Contextualizing Events in TV News Shows

José Luis Redondo García, Laurens De Vocht, Raphaël Troncy, Erik Mannens, Rik Van de Walle
<raphael.troncy@eurecom.fr>

http://goo.gl/xUjDrK

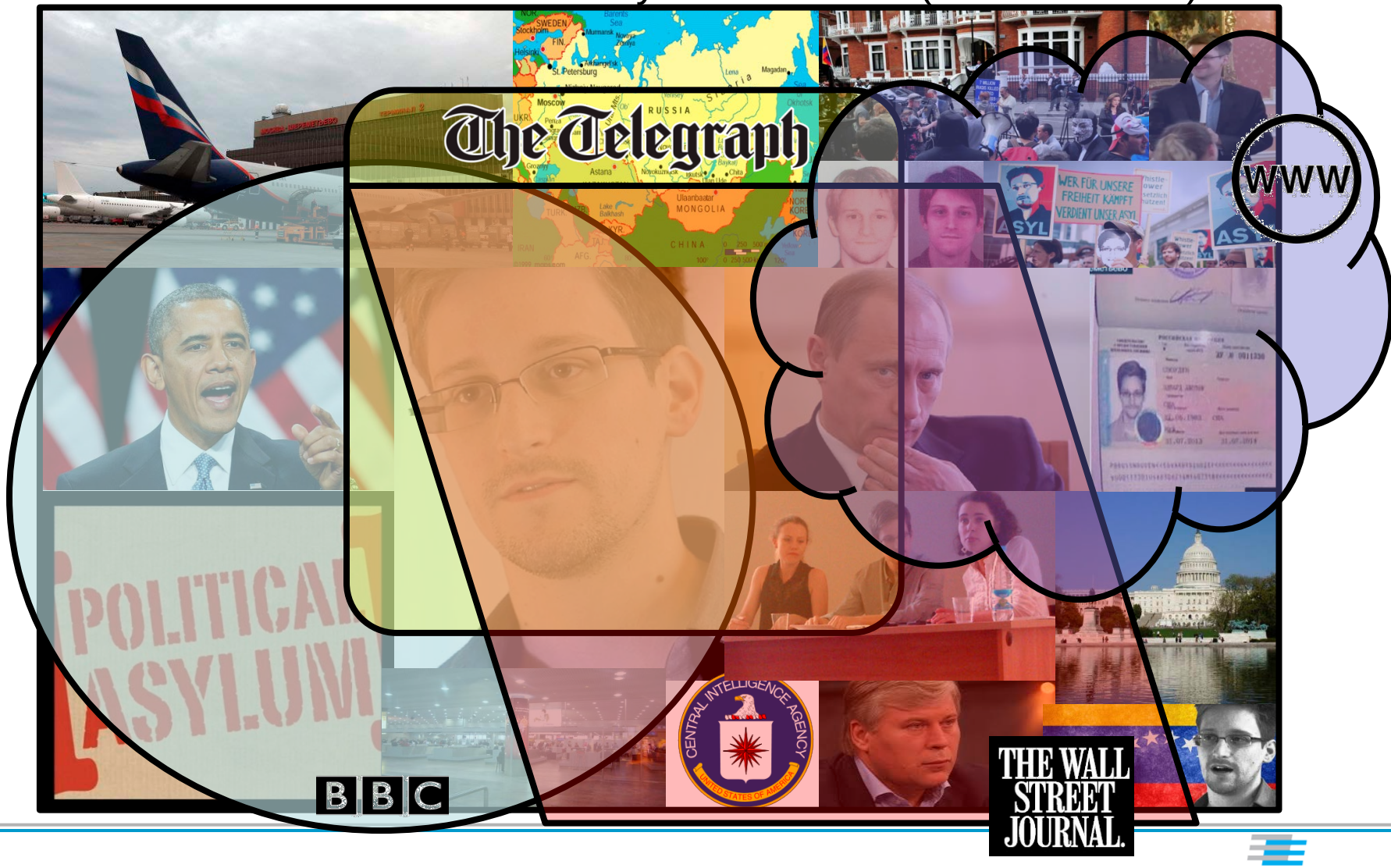
Scan the QR-code or use this code:

8360C07D

Hide second screen link

Problem: User Perspective

Edward Snowden asks for asylum in Russia (04 / 07 / 2013)



EURECOM

Problem: Technological Perspective

```
0:02:57,530 --> 00:03:01,350
nd the reason of course is that it

6
0:03:01,350 --> 00:03:05,960
roblems of very large scale and th

7
0:03:05,960 --> 00:03:12,810
inear programming exactly or with

8
0:03:12,820 --> 00:03:18,280
ubproblem or to solve subproblems

9
0:03:18,280 --> 00:03:24,850
iscrete optimization problems that

0
0:03:24,850 --> 00:03:30,410
ubproblem is a linear programming

1
0:03:30,750 --> 00:03:32,460
omes from the

2
0:03:32,470 --> 00:03:35,120
ractability of the linear programm

3
0:03:35,660 --> 00:03:42,970
o for convex optimization it's mor

4
0:03:42,970 --> 00:03:43,840
```

In which Russian airport is he exactly?

SUBTITLES

METADATA

```
XMLSchema-instance" xsi:schemaLocation="urn:ard:web:tva
<tva:ProgramDescription>
    <tva:ProgramInformationTable>
        <!-- (Description for whole show)-->
```

