



Searching and browsing TED Talks
like you never imagined



José Luis Redondo García – redondo@eurecom.fr GO!

@peputo

@rtroncy

@marSabs

@PasqLisena

Raphaël Troncy – raphael.troncy@eurecom.fr

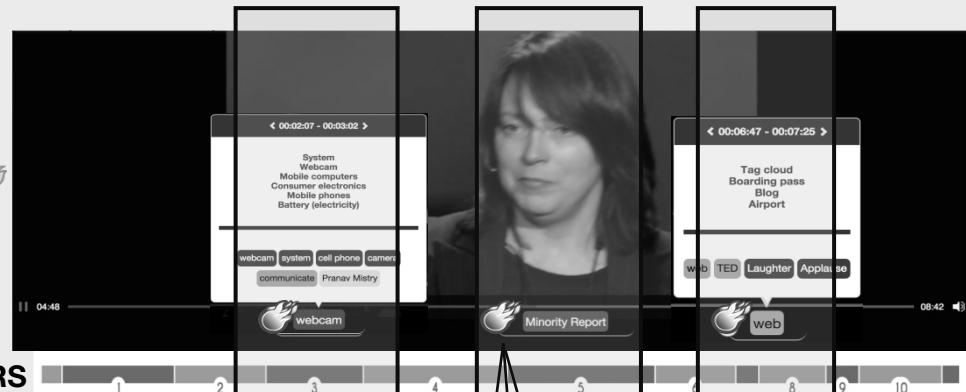
Mariella Sabatino – mariella.sabatino@eurecom.fr

Pasquale Lisena – pasquale.lisena@eurecom.fr

2014

< HYPER TED >

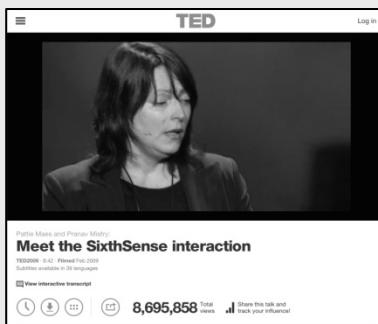
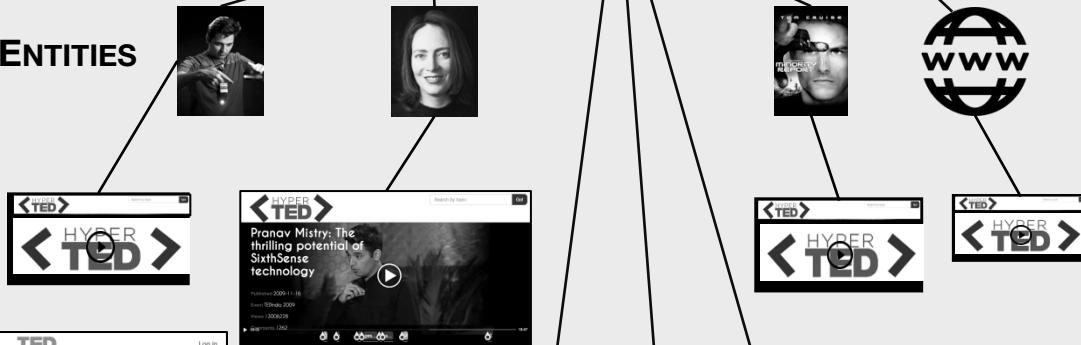
HOT SPOTS



CHAPTERS

RELATED TED'S CHAPTERS

ENTITIES



2006 TED.com

1984 TED Ideas worth spreading



UNDERSTANDING
ENVIRONMENT:
A SYSTEM
APPROACH

THE MYSTERIOUS
FIELD OF
ENGINEERING
SYSTEMS

SYSTEMS
PRACTICE:
MANAGING
SUSTAINABILITY

New Consuming Paradigm

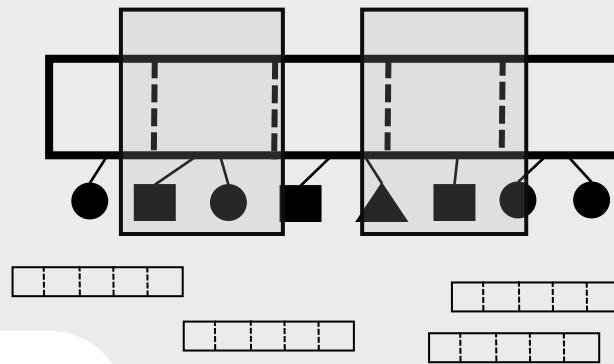
Users
overwhelmed
with audio-
visual content

