

CHIST-ERA Project Seminar

Bruxelles, March 21, 2017





Project Summary

- Collaborative **A**nnotation of multi-**M**odal, **M**ulti-**L**ingual and multi-m**E**dia documents

- CHIST-ERA call 2011 D2K

Oct. 1, 2012 to Nov. 30, 2016

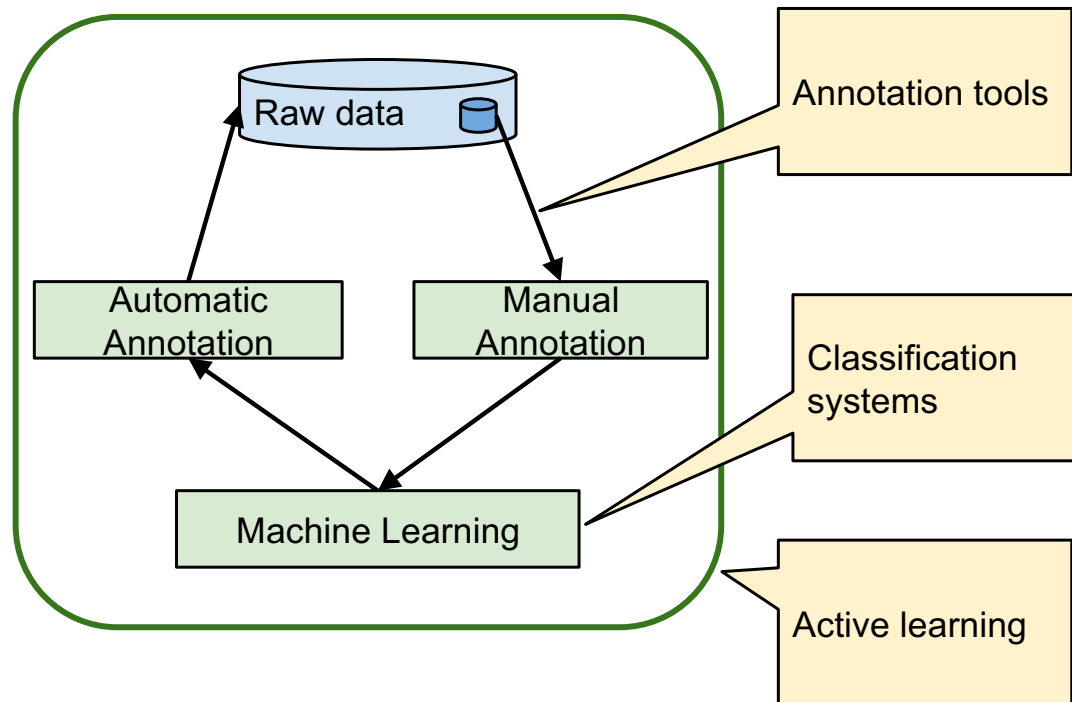
- 6 participants from 4 countries

France, Luxembourg, Spain, Turkey

Key challenges and objectives

- Scientific background

- Growing amount of heterogeneous, human-generated data
- Machine learning (still) needs lots of annotated data

