ Semantic Event Representation
- ➔ Semantic Annotation and Rule-based analysis
- ➔ Indexation and visualization

Workflow Process

Filtrage des Tweets

Serialisation flux

Extraction
Détection

Content
Annotation

Détection des joueurs, actions du match

Pose de metadatas

Indexation

Sortie Annotation et des tags détectés

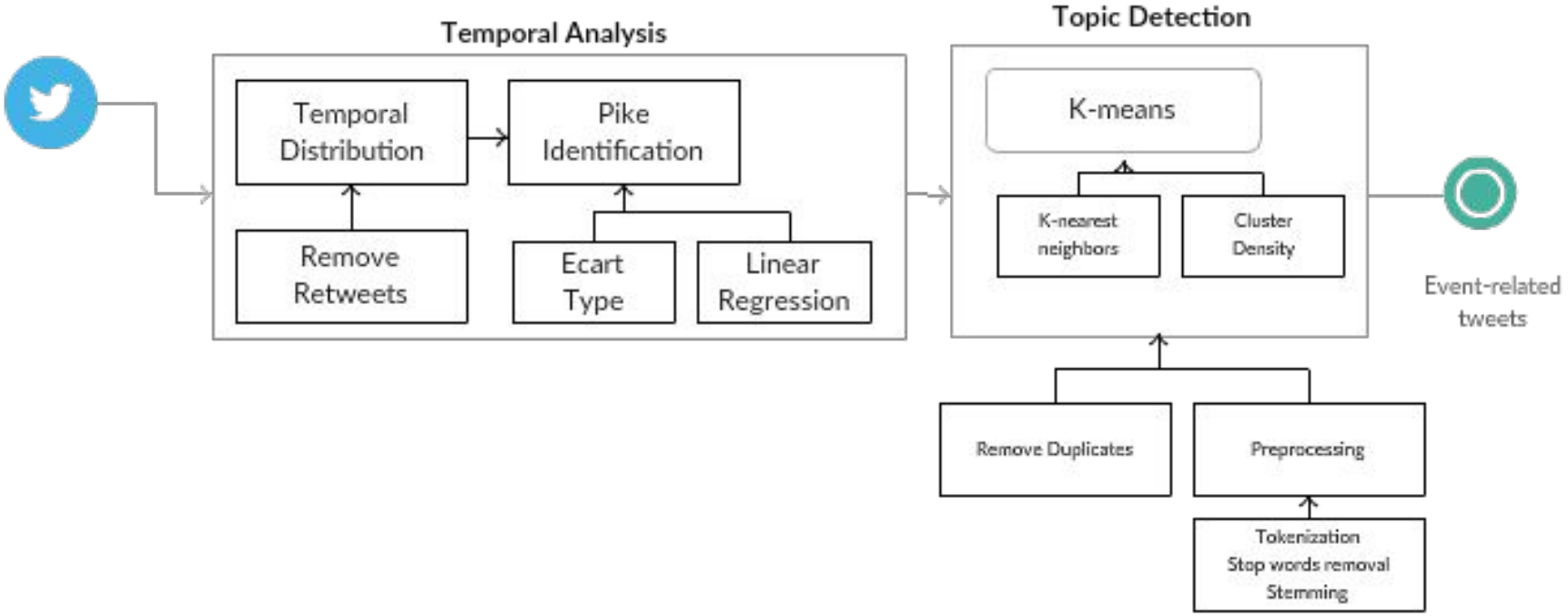
elasticsearch.

