



## Tâche de détection d'évènements

HackaTAL

Djamel MOSTEFA & Nadi TOMEH

[djamel.mostefa@systrangroup.com](mailto:djamel.mostefa@systrangroup.com)

[nadi.tomeh@lipn.univ-paris13.fr](mailto:nadi.tomeh@lipn.univ-paris13.fr)

2 juillet 2016

- ⇒ Détection d'évènements dans des matchs de l'euro 2016 à partir de Twitter
- ⇒ Objectif applicatif : créer automatiquement des résumés de matchs ou des « files live » de commentaires à partir de Twitter
- ⇒ Données collectées sur Twitter en arabe, anglais et français
- ⇒ Type d'évènements à détecter:
  - Début et fin de périodes
  - Buts
  - Cartons
  - Penaltys
  - Tirs au but
- ⇒ évènement = (heure absolue, type d'évènement, information complémentaire)

LABEL	TYPE EVENEMENT	INFORMATION COMPLÉMENTAIRE (ÉVENTUELLE EN FONCTION DU TYPE D'ÉVÉNEMENT)
BUT	BUT	Nom du buteur
CJA	CARTON JAUNE	Nom du joueur sanctionné
CRO	CARTON ROUGE	Nom du joueur sanctionné
CGT	CHANGEMENT	Nom du joueur sortant ; Nom du joueur entrant
D1P	DEBUT 1ERE PERIODE	N/A
D2P	DEBUT 2NDE PERIODE	N/A
F1P	FIN 1ERE PERIODE	N/A
F2P	FIN 2NDE PERIODE	N/A
PEN	PENALTY	Nom du pays pour lequel le penalty est sifflé
TIR	TIR AU BUT	Nom du tireur

- Données collectées en utilisant l'API streaming de Twitter
- Format JSON natif fourni par l'API
- Annotations manuelles produites en se référant aux « fils live » de commentaires des matchs de sites sportifs
- Annotations au format texte tabulé

• HEURE\_ABSOLUE TYPE\_EVENEMENT INFORMATION\_COMPLEMENTAIRE

21:01:00	D1P	
21:02:00	TIR	Matuidi
21:04:00	TIR	Pintilii
21:05:00	TIR	Stancu
21:11:00	TIR	Griezmann
21:12:00	TIR	Giroud
21:15:00	TIR	Griezmann
21:28:00	TIR	Pogba
21:33:00	CJA	Chiriches
21:37:00	TIR	Griezmann
21:45:00	TIR	Payet
21:46:00	CJA	Rat
21:48:00	TIR	Giroud
21:48:00	F1P	
22:04:00	D2P	
22:07:00	TIR	Stancu
22:12:00	TIR	Giroud
22:16:00	TIR	Pogba
22:17:00	BUT	Giroud
22:19:00	TIR	Andone
22:20:00	CGT	Andone Alibec
22:23:00	PEN	Roumanie
22:24:00	BUT	Stancu
22:25:00	CGT	Griezmann ; Coman
22:28:00	CJA	Giroud
22:31:00	CGT	Stanciu ; Chipciu
22:36:00	CGT	Pogba ; Martial
22:37:00	CJA	Popa
22:39:00	TIR	Hoban
22:40:00	TIR	Martial
22:41:00	CGT	Popa ; Torje
22:48:00	BUT	Payet
22:53:00	F2P	

## ⇒ Données de développement

- 24 matchs poule de la 1ere et 2eme journée

## ⇒ Données évaluation

- 12 matchs poule de la 3eme journée

## ⇒ Nomenclature des données:

- <PAYS1>\_<PAYS2>\_<YYYY>\_<MM>\_<DD>\_<HH>h\_<lang>.{json|html|tsv}

## ⇒ Pour un match:

- Données collectées sur Twitter en ar/en/fr au format JSON (1 JSON/langue)
- Les annotations des évènements du match au format TSV
- Le fichier HTML du site sportif qui a servi aux annotations

```
-- Groupe_A
|-- ar
|   |-- Albanie_Suisse_2016-06-11_15h_ar.json
|   |-- France_Albanie_2016-06-15_21h_ar.json
|   |-- France_Roumanie_2016-06-10_21h_ar.json
|   |-- Roumanie_Suisse_2016-06-15_18h_ar.json
|-- en
|   |-- Albanie_Suisse_2016-06-11_15h_en.json
|   |-- France_Albanie_2016-06-15_21h_en.json
|   |-- France_Roumanie_2016-06-10_21h_en.json
|   |-- Roumanie_Suisse_2016-06-15_18h_en.json
|-- fr
|   |-- Albanie_Suisse_2016-06-11_15h_fr.html
|   |-- Albanie_Suisse_2016-06-11_15h_fr.json
|   |-- Albanie_Suisse_2016-06-11_15h_fr.tsv
|   |-- France_Albanie_2016-06-15_21h_fr.html
|   |-- France_Albanie_2016-06-15_21h_fr.json
|   |-- France_Albanie_2016-06-15_21h_fr.tsv
|   |-- France_Roumanie_2016-06-10_21h_fr.html
|   |-- France_Roumanie_2016-06-10_21h_fr.json
|   |-- France_Roumanie_2016-06-10_21h_fr.tsv
|   |-- Roumanie_Suisse_2016-06-15_18h_fr.html
|   |-- Roumanie_Suisse_2016-06-15_18h_fr.json
|   |-- Roumanie_Suisse_2016-06-15_18h_fr.tsv
```

### ⇒ Soumission d'un ou plusieurs système par équipe

- <NOM\_EQUIPE>\_<ID\_SYSTEME>.tgz
  - ou ID\_SYSTEM = SYS1 | SYS2 | ...
  - NOM\_EQUIPE = TEAM\_1 | TEAM\_2 | ...

### ⇒ L'archive doit contenir les 12 fichiers hypothèses au format TSV correspondant aux 12 match avec la nomenclature suivante:

- <PAYS1>\_<PAYS2>\_<YYYY>\_<MM>\_<DD>\_<HH>h\_<ID\_SYSTEM>\_<NOM\_EQUIPE>.tsv

### ⇒ Soumissions par email à [djamel.mostefa@systrangroup.com](mailto:djamel.mostefa@systrangroup.com) et à [nadi.tomeh@lipn.univ-paris13.fr](mailto:nadi.tomeh@lipn.univ-paris13.fr)

- 1 email par soumission
- Objet : [HackaTAL] <NOM\_EQUIPE>\_<ID\_SYSTEME>.tgz



SYSTRAN

## Evaluation

### → Evaluation

- Calcul de la précision, du rappel et la F-mesure (Script d'évaluation fourni) en 3 modes:
  - Mode strict : match exact sur les triplets avec 4 min de tolérance (2 min avant / 2 min après) pour les temps
  - Mode intermédiaire dans lequel on ne considère que le type d'évènement et les annotations
  - Mode relâché dans lequel on ne considère que le type d'évènement
- Script d'évaluation disponible sur le github



**SYSTRAN**

Questions?

⇒ **Bonnes détectations et Bon hackaTAL !**