Can the video
be divided into
meaningful
fragments?

How can those
fragments be
properly
described?

How can users easily find
related documents which
complement the video

What are the
potentially
relevant
fragments ?



HyperTED

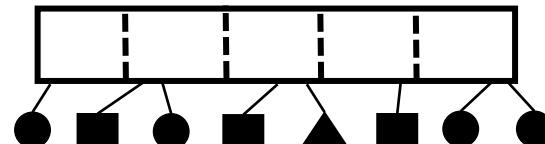
Media Fragment support

- ◆ Chapters
- ◆ Hot Spots



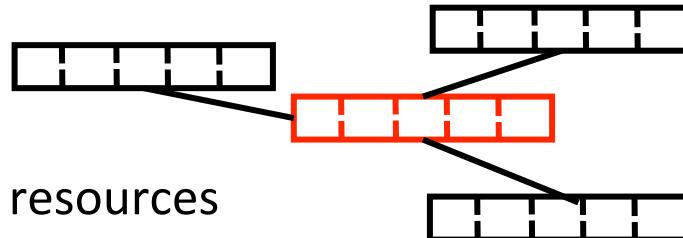
Media Fragment annotations

- ◆ Named Entity Extraction
- ◆ Topic Detection



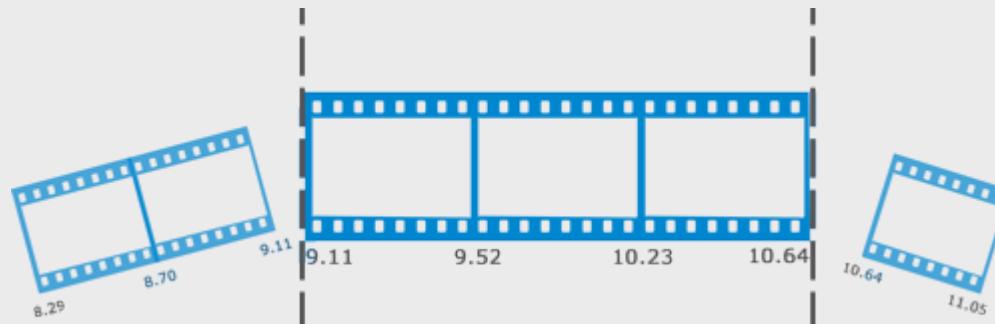
Hyperlinking

- ◆ With TED talks chapters
- ◆ With other educational online resources



◀ Media Fragments ▶

A *Media Fragment* is a portion of a multimedia resource.



Temporal Fragments

sections along the time dimension of the media resource with a start and an end point.

<http://www.w3.org/TR/media-frags/>

MF: Chapters

TED Talks have **paragraphs**:
a **human-made subdivision** of subtitles

The screenshot shows a TED Talk page for a video titled "down" from March 2014. The video thumbnail features a person with colorful braids. The transcript section has a "Transcript" heading and a "Select language" dropdown set to "English". Below the dropdown, there are three paragraphs of subtitles with red arrows pointing to them:

- 0:11 Good afternoon. My name is Uldus. I am a photo-based artist from Russia. I started my way around six years ago with ironic self-portraits to lay open so many stereotypes about nationalities, genders, and social issues — ["I am Russian. I sell drugs, guns, porno with kids!"] ["Vodka = water. I love vodka!"] (Laughter) — using photography as my tool to send a message. ["Marry me, I need a visa."]
 - 0
 - 1
 - 2
- 0:40 Today, I am still performing in front of the camera and trying to be brave like Wonder Woman. I focus on balancing meaningful message, aesthetic, beauty, composition, some irony, and artifacts.
- 0:56 Today, I'm going to tell you about my project, which is named Desperate Romantics. They're my artifacts, or paintings from pre-Raphaelites Brotherhood England mid-19th century. I took the painting and added new, contemporary meaning talking about issues which are surrounding me in Russia, capturing people who are non-models but have an interesting story.

