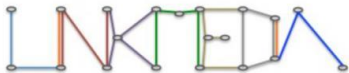


Leveraging topic models for video hyperlinking

in the context of the MediaEval and TRECVID benchmarking initiatives

Anca Şimon



Pascale Sébillot & Guillaume Gravier & Rémi Bois & Ronan Sicre



Sien Moens

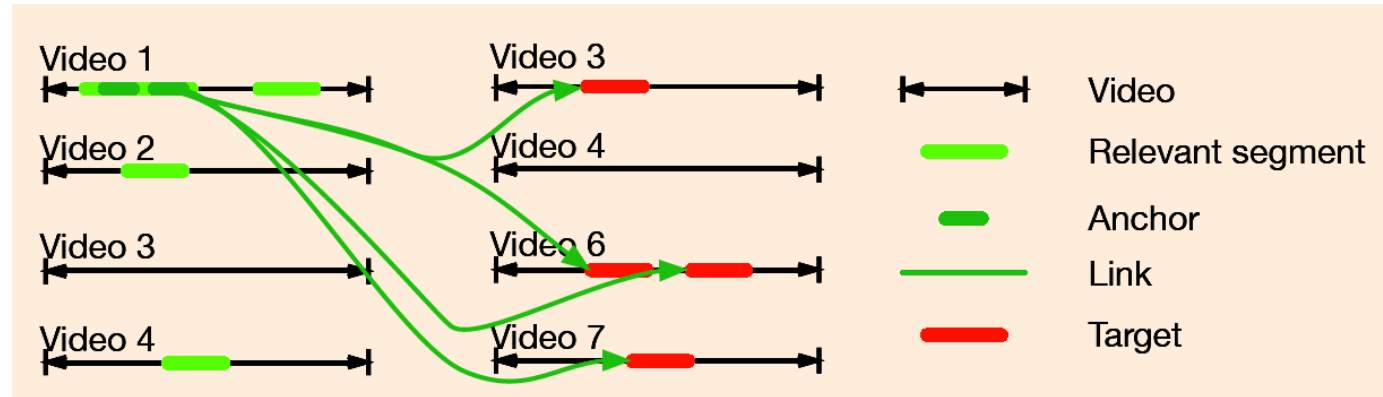


Video Hyperlinking

Use case

Text query

- speech cue
- visual cue



Beyond search,

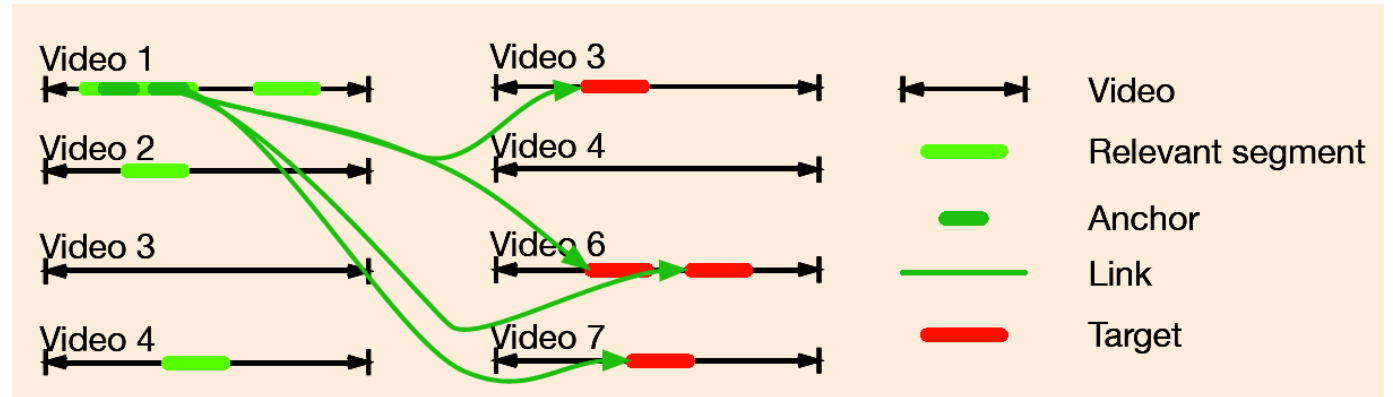
anchor detection + hyperlinking = organizing a collection
for analytics based on interaction with the data