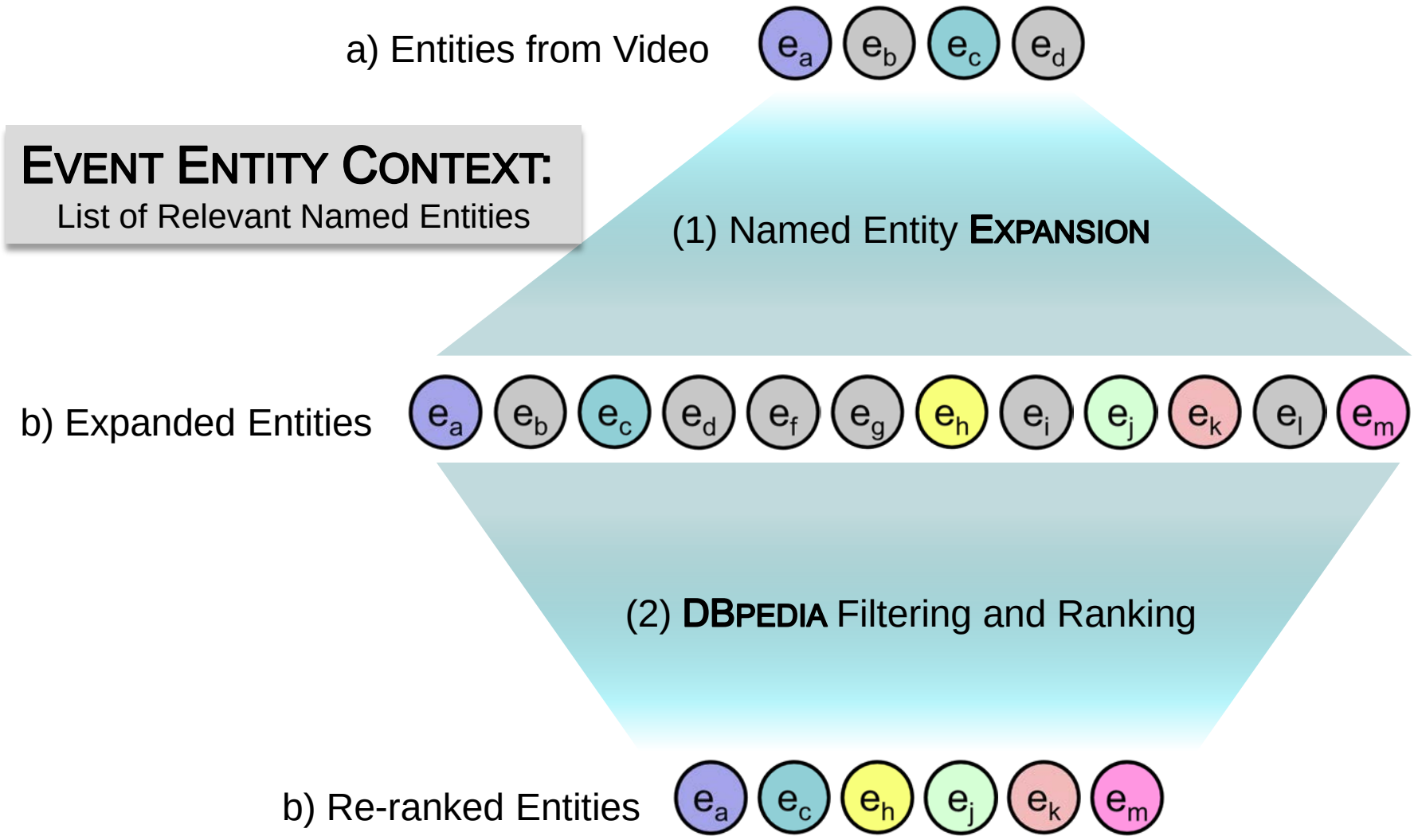
OCR

VISUAL CONCEPTS

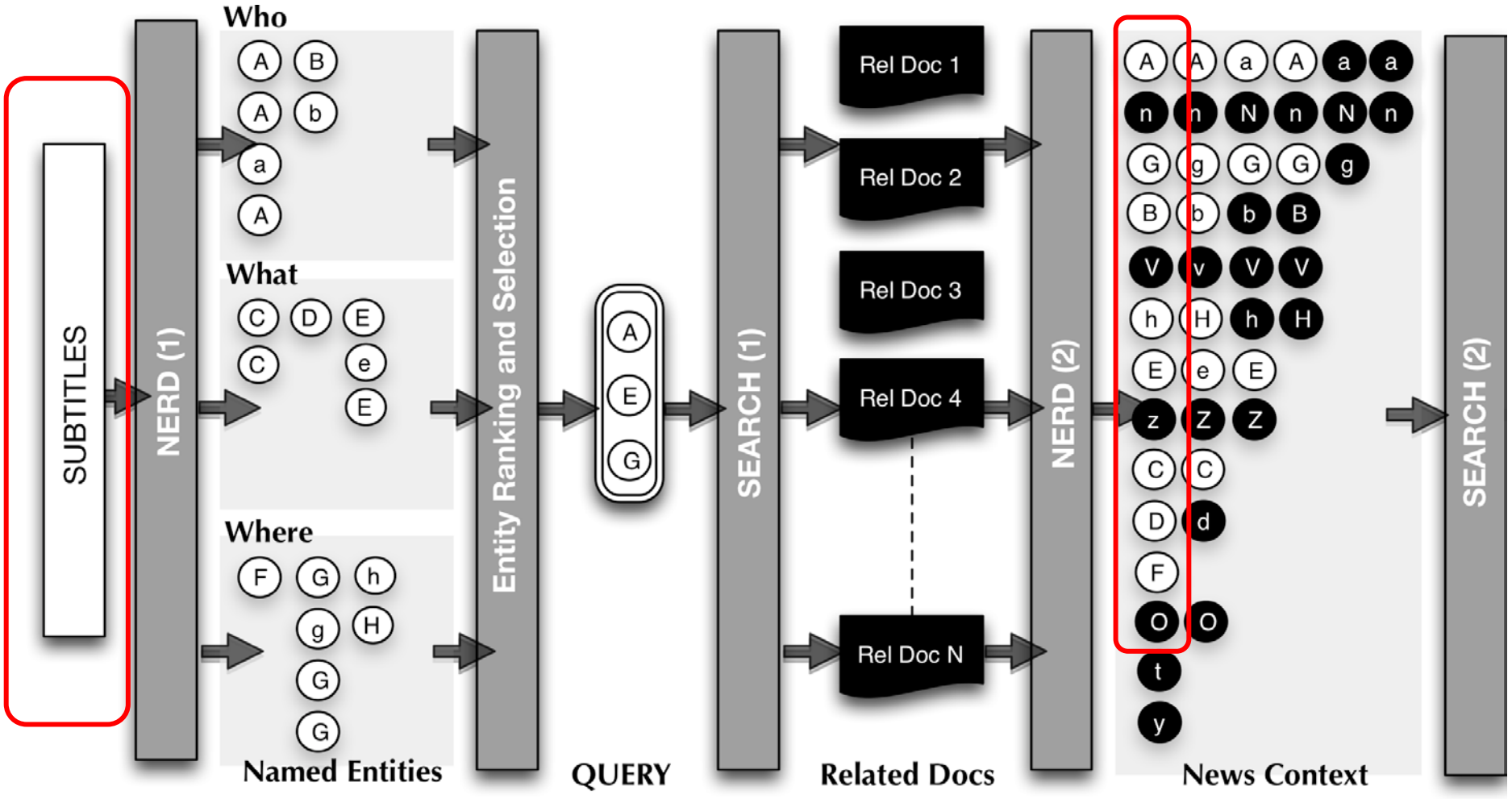
- LSCOM:Face
- LSCOM:Building

EURECOM

Approach

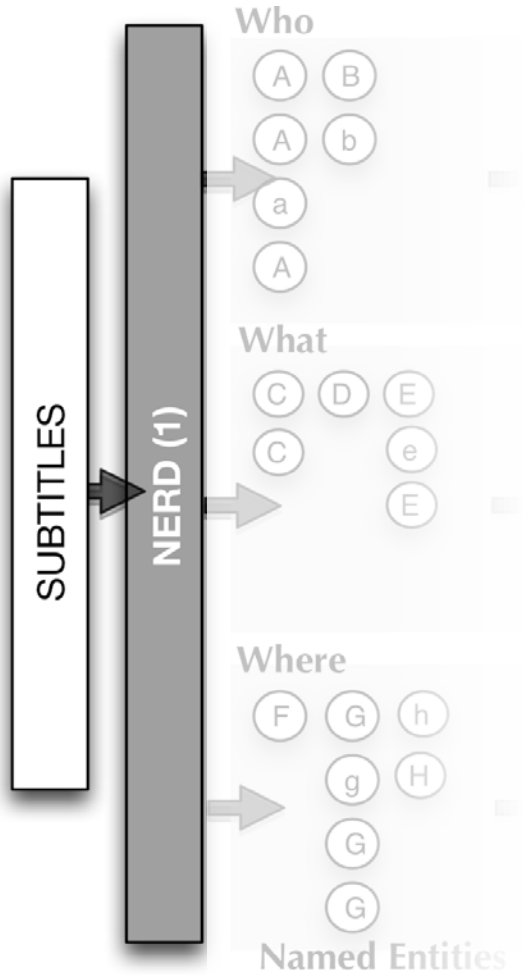

a) Entities from Video
e_a e_b e_c e_d
EVENT ENTITY CONTEXT:
List of Relevant Named Entities
(1) Named Entity EXPANSION
b) Expanded Entities
e_a e_b e_c e_d e_f e_g e_h e_i e_j e_k e_l e_m
(2) DBPEDIA Filtering and Ranking
b) Re-ranked Entities
e_a e_c e_h e_j e_k e_m

Named Entity Expansion



EURECOM

Named Entity Expansion: step 1

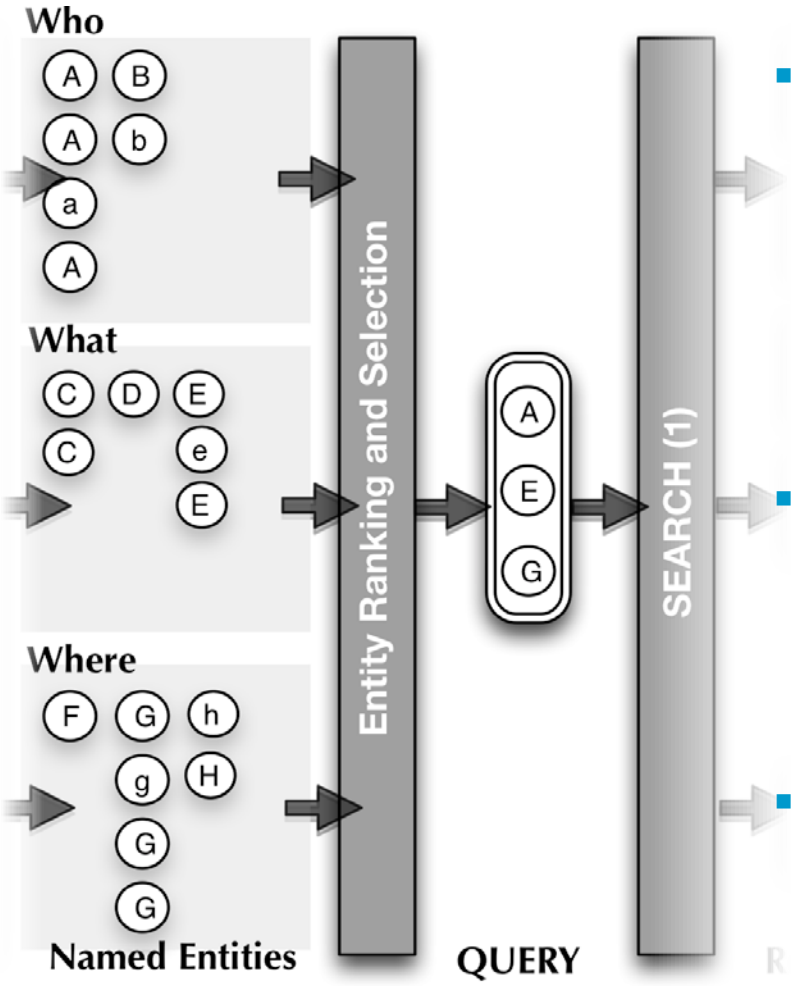


N·E·R·D

ontology[1] REST API[2] UI[3]