MF: Chapters



Search by topic

Go!



◀ 11 chapters ▶

Annotations: Named Entities

"This is Nikita, a security guard from one of the bars in St. Petersburg."



*"This is **Nikita**, a **security guard** from one of the bars in **St. Petersburg**."*



Category:
type in the NER task.

Example taken from the transcript of
<https://www.ted.com/talks/2089>

Natural Language Processing (NLP)
Task → disambiguating URL in
a knowledge base.
E.g.
[http://dbpedia.org/resource/
Saint Petersburg](http://dbpedia.org/resource/Saint_Petersburg).

NER Extractors



TextRazor. Wikimeta

YAHOO! DEVELOPER NETWORK

saplo

ontotext

N-E-R-S
NAMED ENTITY RECOGNITION AND DISAMBIGUATION

- **Integrates** different NER tools available on Web.
- Unify NER extractors in a **common output**.

wikipedia

10DATI
ALL YOU NEED IS DATA

CALAIS
Powered by Thomson Reuters

THD

Zemanta
Blog Me Up!

Annotations: Named Entities

subtitles

chapter 5

So one important difference is that it's totally mobile.

Another even more important difference is that in mass production

this would not cost more tomorrow than today's cell phones

and would actually not sort of be a bigger packaging --

could look a lot more stylish

than this version that I'm wearing around my neck.

But other than letting some of you live out your fantasy

of looking as cool as Tom Cruise in "Minority Report,"

the reason why we're really excited about this device

is that it really can act as one of these sixth-sense devices

that gives you relevant information

about whatever is in front of you.

So we see Pranav here going into the supermarket

and he's shopping for some paper towels.

Annotations: Topics

Chapter 3

"I'm wearing a camera, just a simple webcam, a portable, battery-powered projection system with a little mirror. These components communicate to my cell phone in my pocket which acts as the communication and computation device. And in the video here we see my student Pranav Mistry, who's really the genius who's been implementing and designing this whole system..."



TextRazor.

Consumer electronics

Battery (electricity)

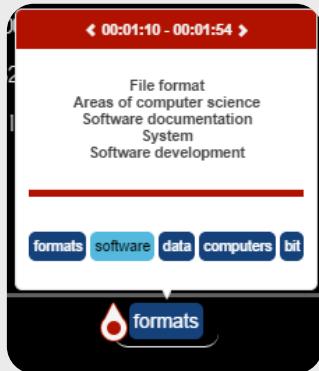
Mobile computers

Example taken from the transcript of
https://www.ted.com/talks/pattie_maes_demos_the_sixth_sense



MF: Hot Spots

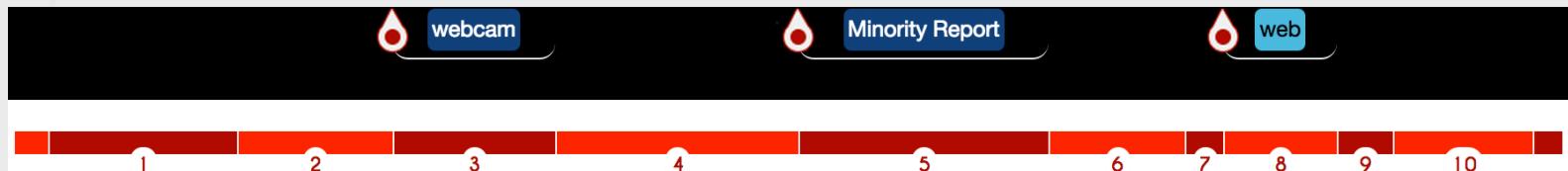
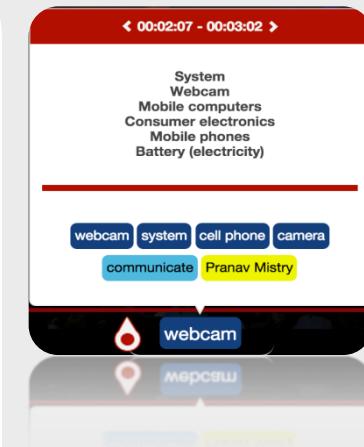
Hot Spot 1