Visualisation

Ontologie des évènements et référentiels

Alimente les composants

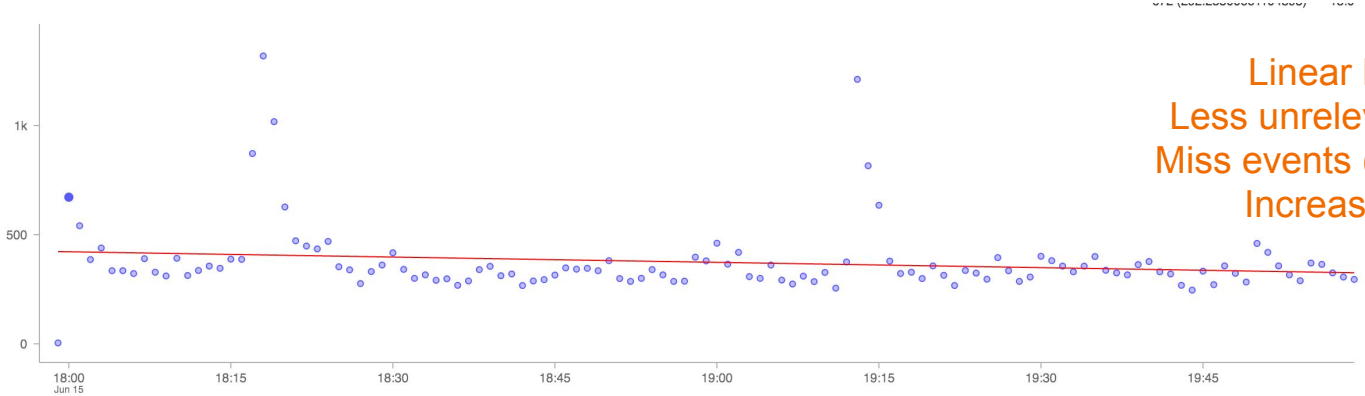
Event Filtering



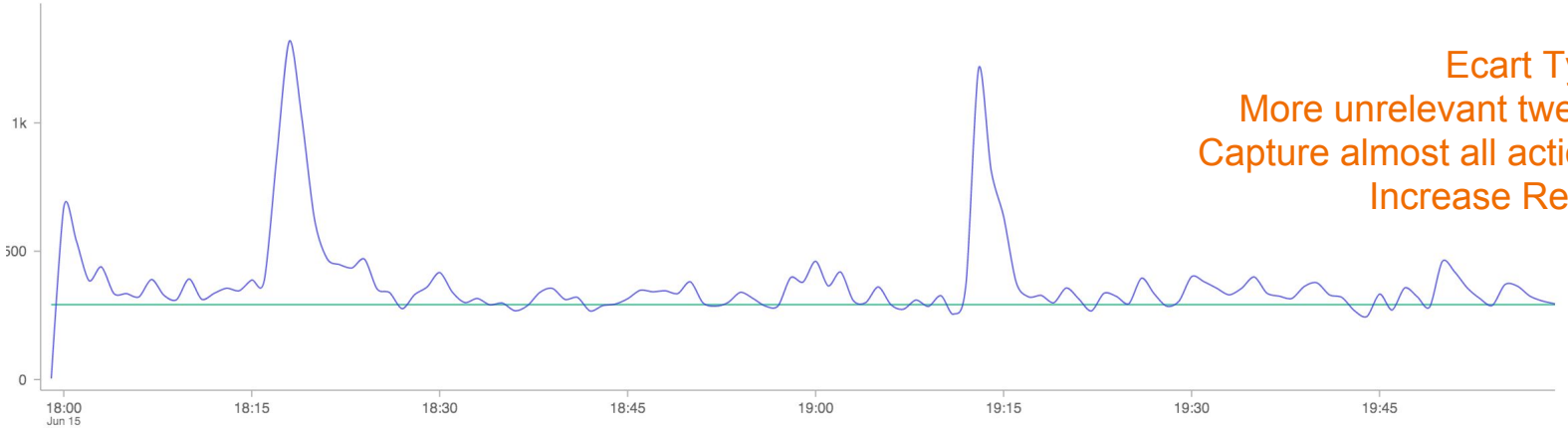
Temporal Analysis and Pike Detection

- ★ Retrospective analysis of a collection of tweets
- ★ Temporal distribution of tweets based on a fixed time-window
- ★ Ignore time-window with low frequency
  - ○ Ecart Type
  - ○ Linear Regression

Pikes Detection (Romania VS Switzerland)



Linear Regression
Less unrelevant tweets
Miss events (e.g. shoot)
Increase Precision