[1] http://nerd.eurecom.fr/ontology
[2] http://nerd.eurecom.fr/api/application.wadl
[3] http://nerd.eurecom.fr

Named Entity Expansion: step 2

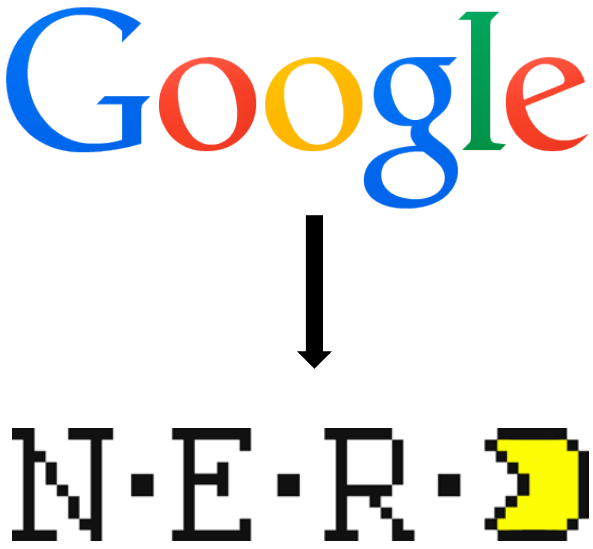


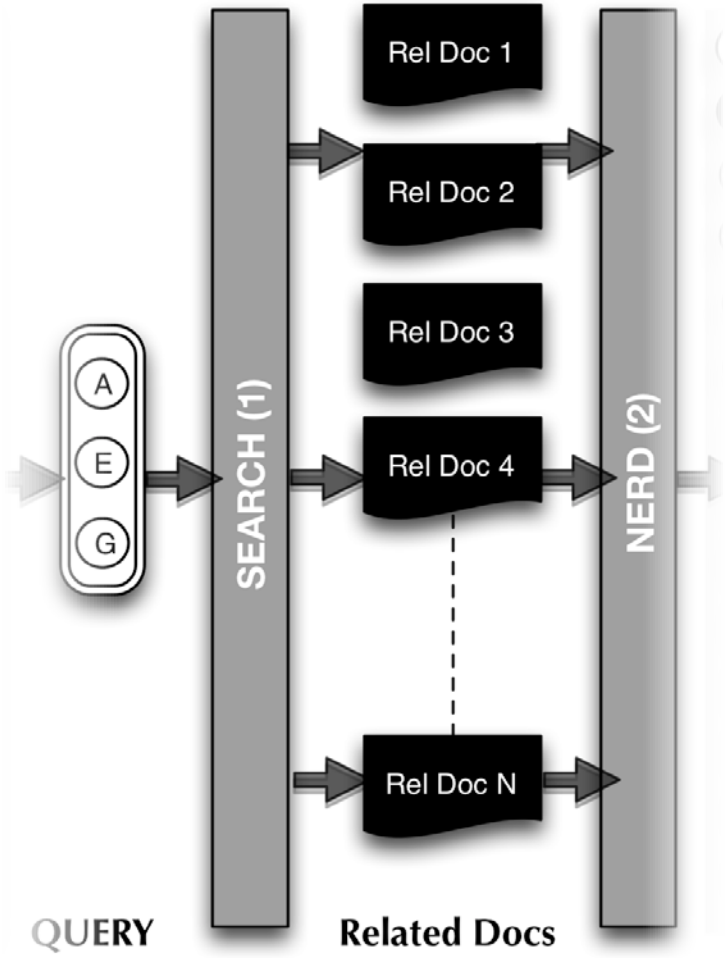
- **Five W´s * ➔ Four W´s**
  - **Who:** nerd:Person, nerd:Organization
  - **What:** nerd:Event, nerd:Function, nerd:Product
  - **Where:** nerd:Location
  - **When:** news program metadata
- **Entity Ranking and Selection:**
  - Ranking according extractor's confidence
  - Relative confidence falls in the upper quarter interval
- **Final Query:**
  - Concatenate **Labels** of the selected entities in **Who**, **What**, **Where**, for a time **t**

(*) J. Li and L. Fei-Fei. What, where and who? classifying events by scene and object recognition

Named Entity Expansion: step 3

QUERY

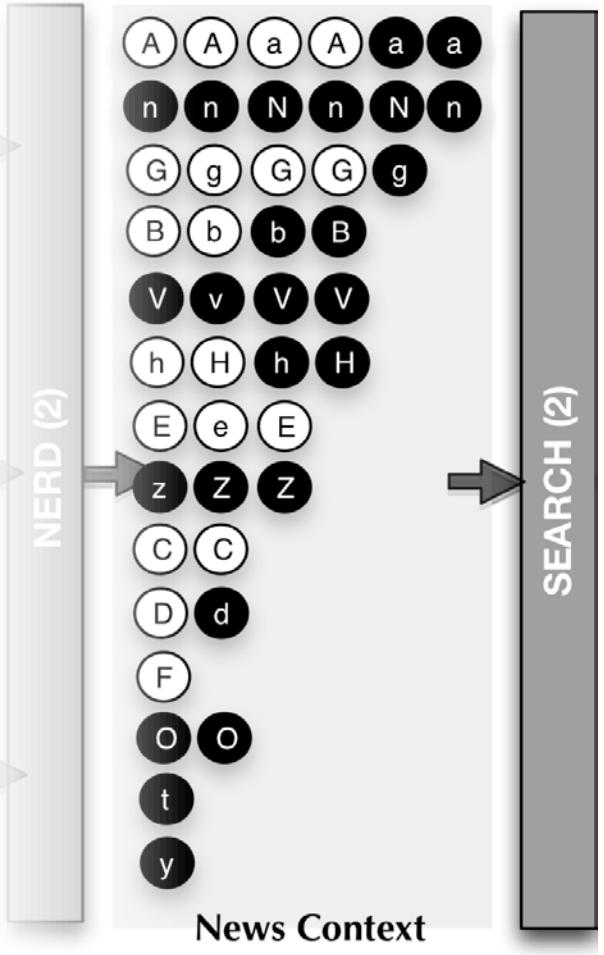




EURECOM

Named Entity Expansion: step 4

- **Entity clustering:**
  - **Centroid-based** approach
  - Distance metric:
    - Strict string similarity over the URL's
    - Jaro-Winkler string distance over labels
- **Entity re-ranking** according to**:**
  - Relative frequency in the transcripts
  - Relative frequency over the additional documents
  - Average confidence score from the extractors
- **Output:**
  - Frequent entities are promoted
  - Entities not disambiguated can be identified with a URL by transitivity
  - Same happens with erroneous labels
  - Relevant but non-spotted entities arise (example: **N**)



