# M2 SID Challenge 2016

# Inspiré de la campagne



# Etiquetage de vidéos en genre

Isabelle FERRANÉ: isabelle.ferrane@irit.fr

Jérôme FARINAS : jerome.farinas@irit.fr

Julien PINQUIER: julien.pinquier@irit.fr

Lynda TAMINE-LECHANI : lynda.lechani@irit.fr

José MORÉNO: jose.moreno@irit.fr





#### Tâche issue de MediaEval

- ♦ MediaEval: campagne d'évaluation <a href="http://www.multimediaeval.org/">http://www.multimediaeval.org/</a>
  - Objectif : explorer de nouvelles techniques sur un corpus semi-professionnel
  - Contenu brut (audiovisuel)
     Web TV, vidéos personnelles,...
     5288 fichiers (> 2000h)
  - Contenu complémentaire
     Métadonnées (description succincte du contenu)
     Transcription de la parole (ASR LIMSI)
     Images clés (détection de plan)
     Données sociales (tweeter)
  - Genres : 25 + 1 « poubelle »



# ♦ Metadonnées (fichier xml)



Video: Crashsolo-OneMinuteRumpoleAndTheAngelOfDeath659.flv

Tag: 1012 literature

Etiquette GENRE issue de blip.tv

# ♦ Metadonnées (fichier xml)

```
blip.tv
Your video, your way
```

Tags donnés par l'utilisateur ayant uploadé la vidéo désigné par <uid> et <login>

</speechSegment>

</SegmentList>

</AudioDoc>

Information sur le locuteur

Genre et numéro : (M/F)S#

# ♦ Transcription et segmentation en locuteurs

```
<SpeakerList> <Speaker ch="1" dur="55.08" gender="1" spkid="MS1"</pre>
  lang="eng-usa" lconf="1.00" nw="166" tconf="0.75"/> </SpeakerList>
<SegmentList>
<SpeechSegment ch="1" sconf="1.00" stime="2.70" etime="57.78" spkid="MS1"</pre>
lang="eng-usa" lconf="1.00" trs="1">
<Word stime="3.00" dur="0.35" conf="0.378"> Right </Word>
<Word stime="3.49" dur="0.23" conf="0.920"> from </Word>
<Word stime="3.72" dur="0.24" conf="0.240"> falling </Word>
<Word stime="4.18" dur="0.30" conf="0.915"> Angel </Word>
<Word stime="4.48" dur="0.13" conf="0.920"> of </Word>
<Word stime="4.67" dur="0.23" conf="0.909"> death </Word>
<Word stime="57.30" dur="0.36" conf="0.438"> ears </Word>
```

Information sur les segments de parole

- début/fin,
- qui parle en quelle langue

#### Transcription

- début / durée
- Score de confiance
- Hypothèse de mot

#### ♦ Transcription et réseaux de confusion



```
<Word stime="9.82" dur="0.22" conf="0.910"> Ron </Word>
<Word stime="10.04" dur="0.24" conf="0.894"> Paul </Word>
<Word stime="10.67" dur="0.23" conf="0.378"> worst </Word>
<Word stime="10.67" dur="0.23" conf="0.376"> worse </Word>
</word stime="56.94" dur="0.17" conf="0.960"> easy </word>
<Word stime="56.94" dur="0.11" conf="0.407"> and </Word>
<Word stime="57.14" dur="0.11" conf="0.340"> on </Word>
</word stime="57.14" dur="0.11" conf="0.340"> on </word>
```

Même instant de début et même durée

</segmentation>

# ♦ Données vidéo \_ Detection de plan et image clé (milieu de plan)

```
<?xml version="1.0" encoding="utf-8" ?>
  <Segmentation>
   <CreationID>
     Crashsolo-OneMinuteRumpoleAndTheAngelOfDeath659
   </CreationID
        <InitialFrameID> Crashsolo-OneMinuteRumpoleAndTheAngelOfDeath659/394.jpg
                                                                              plan
       </InitialFrameID>
   <Segments type="SHOT">
       <Segment start="T00:00:00:00:000F1000" end="T00:00:56:357F1000">
       <Index> 0 </Index>
         <KeyFrameID time="T00:00:28:142F1000">
              CrashsoloOneMinuteRumpoleAndTheAngelOfDeath659/394.jpg
        </KeyFrameID>
     </Segment>
     <Segment start="T00:00:56:357F1000" end="T00:01:00:928F1000">
                                                                1 image clé par plan
       <Index> 1 </Index>
         <KeyFrameID time="T00:00:58:642F1000">
            Crashsolo-OneMinuteRumpoleAndTheAngelOfDeath659/821.jpg
        </KeyFrameID>
       </Segment>
 </Segments>
                                      M2 SID Challenge 2016 -- Etiquetage de vidéos en genre
```



Frontières du segment correspondant à un plan

**KEYFRAMES** 



oneminutecritic .wordpress .com

#### ♦ Données sociales



Code: 1271048

Post: http://blip.tv/file/1271048

http://twitter.com/crashsolo

Posted 'One Minute Rumpole and the Angel Of Death' to blip.tv:

http://blip.tv/file/1271048



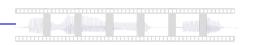
# **Genres disponibles**



Genre labels from Blip.tv	
1000 art	1001 autos_and_vehicles
1002 business	1003 citizen_journalism
1004 comedy	1005 conferences_and_other_events
1006 default_category	1007 documentary
1008 educational	1009 food_and_drink
1010 gaming	1011 health
1012 literature	1013 movies_and_television
1014 music_and_entertainment	1015 personal_or_auto-biographical
1016 politics	1017 religion
1018 school_and_education	1019 sports
1020 technology	1021 the environment
1022 the mainstream media	1023 travel
1024 videoblogging	1025 web_development_and_sites