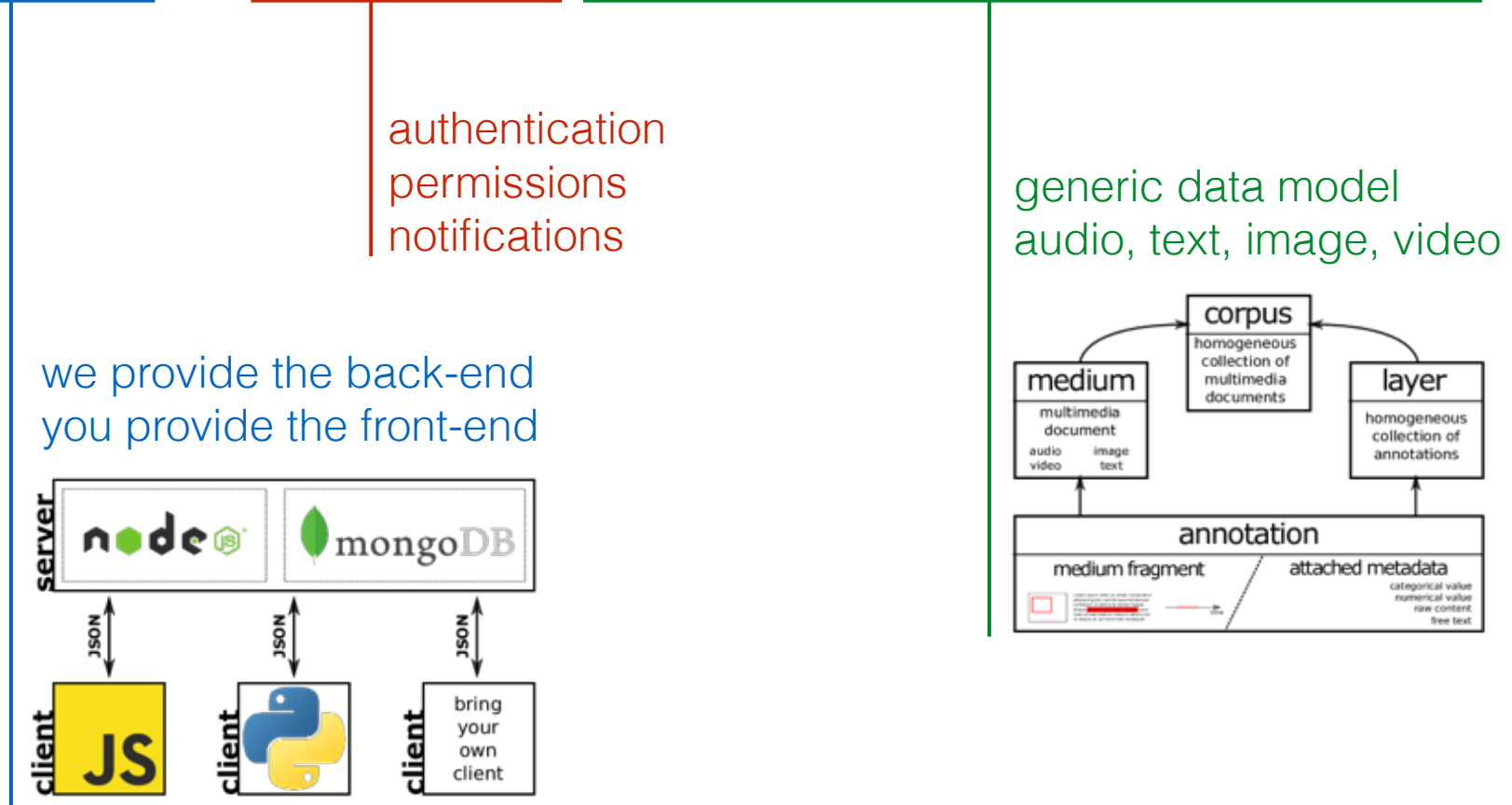


Key challenges and objectives

- Challenges
 - Collaborative annotation of multimodal, multilingual data
 - Combine automatic and manual annotations
- Unifying use case: multimodal person identification
 - Automatic recognition (speaker id, visual person recognition and fusion)
 - Active learning strategies
 - Annotation framework and evaluation campaigns

Collaborative Annotation Framework

A platform for collaborative annotation of multimedia documents



Collaborative Annotation Framework



OPEN-SOURCE server and client code on GitHub (camomile-project/camomile-server)

The screenshot shows the camomile web application. On the left is a dark sidebar with a search bar and a list of navigation items: prerequisite, Authentication, User, Group, Corpus, Create a corpus, Get list of corpus, Get a corpus, Update a corpus, Delete a corpus, Add media into a corpus, Create a layer into a corpus (highlighted in blue), Get list of medias into a corpus, Get list of layers into a corpus, get ACL of a corpus, update user ACL of a corpus, update group ACL of a corpus, remove user from ACL of a corpus, and remove group from ACL of a corpus. The main content area is titled 'Create a layer into a corpus'. It includes a restriction message: 'Restriction: user/group rights is 'O' or 'W' for this corpus'. Below this, it shows the HTTP REQUEST: 'POST http://example.com/corpus/<id_corpus>/layer'. It also lists QUERY PARAMETERS with a table:

Parameter	Type	Description
id_corpus	String	The corpus identifier (required)

. Then it lists DATA PARAMETERS with a table:

Parameter	Type	Description
name	String	Name of the layer (required, unique in the corpus)
description	free	A description of the layer
fragment_type	String	The fragment type (required)
data_type	String	The data type (required)
annotations	List	list of annotations (OR , ,) to add to the new layer

. On the right, there is a code editor with tabs for 'shell', 'python', and 'javascript'. The 'shell' tab is active, showing a curl command:

```
curl -i -X POST http://example.com/corpus/<id_corpus>/layer --data '{"name":<name>, "description":<description>, "fragment_type":<fragment_tpy>, "data_type":<data_tpy>, "annotations":[{"media_name":<media_name>, "fragment":<fragment>, "data":<data>}]}' -H "Content-Type: application/json" --cookie "cookies.txt"
```

. Below the code, it says 'The above command returns JSON structured like this:' followed by a JSON object:

```
{  "ACL": {    "users": {      "5305d1a044dc07e805000003": "0"    }  },  "_id": "545387f8777a800000766bf3",  "description": {    "abc": "def"  },  "history": [    {      "date": "2014-10-31T13:00:40.604Z",      "id_user": "5305d1a044dc07e805000003",
```

New MediaEval 2015 and 2016 task

Multimodal person discovery



Multimodal

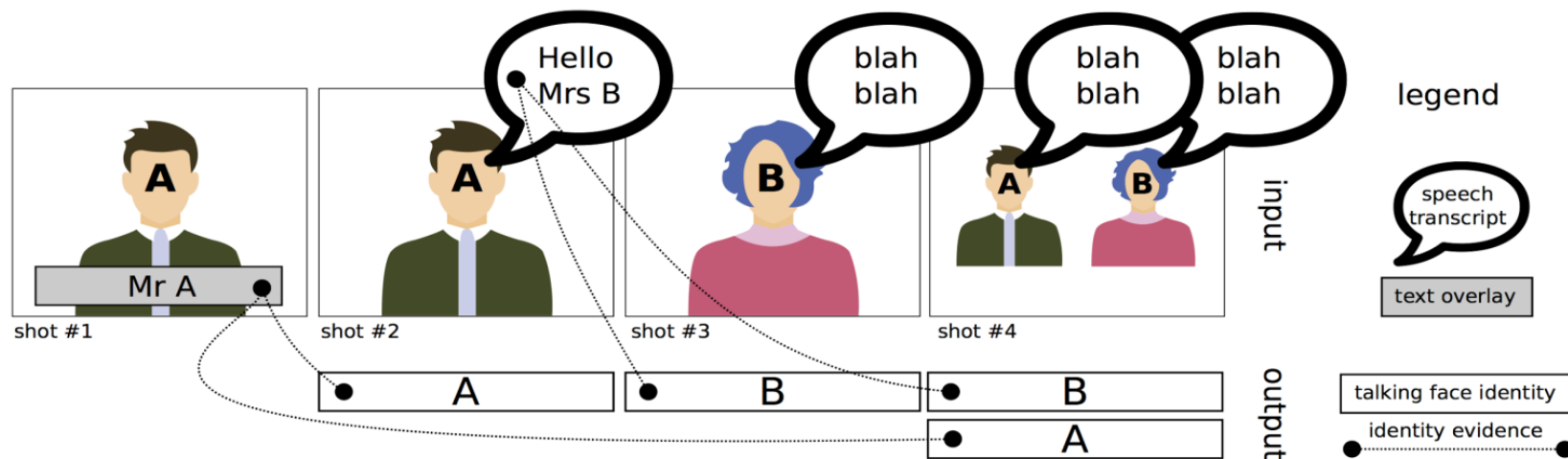
Identify persons of interest (audible **AND** visible at the same time) within a shot
Combine speaker identification, face recognition and speech/overlaid text transcription

Person discovery

Discover names in speech and overlaid text transcription
Unsupervised approaches only (no prior biometric models)