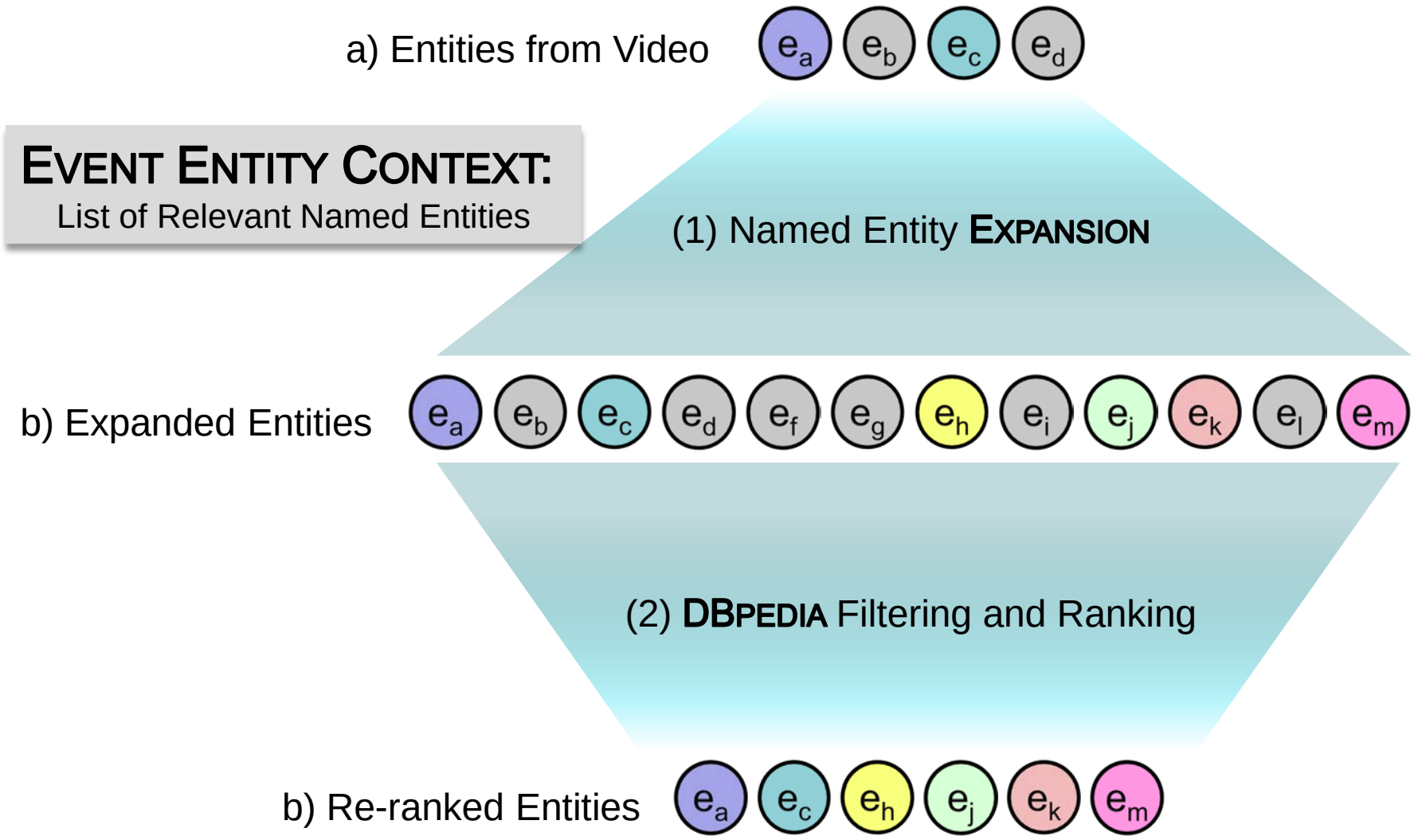
Named Entity Expansion: Results

Label	Relevance	Sentiment	Type	URI
Russia	0.809216	Mixed	Location	DBpedia:Russia
Edward Snowden	0.717369	Mixed	Person	DBpedia:/Edward_Snowden
South America	0.56586	Mixed	Location	DBpedia:South_America
president Putin	0.459811	positive	Person	DBpedia:Vladimir_Putin
president	0.401138	negative	JobTitle	DBpedia:President
Moscow	0.352101	Mixed	City	DBpedia:Moscow
CIA	0.334887	neutral	Organization	DBpedia:CIA
Bolivia	0.324607	neutral	Location	DBpedia:Bolivia
Obama	0.321901	negative	Person	DBpedia:Barack_Obama

Named Entity **EXPANSION**

	Label	Relevance	F_{video}	F_{Docu}	Type	URI
→	Russia	1.0	7	264	Location	DBpedia:Russia
→	Edward Snowden	0.80479	2	227	Person	http://dbpedia.org/resource/Edward_Snowden
→	US	0.61643	5	160	Location	DBpedia:United_States
→	Vladimir Putin	0.39383	1	111	Person	DBpedia:Vladimir_Putin
+	asylum	0.32876	4	80	Thing	DBpedia:Right_of_asylum
↗	Barack Obama	0.31506	1	88	Person	DBpedia:Barack_Obama
↗	Moscow, Russia	0.30479	1	85	Location	DBpedia:Moscow
↘	American president	0.19178	2	48	Thing	DBpedia:President_of_the_United_States
↗	Central Intelligence Agency	0.19178	0	56	Location	DBpedia:Central_Intelligence_Agency
+	Anatoly Kucherena	0.147260	0	43	Person	–
+	extradition	0.116438	2	26	Thing	DBpedia:Extradition
+	White House	0.10616	0	31	Location	DBpedia:White_House
+	Sheremetyevo	0.0890	0	26	Location	DBpedia:Sheremetyevo_International_Airport
+	WikiLeaks	0.08219	0	24	Organization	DBpedia:WikiLeaks
+	Washington's	0.075342	0	22	Location	DBpedia:Washington,_D.C.

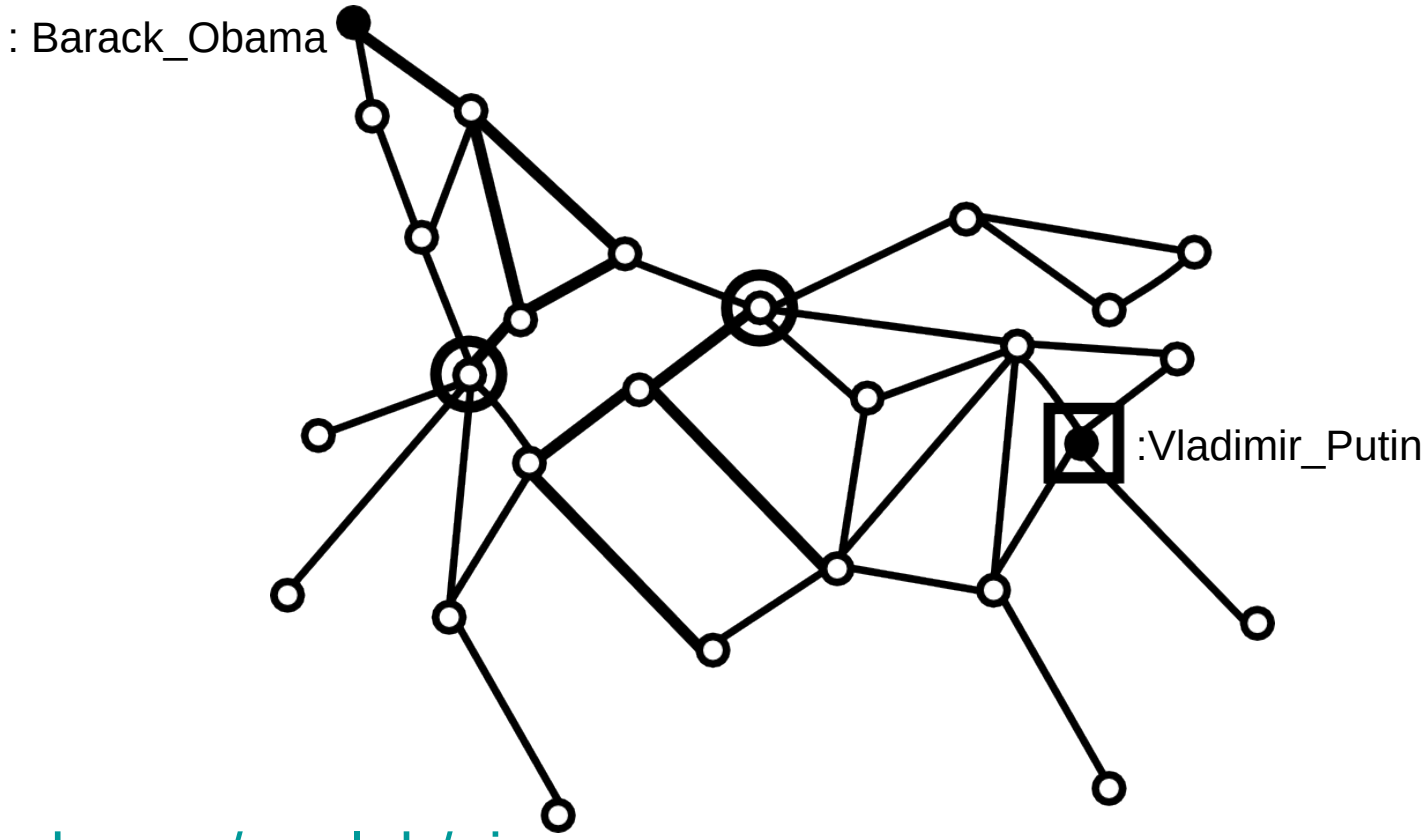
Approach



Refining via EiCE

For each pair of results:

Iteratively generate DBpedia paths using the **EiCE** engine [1]

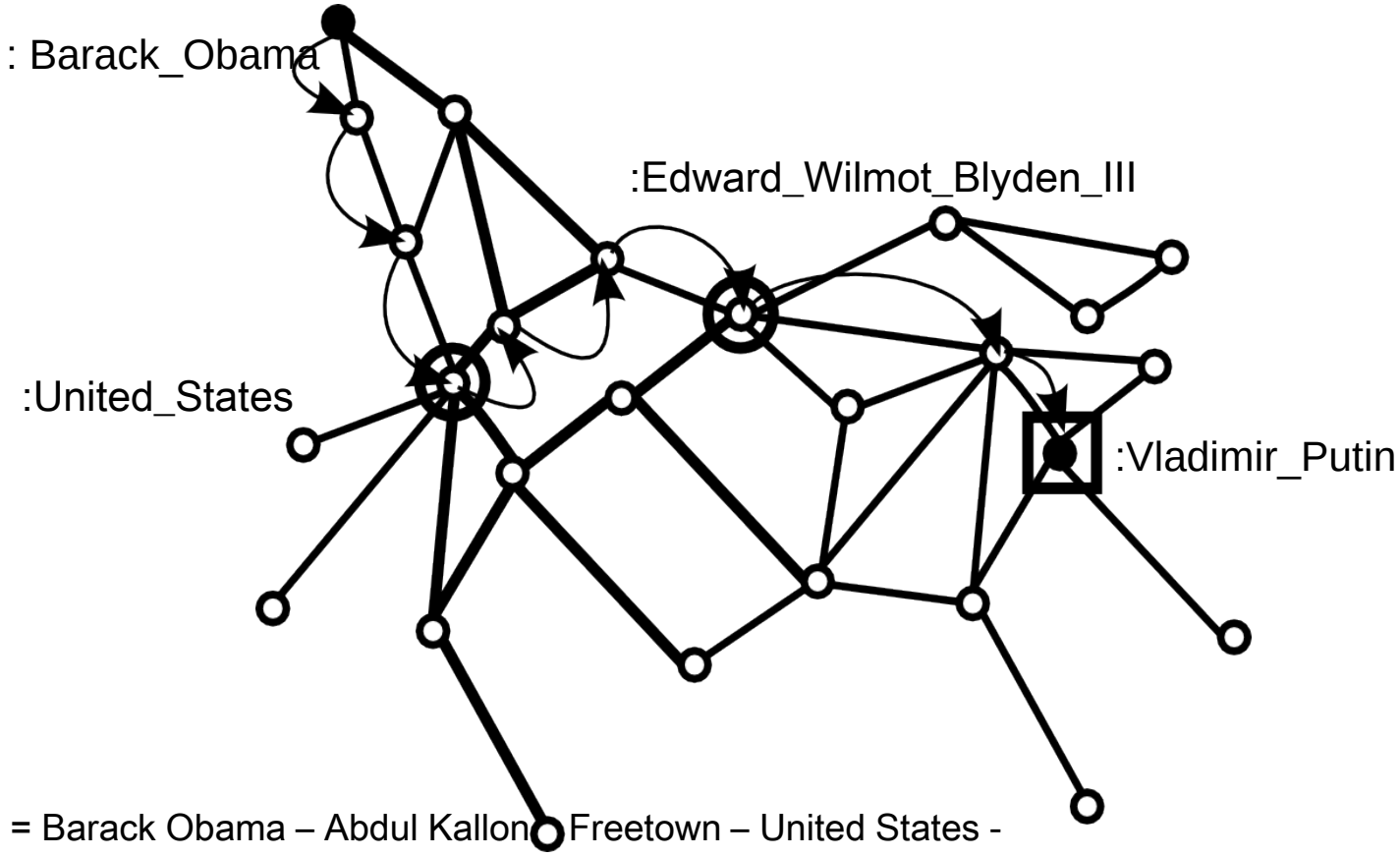


[1] http://github.com/mmlab/eice

Refining via EiCE

Computing of initial path:

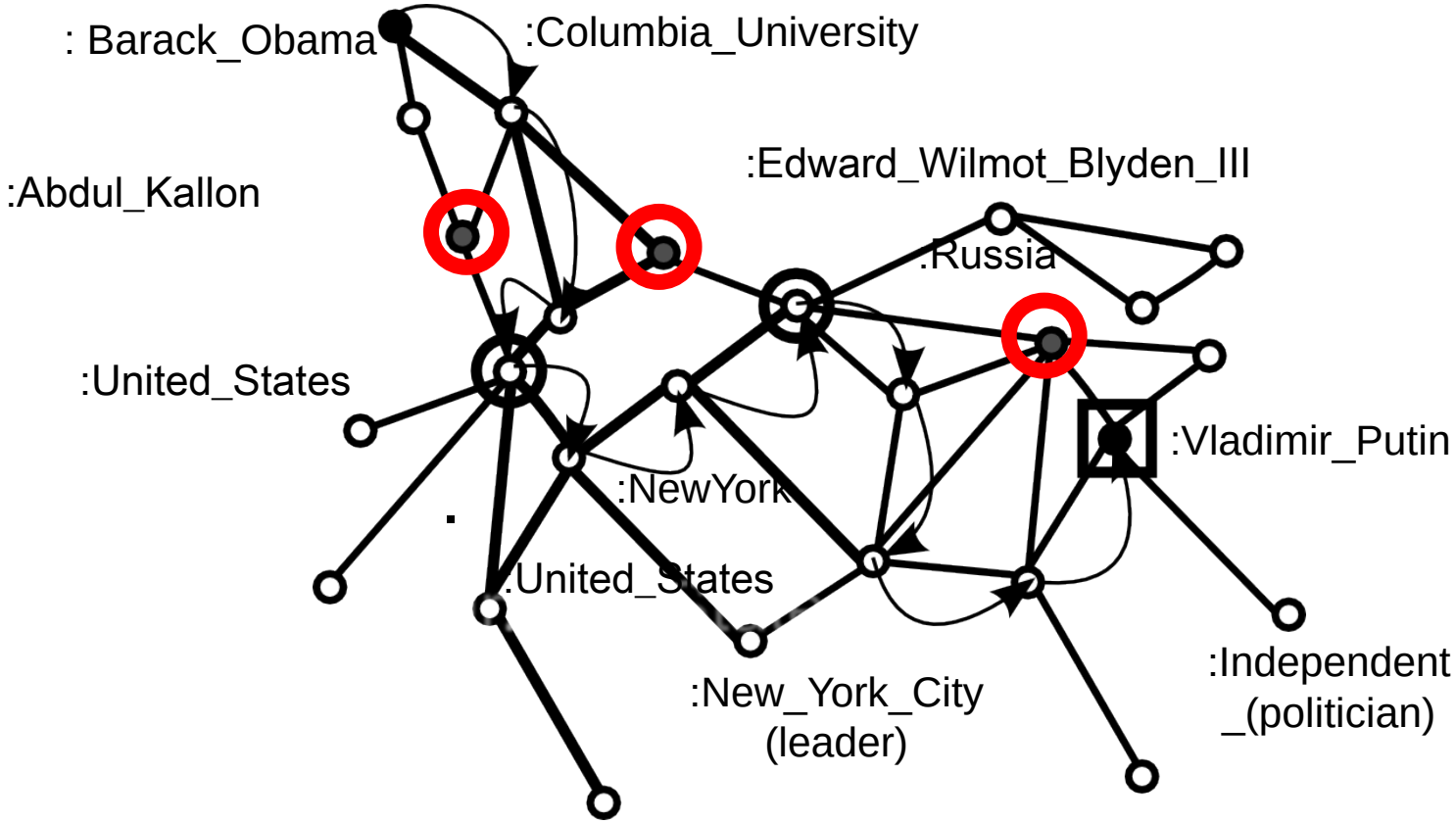
- Preferring links with lower weight first (thinner lines)