Hot Spots
Chapters

1. Clustering of consecutive **chapters** which talk about similar **topics** and **entities**
2. Ordering of those fragments based on **annotation** relevance (TF-IDF)
3. Filtering: Hot Spots are fragments whose relative relevance falls under the **first quarter** of the final score distribution

Hot Spot 2



The logo features a stylized orange and yellow flame icon on the left, followed by the text "MF: Hot Spots" in a large, bold, red sans-serif font. The "MF" is in a smaller font size than "Hot Spots".

1. Clustering of consecutive **chapters** which talk about similar **topics** and **entities**

$$d(Ch_1, Ch_2) = w_{topic} \cdot \left(\frac{|T_1 \cap T_2|}{\max \{|T_1|, |T_2|\}} \right) + w_{entity} \cdot \left(\frac{|E_1 \cap E_2|}{\max \{|E_1|, |E_2|\}} \right) \quad (1)$$

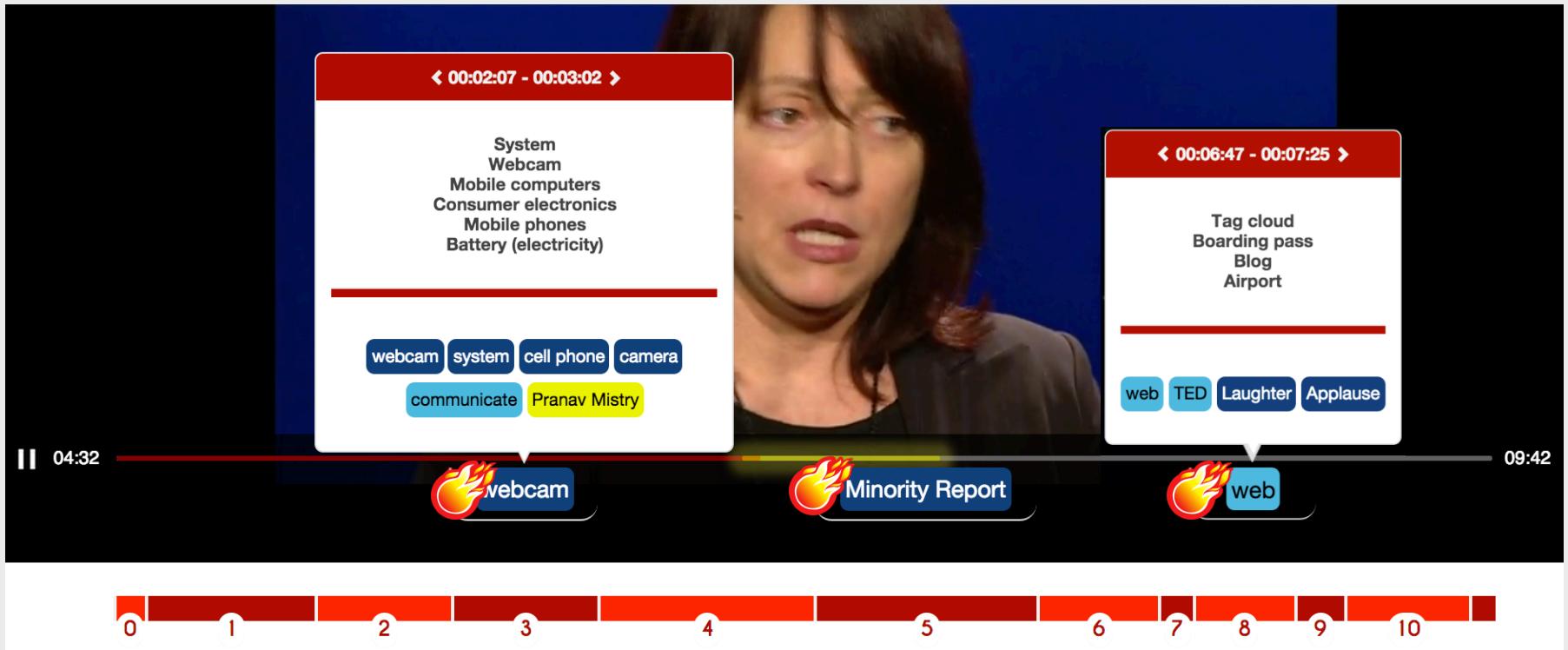
2. Ordering of those fragments based on **annotation** relevance (TF-IDF)