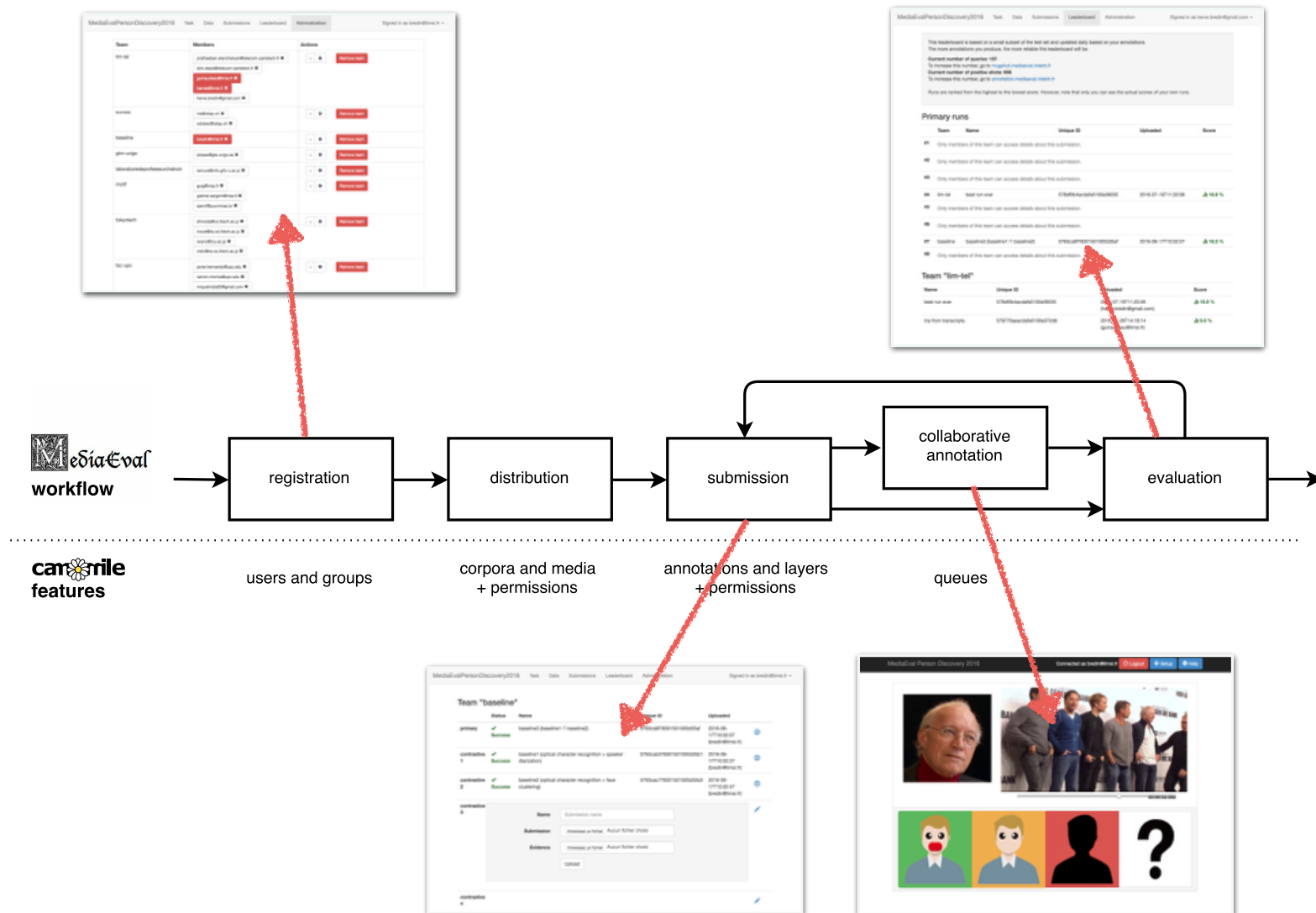
in TV Broadcast

2015 : INA French TV Broadcast News corpus (100 hours)



Distribution of a baseline system (monomodal and fusion components)
A posteriori collaborative annotation using Camomile framework