Path 1 = Barack Obama – Abdul Kallon – Freetown – United States - Edward_Wilmot_Blyden_III – Russia Vladimir_Putin

Refining via EiCE

Visited nodes being ignored in next iteration

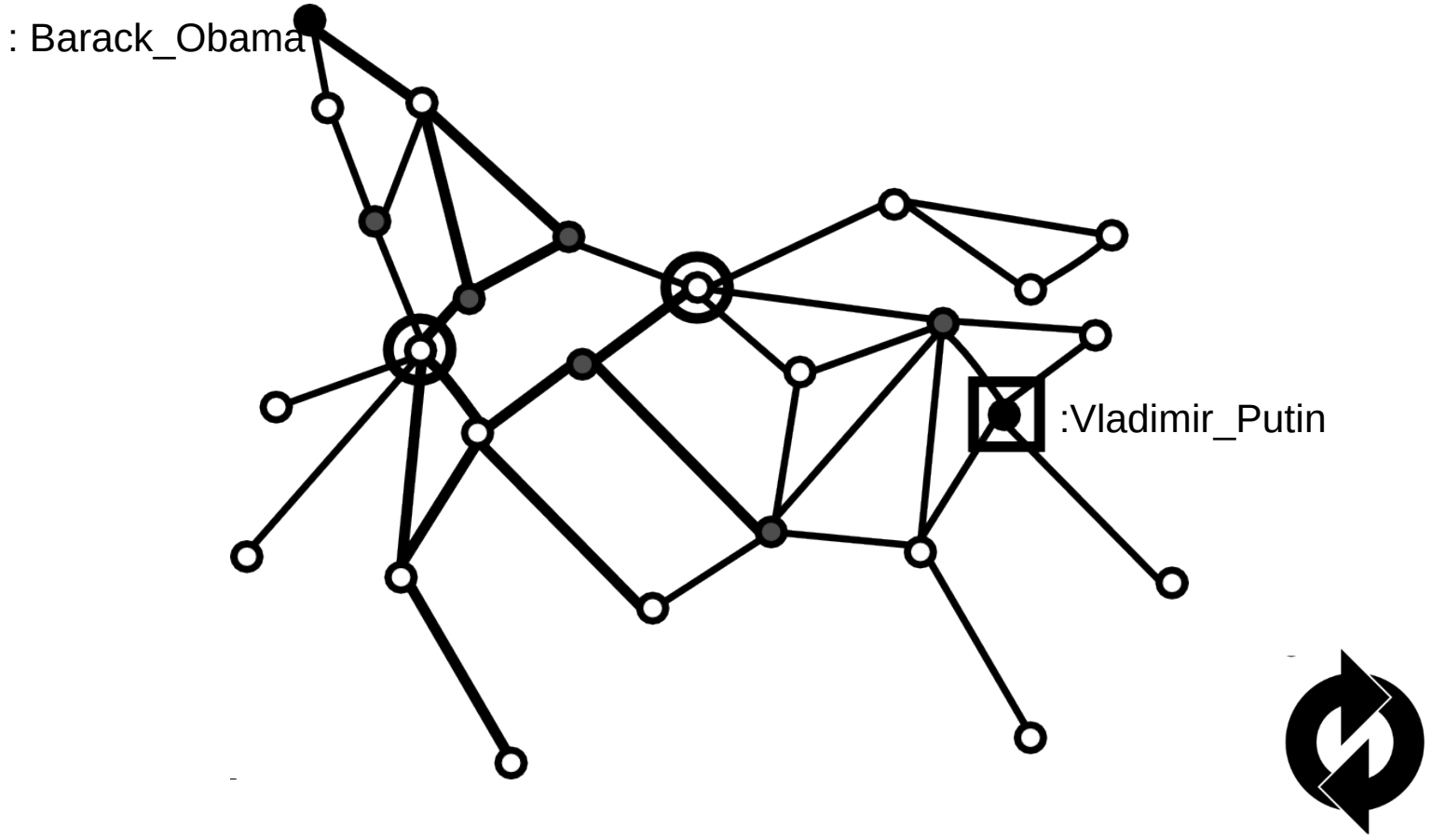


Path 2 = Barack Obama – Columbia University – United States – New York City– New York City(Leader) – Independent Politician -Vladimir_Putin

EURECOM

Refining via EiCE

Iterating continues until no more paths can be found



Refining via EiCE

Computation of the average path lengths after k iterations:

Length(Path 1) = 6
Length(Path 2) = 7
…
Length(Path k) = n

Average_Length (:Barack_Obama, :Vladimir_Putin) **=** (13 + … + n)/k

Distance (:Barack_Obama, :Vladimir_Putin) (*) **=** ***normalize*** (Average_Length)

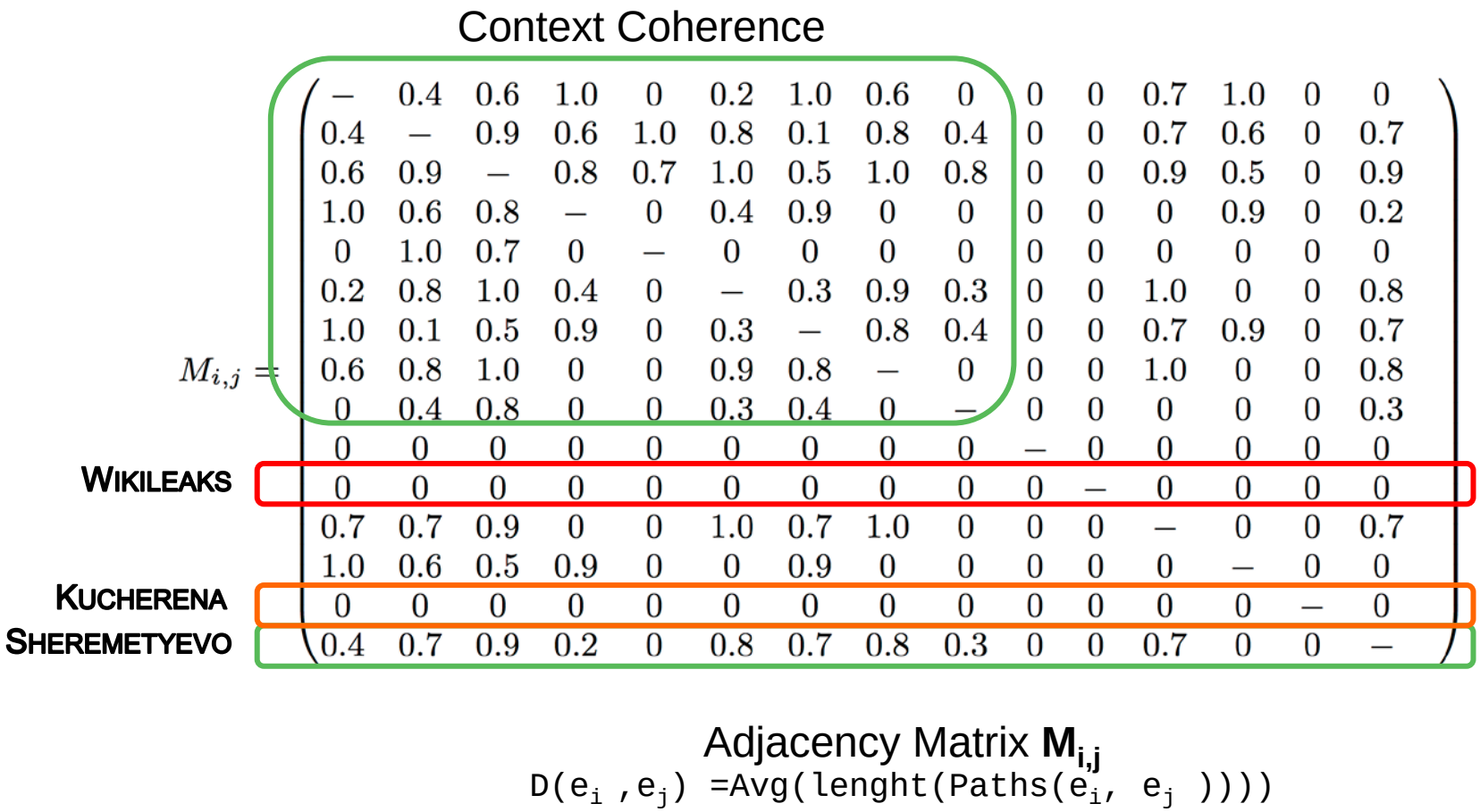
(*) http://demo.everythingisconnected.be/estimated_normalized_distance