$$\text{Relevance}(\text{Fragment}) = \frac{w_{topic} \cdot \sum_{t \in T} Rel_t + w_{entity} \cdot \sum_{e \in E} Rel_e}{\text{Duration}(Ch)} \quad (2)$$

3. Filtering: fragments whose relative relevance falls under the **first quarter**



MF: Hot Spots



Hyperlink: Indexing TED Talks

Granularity level:

- Chapter



Features indexed :

- Topics
- Entities
 - *time code references (startNPT and endNPT)*
 - *extractor confidence*
- Resource identifier
- Full text transcript

Hyperlink: Indexing TED Talks



A video player interface showing a TED talk by Tim Berners-Lee. The video frame shows him from the chest up, wearing a blue shirt, gesturing with his right hand. The background is a dark stage with vertical blue light bars. A progress bar at the bottom left indicates the video is at 01:52. A button labeled "GENERATE HOTSPOTS" is visible at the bottom center. The HyperTED logo is in the top left corner of the video frame.

Search by topic

watch also

Tim Berners-Lee: The year open data went ...



- ▶ Chapter 11 (00:03:28 - 00:04:02)
- ▶ Chapter 3 (00:00:50 - 00:01:17)

Geraldine Hamilton: Body parts on a chip



- ▶ Chapter 15 (00:11:36 - 00:12:11)

Avi Rubin: All your devices can be hacked



- ▶ Chapter 19 (00:11:51 - 00:12:57)

Tan Le: A headset that reads your brainwa...



- ▶ Chapter 14 (00:08:56 - 00:09:38)

Hyperlink: Finding related Courses

Datasets:

- Open Courseware
- Open University



Anchors used in search:

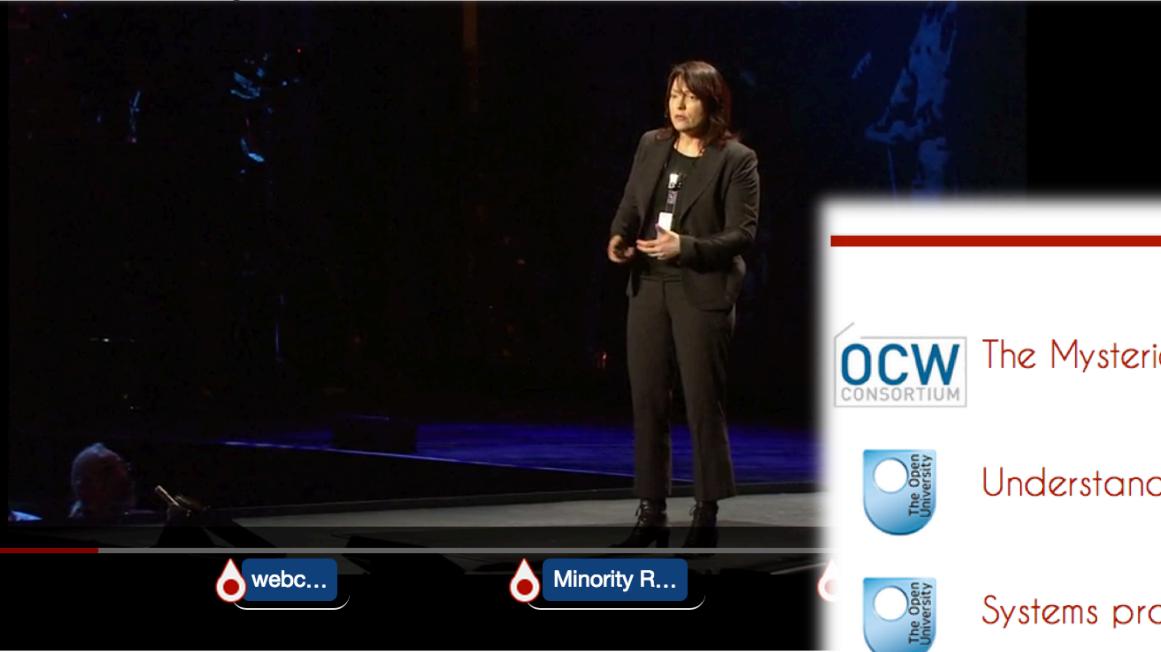
- Entities → Too specific.
- Topics → Courses about the same thematic

Attributes used in search:

- Title
- Description
- Subject, thematic ...

SPARQL is not an IR Language

< Hyperlink: Finding related Courses >



Search by topic Go!