# **Objectifs**



- ♦ Mettre en application les connaissances acquises dans les modules
  - Programmation
  - Analyse et exploitation de données
  - •
- ❖ Se confronter à un problème conséquent
- ❖ Travailler en groupe et effectuer une gestion de projet
- → Utilisation de Matlab/Java pour simplifier et réutiliser le travail effectué en TP



#### Corpus à utiliser



- ❖ Afin de limiter l'encombrement et la puissance de calcul nécessaire pour travailler sur le corpus MEDIAEVAL, nous vous proposons de travailler sur un sous corpus, constitué des catégories :
  - Autos and vehicule (code 1001)
  - Health (code 1011)
  - Movies and television (code 1013)
  - Religion (code 1017)
  - Food and drink (code 1009)
  - Litterature (code 1014)
  - Sports (code 1019)
  - Politics (code 1016)
- ♦ Nous avons également limité à 6 Go la taille totale du sous-corpus.



- ♦ Mise en place d'un système de base
  - Descripteurs s'appuyant sur les données mises à disposition et vos acquis

Vidéo : couleur et texture des images clés

Audio: informations sur les locuteurs

Texte : représentation vectorielle des descriptions avec le schéma de pondération TF-IDF. Pour cela constituer le corpus des descriptions et construire l'index et posting associés

- → Pour chaque média, un vecteur de descripteurs représente le contenu (fichier)
- Classification

Traitement effectué indépendamment sur chaque média

Utilisation d'une méthode non supervisée : k-means

Evaluation des résultats par média : pureté des clusters obtenus





#### ♦ Amélioration des descripteurs

Prise en compte du contenu

Vidéo: couleur et texture sur la vidéo complète

Audio: calcul d'énergie, ZCR, MFCC...

#### Texte:

- Enrichir les descriptions de textes en utilisant wikipédia simple (vu en TP)
- Vérifier les entités nommées candidates : si elles existent, alors récupérer leurs descriptions; si elles n'existent pas, utiliser des expressions régulières pour récupérer des descriptions additionnelles de l'entité. (eg. is \*\*\*\*), citées dans une autre page wikipédia
- Enrichir les classes : les classes ont des titres qu'il faut enrichir à partir des pages wikipédia.



- → Pour chaque média, un (ou plusieurs) vecteur(s) de descripteurs représente(nt) le contenu
- ♦ Pour le texte, utiliser :
  - Vecteur tf idf métadonnées
  - Vecteur tf-idf enrichi
  - Vecteur tf idf des classes enrichies
  - Score de similarité entre le vecteur des métadonnées enrichi et le vecteur des classes
  - Classification Traitement effectué indépendamment sur chaque média

Utilisation d'une méthode supervisée : k plus proches voisins

Evaluation des résultats par un taux de reconnaissance, suite à une validation croisée :

leave-one-out

Taux de reconnaissance = nb fichiers reconnus / nb total de fichiers

Prise en compte du nombre de vecteurs par fichier, pour chaque méthode de classification





#### → Fusion des média

#### Classification tardive

Traitement effectué en fusionnant les résultats de chacun des médias Pondération possible suivant la qualité de résultats des médias (issue des systèmes précédents)

#### Précoce

Traitement effectué en fusionnant le vecteur de descripteurs de chacun des médias en un « super-vecteur » → Attention à la normalisation des descripteurs !

#### Présentation des résultats

Mode non-supervisé : pureté des clusters

Mode supervisé : taux de reconnaissance

#### **Organisation** (1/2)

- ♦ Groupe de 5 étudiants :
  - Un personne par activité : Vidéo, Audio, Texte, Classification
  - Un responsable de projet

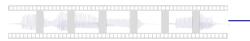
Trello
Organize anything, together

- ♦ Gestion et suivi de projets : Trello
  - Découpage du travail en tâches
  - Affichage de l'avancement et de la personne en charge de la tâche
  - Visible par le groupe et par les enseignants
  - Mise à jour régulière : quotidienne ou presque...



#### **Organisation** (2/2)

- ♦ Vous aurez également à remplir une feuille de suivi individuelle sur google drive
  - Chaque étudiant recopiera la liste des taches a faire en début de journée
  - Chaque étudiant indiquera en fin de journée :
    - L'avancement réalisé des taches sur la journée
    - Les difficultés rencontrées (cela permettra d'aider au dimensionnement des taches le lendemain)
- ♦ Le rôle du chef de projet est :
  - Organiser la réunion du matin avec tous les membres de son groupe
  - Vérifier que les fiches de suivi soient bien remplies
  - De participer comme les autres responsables au développement général.
- ❖ Le rôle d'un responsable (texte/audio/vidéo/classif)
  - De planifier le développement de l'activité
  - Est de vérifier que son activité avance comme prévu
  - De participer au développement général.



#### **Organisation**



- **♦** Support
  - 5 enseignants assureront le suivi
  - Contacts par mail ou sur rdv
  - Présence de 1h à 2h / semaine dans la salle

A préciser...

#### **Evaluation**



- ♦ Soutenance le 25/11/2016 avec tout le monde
  - 30 minutes par groupe
     Démonstration de l'outil

Chaque membre prend la parole / répond aux questions

- **♦** Rapport
  - Choix des méthodes
  - Organisation
  - Manuel utilisateur
  - Exemples de résultats
  - Qualité du code
  - Problèmes rencontrés et solutions envisagées
  - Bilan personnel

#### **Questions?**