…
…

Adjacency Matrix (**)

(**) http://demo.everythingisconnected.be/estimated_normalized_distance_matrix

Refining via EiCE: Adjacency Matrix



$$M_{i,j} = \begin{pmatrix}
- & 0.4 & 0.6 & 1.0 & 0 & 0.2 & 1.0 & 0.6 & 0 & 0 & 0 & 0.7 & 1.0 & 0 & 0 \\
0.4 & - & 0.9 & 0.6 & 1.0 & 0.8 & 0.1 & 0.8 & 0.4 & 0 & 0 & 0.7 & 0.6 & 0 & 0.7 \\
0.6 & 0.9 & - & 0.8 & 0.7 & 1.0 & 0.5 & 1.0 & 0.8 & 0 & 0 & 0.9 & 0.5 & 0 & 0.9 \\
1.0 & 0.6 & 0.8 & - & 0 & 0.4 & 0.9 & 0 & 0 & 0 & 0 & 0 & 0.9 & 0 & 0.2 \\
0 & 1.0 & 0.7 & 0 & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.2 & 0.8 & 1.0 & 0.4 & 0 & - & 0.3 & 0.9 & 0.3 & 0 & 0 & 1.0 & 0 & 0 & 0.8 \\
1.0 & 0.1 & 0.5 & 0.9 & 0 & 0.3 & - & 0.8 & 0.4 & 0 & 0 & 0.7 & 0.9 & 0 & 0.7 \\
0.6 & 0.8 & 1.0 & 0 & 0 & 0.9 & 0.8 & - & 0 & 0 & 0 & 1.0 & 0 & 0 & 0.8 \\
0 & 0.4 & 0.8 & 0 & 0 & 0.3 & 0.4 & 0 & - & 0 & 0 & 0 & 0 & 0 & 0.3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & - & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & - & 0 & 0 & 0 & 0 \\
0.7 & 0.7 & 0.9 & 0 & 0 & 1.0 & 0.7 & 1.0 & 0 & 0 & 0 & - & 0 & 0 & 0.7 \\
1.0 & 0.6 & 0.5 & 0.9 & 0 & 0 & 0.9 & 0 & 0 & 0 & 0 & 0 & - & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & - & 0 \\
0.4 & 0.7 & 0.9 & 0.2 & 0 & 0.8 & 0.7 & 0.8 & 0.3 & 0 & 0 & 0.7 & 0 & 0 & -
\end{pmatrix}$$

Adjacency Matrix $\mathbf{M_{i,j}}$

$D(e_i, e_j) = Avg(lenght(Paths(e_i, e_j)))$

Refining via EiCE: Results

Frequent **Nodes** and **Properties**

Nodes		Properties	
URL	Frequence	URL	Frequence
DBpedia:Wilmington,_North_Carolina	14	http://dbpedia.org/ontology/country	44
DBpedia:United_States	38	http://dbpedia.org/ontology/birthPlace	64
DBpedia:Russia	12	http://purl.org/dc/terms/spatial	42
DBpedia:conference/AIPR/2008	11	http://dbpedia.org/ontology/almaMater	26
DBpedia:Washington,_D.C.	10	http://dbpedia.org/ontology/location	21
DBpedia:Igor_Panarin	8	http://dbpedia.org/ontology/profession	20
DBpedia:Chap_Petersen	8	http://dbpedia.org/ontology/nationality	14
DBpedia:North_Carolina	8	http://dbpedia.org/property/leaderTitle	14
DBpedia:Independent_(politician)	8	http://dbpedia.org/ontology/occupation	14

Some **frequent classes** already detected are further **promoted**

In the context of the news program, **geographical** and **political** themes are dominant

Refining via EiCE: Results

NE Expansion

Label
Russia
Edward Snowden
US
Vladimir Putin
asylum
Barack Obama
Moscow, Russia
American president
Central Intelligence Agency
Anatoly Kucherena
extradition
White House
Sheremetyevo
WikiLeaks
Washington's

(2) DBPEDIA Ranking

DBPedia Connectivity

Label	Rel_{Final}	$Conectivity$	$F_{inPaths}$
Russia	1.0	6.10	12
Edward Snowden	0.79469	7.60	-
US	0.77188	9.40	38
Vladimir Putin	0.38383	5.50	-
Barack Obama	0.37231	6.69	-
Washington's	0.21544	6.10	10
Moscow, Russia	0.35991	6.90	-
asylum	0.25036	1.70	-
American president	0.21571	6.60	-
White House	0.12096	6.30	-
Sheremetyevo	0.11905	3.90	-
Central Intelligence Agency	0.09171	2.80	-
Anatoly Kucherena	0.08780	0.00	-
extradition	0.06693	0.00	-
WikiLeaks	0.04432	0.00	-

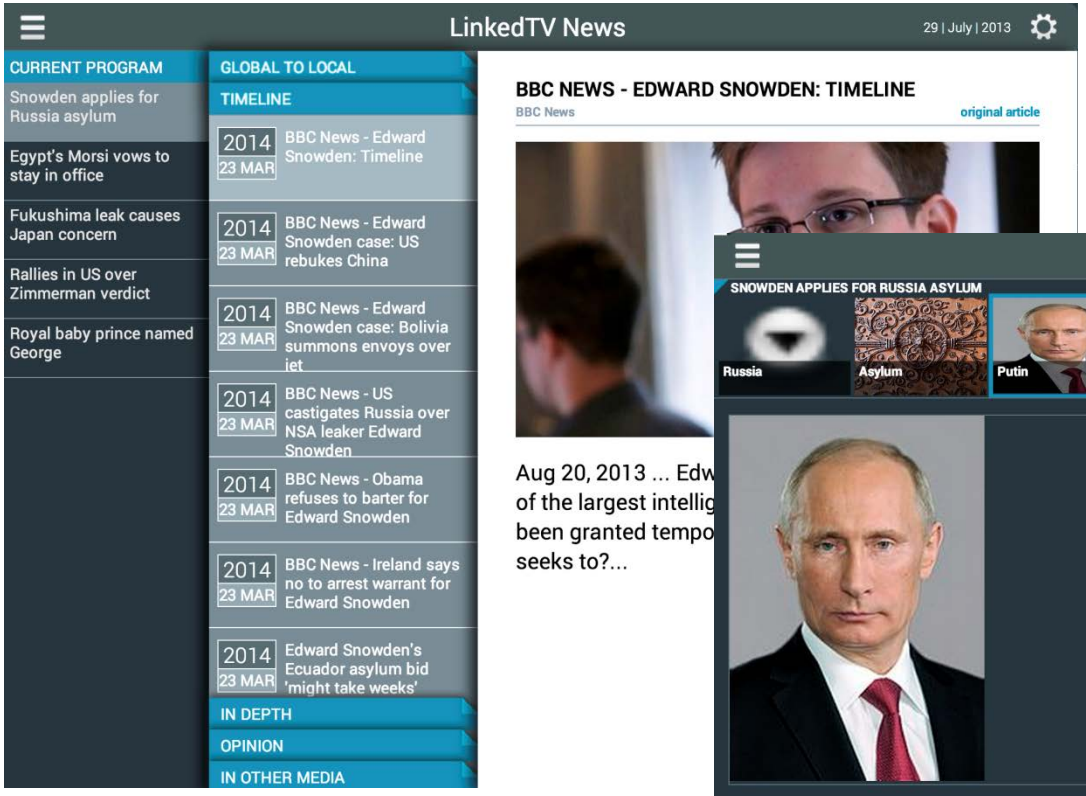
Evaluation: Edward Snowden Asylum

- **Setup:** Expert defines the list of relevant concepts for the newscast based on a deep analysis of the main argument and the feedback of 8 users participating in a user experience study [*]