Workflow for Multimodal person discovery



Evidence annotation for Multimodal person discovery

MediaEval Person Discovery 2016

Connected as dev_mediaevalpersondiscovery2016

Logout Setup Help

firstname_lastname barry_corbin

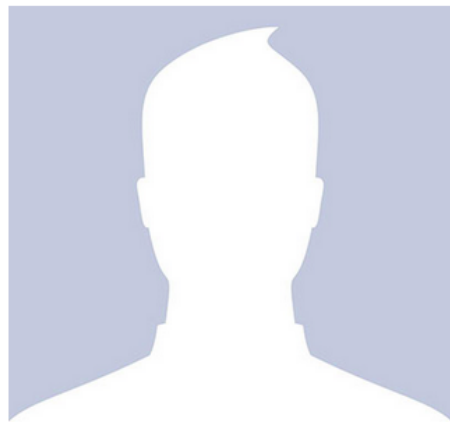


Barry Corbin

Play 00:51:47.840

Wrong person I am not sure

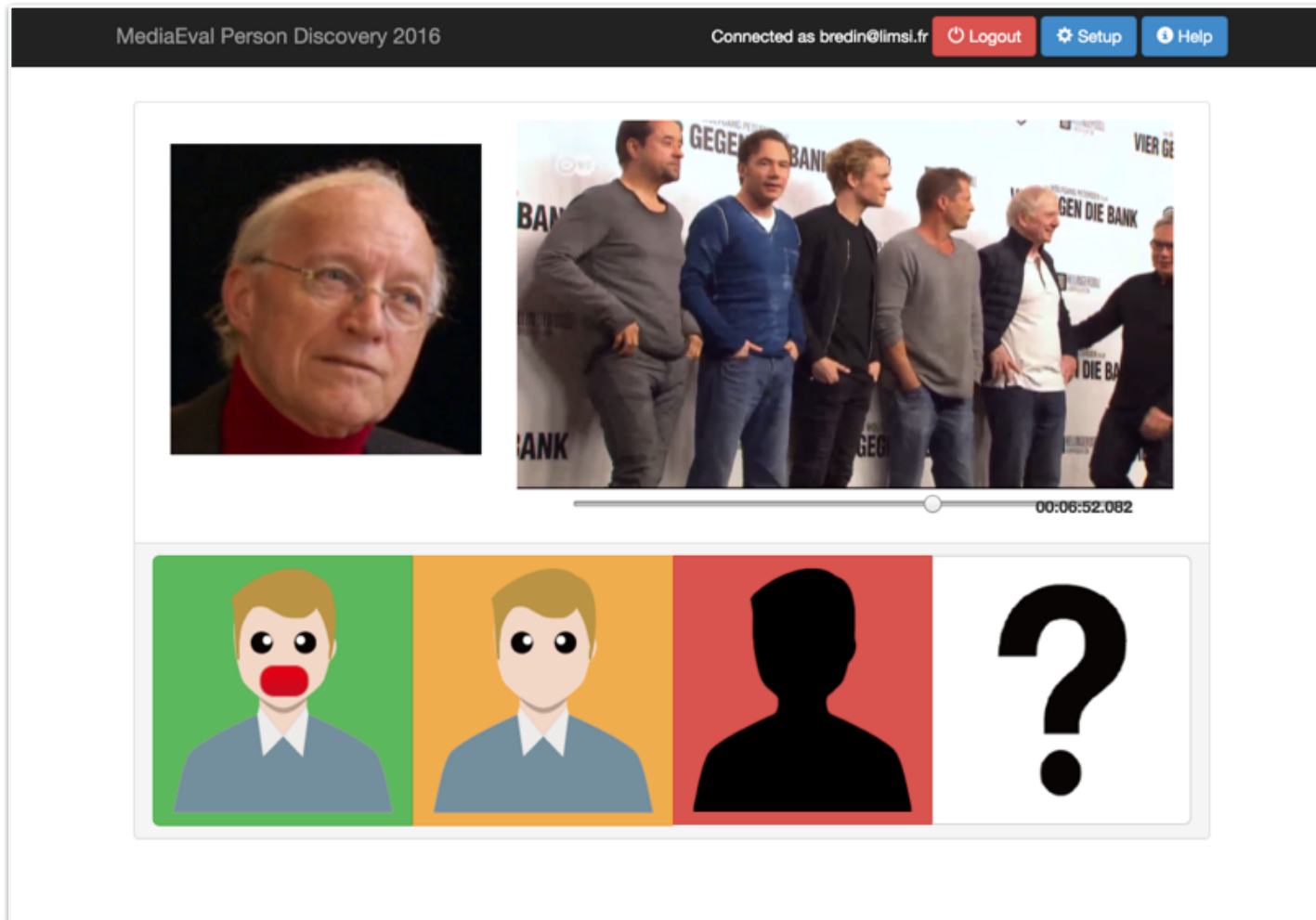
barry_corbin



Save

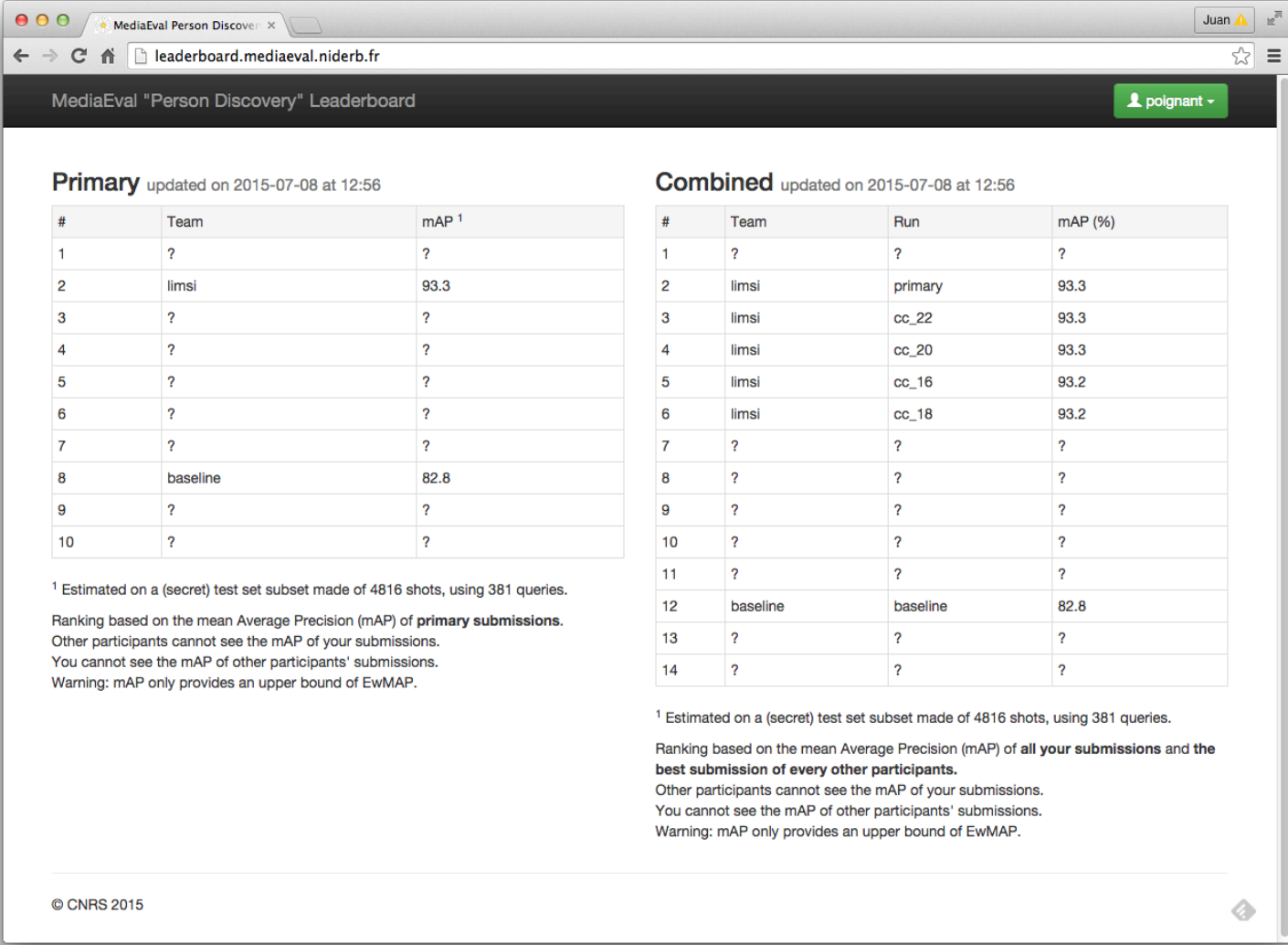
> 7k annotated evidences on 2015 data (10.2s/annotation)