learn also



The Mysterious Field of Engineering Systems

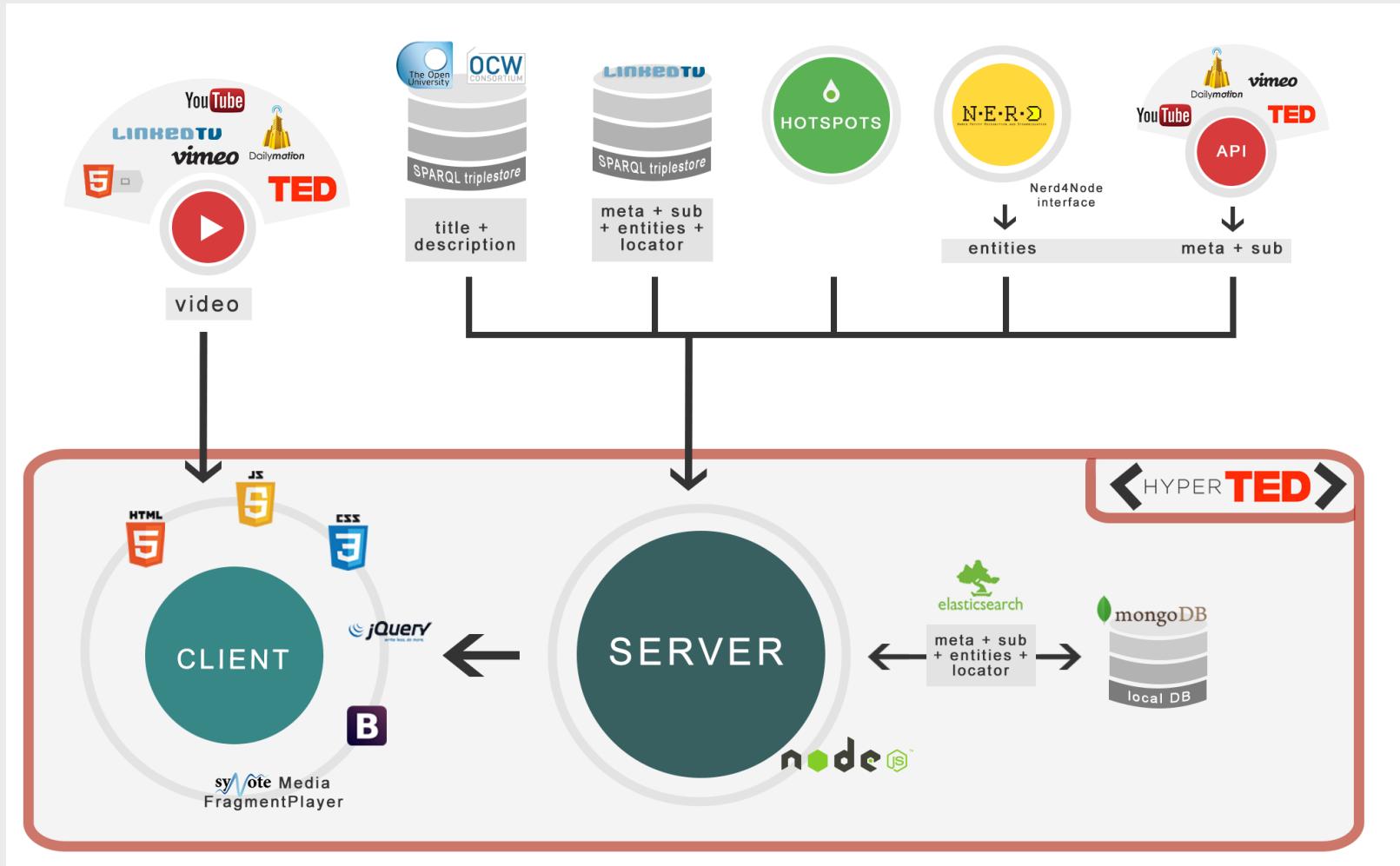


Understanding systems: a systems approach



Systems practice: Managing sustainability

Architecture



< DEMO time >

The screenshot shows a TED video player interface. At the top left is the 'HYPER TED' logo. The main title is 'Sally Kohn: Don't like clickbait? Don't click'. Below the title, there are details: 'Published 2014-08-28', 'Event TED@NYC', 'Views 426164', and 'Comments 57'. The video frame shows Sally Kohn speaking on stage. A play button icon is overlaid on the video. Below the video frame is a progress bar with a red scrubber, showing the current position at 03:16 and the total duration at 04:40. There are also 'tweeting' and 'Tweet' buttons. Below the progress bar, it says 'chapter 5 of 7'. To the right of the video frame, there is a text box containing a quote from Sally Kohn about clickbait.

published on 2014-08-28 from TED@NYC

Sally Kohn: Don't like clickbait? Don't click

Doesn't it seem like a lot of online news sites have moved beyond reporting the news to openly inciting your outrage (and your page views)? News analyst Sally Kohn suggests — don't engage with news that looks like it just wants to make you mad. Instead, give your precious clicks to the news sites you truly trust.

This screenshot shows a video player interface with a white background. At the top, there are two buttons: 'subtitles' on the left and 'watch also' on the right. Below the 'subtitles' button, there is a subtitle box containing text from Sally Kohn's speech. Below the 'watch also' button, there is a section titled 'Massimo Banzi: How Arduino is open-sourc...' with a thumbnail image of Massimo Banzi and a link to 'Chapter 9 (00:06:39 - 00:08:04)'. At the bottom of the interface, there is a navigation bar with several small video thumbnails and their corresponding chapter numbers.

<http://linkedtv.eurecom.fr/HyperTED>

Conclusions

- Media Fragment player: annotated fragments
- **Hot Spots** detection based on topics and entities
- Hyperlinking between **chapters** of TED talks + **external education resources**
- Responsive and Web compliant **UI**
- **Future:** formal evaluation of recommended content, better SPARQL queries and content indexing



Thank you!

Source Code: <https://github.com/jluisred/HotSpots>
<https://github.com/pasqLisena/MediaFragPlayerDemo>

José Luis Redondo García – redondo@eurecom.fr

GO!



@peputo

@rtroncy

@marSabs

@PasqLisena

Raphaël Troncy – raphael.troncy@eurecom.fr

Mariella Sabatino – mariella.sabatino@eurecom.fr

Pasquale Lisena – pasquale.lisena@eurecom.fr