GT Entity	Expert's Comment	NE Extraction	NE Expansion	DBpedia Connectivity
Edward Snowden	Public figure. He is the "who" of the news	✔	✔	✔
Russia	The location, but also an actor, the indirect object of the main sentence ""to whom"	✔	✔	✔
Political Asylum	This is related to the "what" of the news. This is the Snowden's request, the direct object.		✔	✔
CIA	Background information on related to Snowden, since he is an ex-CIA employee. An axe in a wider sense, not this item in particular, but on Snowden's history.	✔	✔	✔
Sheremetyevo	Airport, specific location of the news		✔	✔
Anatoly Kucherena	Secondary actor and speaker in the video. Information about an interview of person expressing his opinion.		✔	✔
US Department of State	Involved organization. Not mentioned but related with the main subject			
Other Entities			**"human-rights"** and **"extradition"**	Minister of Russia, "**Igor Shuvalov"**
Other Comments	**[*]** L. Perez Romero, R. Ahn, and L. Hardman. *LinkedTV News: designing a second screen companion for web-enriched news broadcasts.*			**Better ranking:** Entities like "Sheremetyevo" scored higher

EURECOM

Feeding a companion second screen app

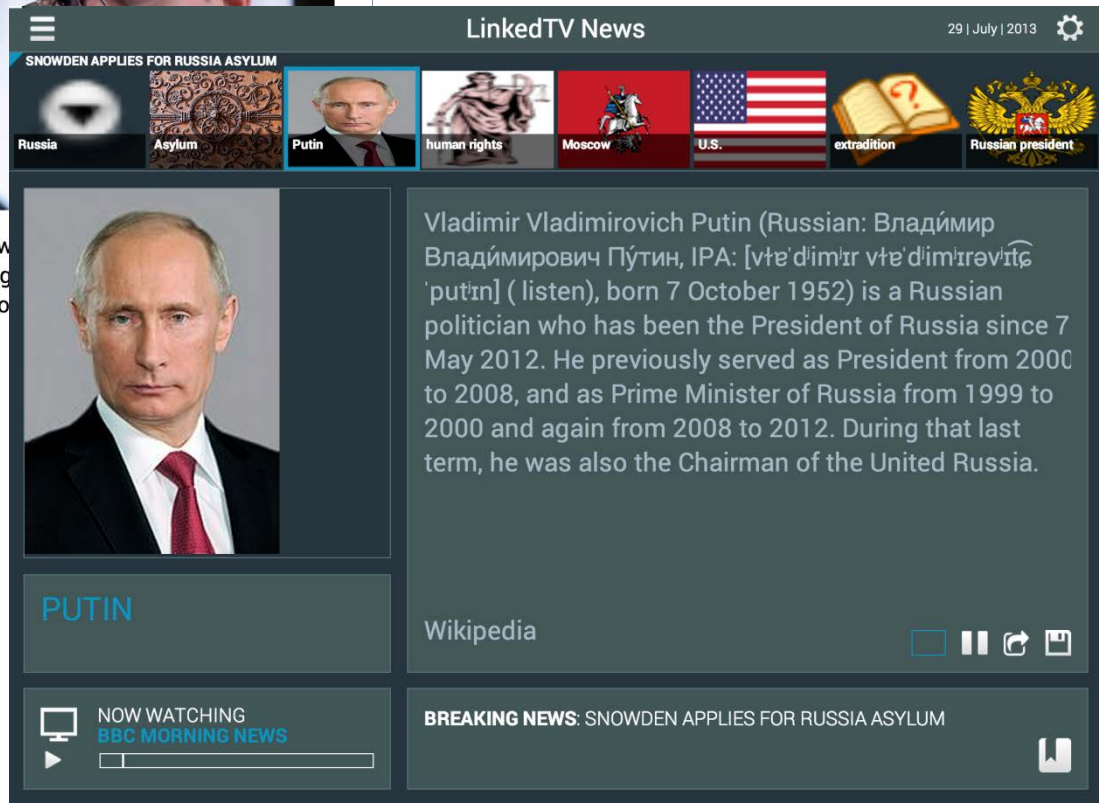


Second Screen @

- http://linkedtv.project.cwi.nl/news/

Main Screen @

- http://linkedtv.project.cwi.nl/news/mainscreen.html



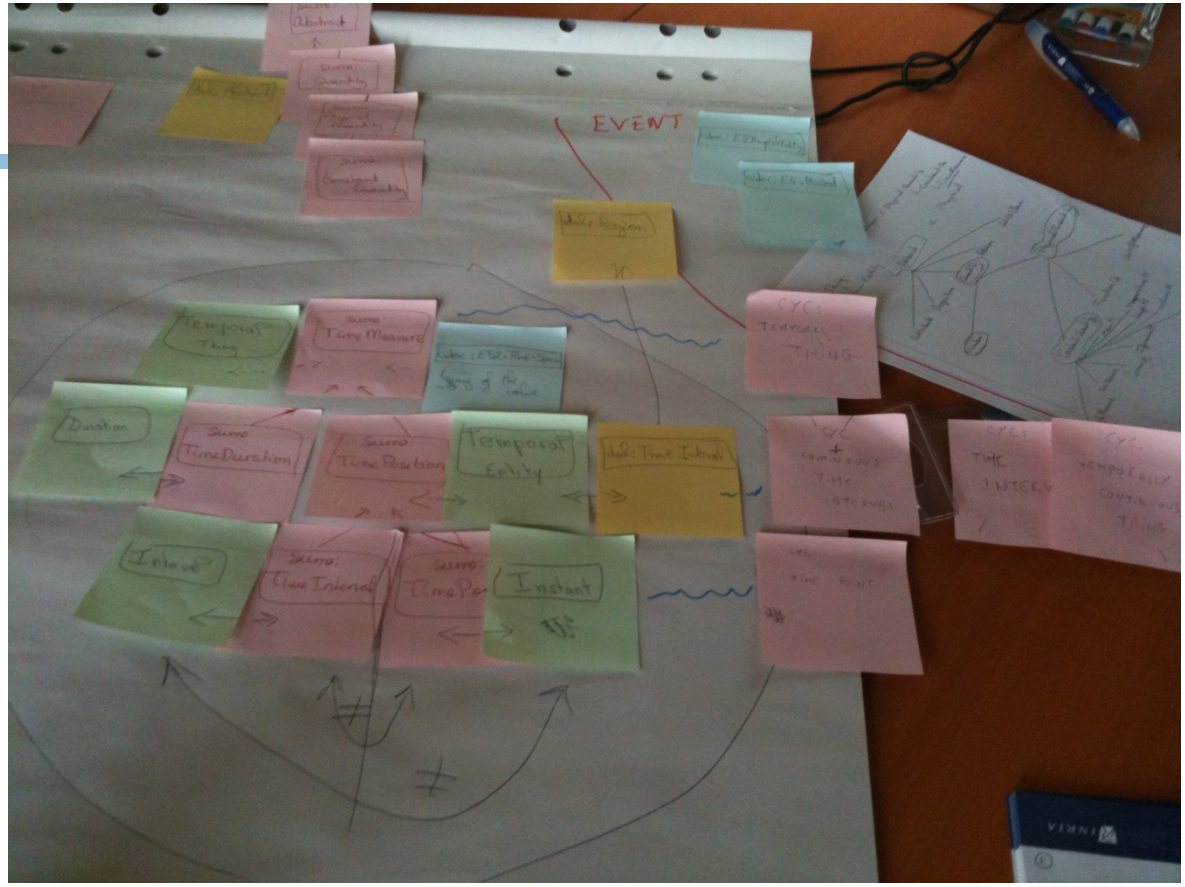
EURECOM

Conclusion

- **Approach for context-aware annotating news events:**
  - Start from named entities recognized in timed text
  - Expand this set by analyzing documents about the same event
  - Complete and re-rank exploring path connectivity in DBpedia
- **Preliminary results indicate that:**
  - The initial set of entities is enhanced with relevant concepts not present in the original program
  - By exploring DBpedia paths we:
    - ☞ Obtain a more accurate ranking of the relevant concepts
    - ☞ Bring forward more related entities and filter out the ones which are less representative in the broader context of an event
- **This *Entity Context* provides valuable data for a second screen application that enhances user experience**

Future Work

- **Evaluation involving more users and other programs**
- **Named Entity expansion:**
  - Study the adequacy of different extractors in NERD when:
    - Annotating the original transcripts of the video (*representative* entities)
    - Annotating additional documents found in the Web (*relevant* entities)
  - Introduce less ambiguity when generating the query by not only considering the surface form
- **Connectivity in DBpedia:**
  - Analysis of the relevant properties for promoting other entities
  - Cluster algorithms over the Adjacency Matrix for detecting meaningful groups of entities



slideshare BETA

http://www.slideshare.net/troncy