Label annotation for Multimodal person discovery



28k shots annotated (4.4s/annotation) with (at least) double annotation

Leaderboard for Multimodal person discovery



MediaEval "Person Discovery" Leaderboard

Primary updated on 2015-07-08 at 12:56

#	Team	mAP ¹
1	?	?
2	limsi	93.3
3	?	?
4	?	?
5	?	?
6	?	?
7	?	?
8	baseline	82.8
9	?	?
10	?	?

¹ Estimated on a (secret) test set subset made of 4816 shots, using 381 queries.
Ranking based on the mean Average Precision (mAP) of **primary submissions**.
Other participants cannot see the mAP of your submissions.
You cannot see the mAP of other participants' submissions.
Warning: mAP only provides an upper bound of EwMAP.

Combined updated on 2015-07-08 at 12:56

#	Team	Run	mAP (%)
1	?	?	?
2	limsi	primary	93.3
3	limsi	cc_22	93.3
4	limsi	cc_20	93.3
5	limsi	cc_16	93.2
6	limsi	cc_18	93.2
7	?	?	?
8	?	?	?
9	?	?	?
10	?	?	?
11	?	?	?
12	baseline	baseline	82.8
13	?	?	?
14	?	?	?

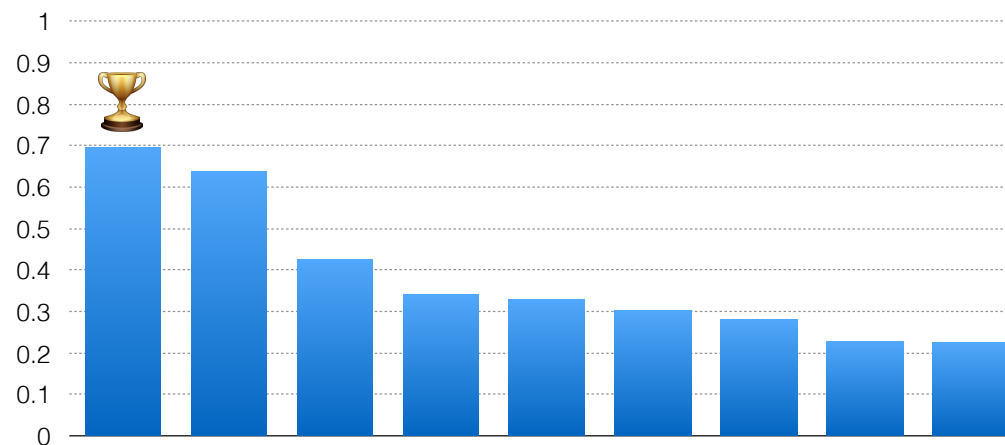
¹ Estimated on a (secret) test set subset made of 4816 shots, using 381 queries.
Ranking based on the mean Average Precision (mAP) of **all your submissions and the best submission of every other participants**.
Other participants cannot see the mAP of your submissions.
You cannot see the mAP of other participants' submissions.
Warning: mAP only provides an upper bound of EwMAP.

© CNRS 2015

2016 evolutions of Multimodal person discovery

- more diverse corpus: French (extracted from 1 full week from INA) ; German and English (Deutsche Welle), Catalan.
- collaborations: INA for French corpus, UPC for OCR and Catalan ASR, LIUM for French/German/English ASR, IDIAP for corpus selection and OCR...

MAP @ 10 | **599 queries**





Questions?