Ecart Type
More unrelevant tweets
Capture almost all actions
Increase Recall

Topic Detection

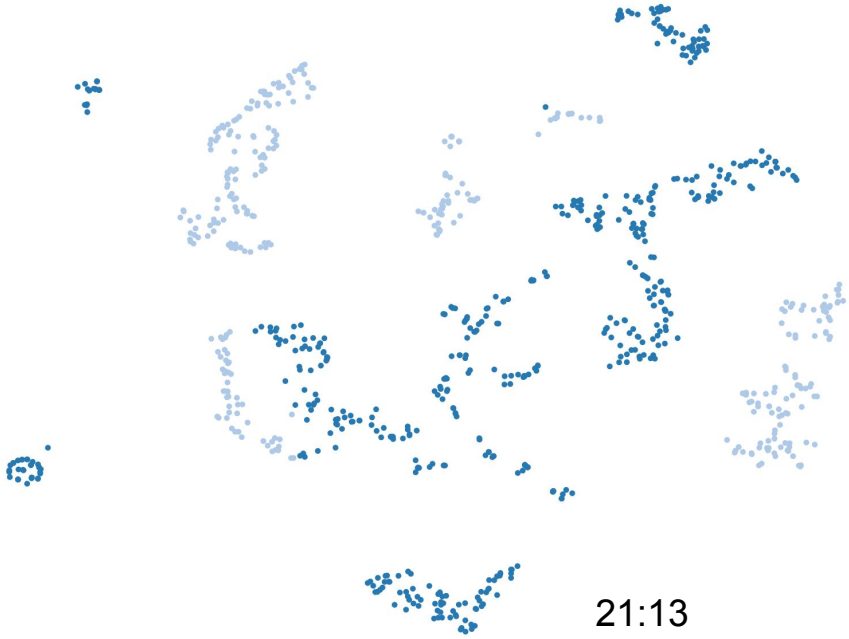
Our goal is to detect topic in relevant time-windows

➔ We use K-means to create tweets clusters in each spike

- **Common Preprocessing tasks** (Stop words removal - Near duplicate removal - Stemming)

➔ Identify cluster with highest density

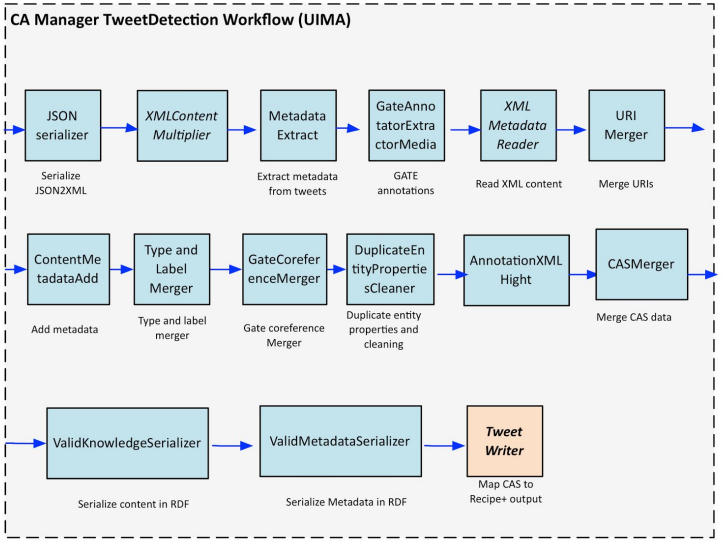
Romania Vs. Switzerland Goal at 21:15



21:13

21:15

Semantic Annotation



DEMO

Useful Resources & Tools

- ➔ Shared Domain ontology: http://goo.gl/XBfQHc (in fr/en; some in ar)
- ➔ Shared External datasets in RDF: https://github.com/eamosse/mondephia/tree/master/rdf (mainly fr/en; some resources in ar)
- ➔ Shared Filtered datasets in JSON: https://www.dropbox.com/sh/c1kmfht4v39nwc5/AAD5LDO8u47u0mQMFrshvINga?dl=0
- ➔ More on our Github: https://github.com/eamosse/mondephia/
- ➔ Python: NLTK, Numpy, SK-Learn
- ➔ Mondeca : CAM, ITM

Conclusion & Perspectives

An approach for detecting events on Twitter

- Natural Language Processing Analysis with Machine Learning
- Semantic Enrichment with Rule-Based Systems

We plan to evaluate this approach using the Hackatal 2016 dataset as ground truth

Submit a research paper

Thank You

Questions?