Video Hyperlinking

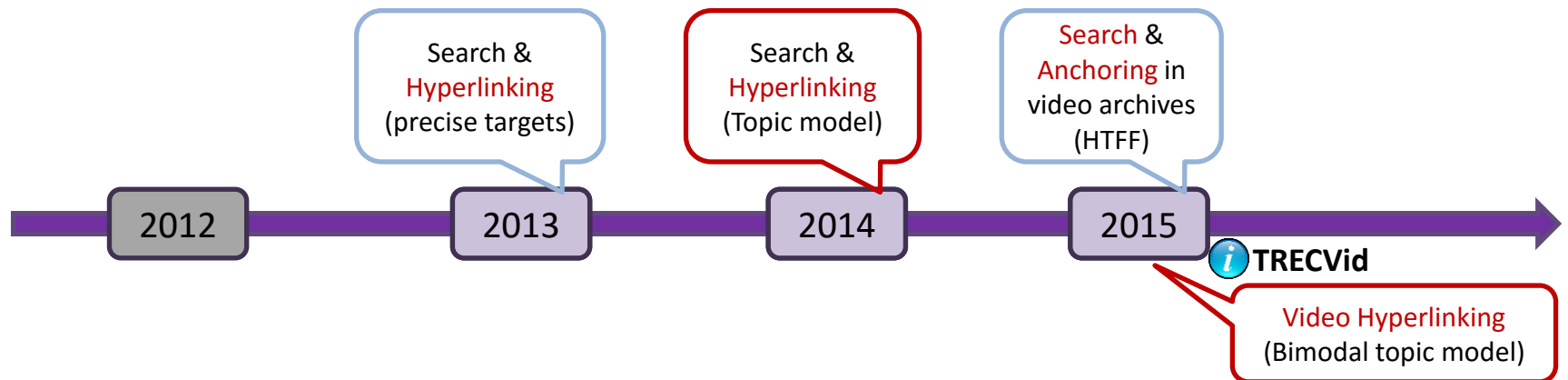
Use case

Text query

- speech cue
- visual cue



MediaEval benchmarking initiative: Search and Hyperlinking task



A search & hyperlinking scenario

Query

things to see in london



Search results:

<u>Video</u>	<u>Start</u>	<u>End</u>
video1	05:20	06:30
video2	03:00	04:45
...		
videon	12:30	15:00



London -10 Things You Need To Know - Hostelworld Video

by Hostelworld

6 years ago • 1,683,538 views

'Find out how to get around, save money and see all the best attractions. Book a Hostel in London today: <http://hwrlid.cm/1oJooe2> ...

CC



London 10 Quirky Places

by Chris Lawson

5 years ago • 776,411 views

An alternative sightseeing trip round London, taking a look at some of it's quirkiest sights. From the Mandella Tank to the Traffic ...

HD



Free Fun In London - what to do and where to go

by Julian Heald

3 years ago • 59,438 views

We put together this video to show you what can be seen in London for FREE. And there certainly is a lot! Museums, markets ...

HD



London, England Travel Guide - Must-See Attractions

by BookingHunterTV

1 year ago • 25,285 views

<http://bookinghunter.com> London is one of the world's most remarkable and exciting cities and has something to offer every type of ...

HD



Travel Tips : List of Top Things to See in London

by eHow

5 years ago • 45,211 views








When traveling to London, some of the top sightseeing attractions include Buckingham Palace, Westminster Abbey and the Tower ...

HD

A search & hyperlinking scenario



Recommened videos

-  Visit England-5 Things You Will Love& Hate About Visiting England (0:52)
-  Unusual Facts About London (11:06)
-  London Tour (19:44)
-  Sydney- 10 things you need to know (7:34)
-  Around the World in 156 Seconds (2:57)
-  How to save money in Paris (2:17)
-  10 Common Expressions in English (9:33)

Anchor

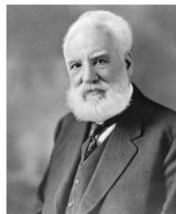
Targets

London- 10 Things You Have To Know

Details on demand



How it was made



Alexander Graham Bell biography



Fireworks in a phone booth



Doctor Who: 50 years traveling through space and time



Phone booth trailer



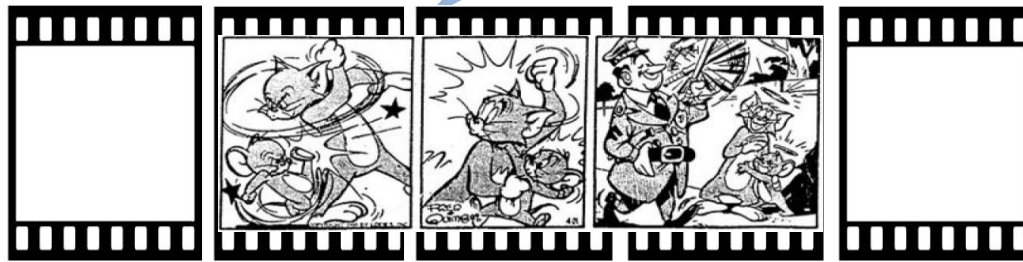
Disappearing London - Red Phone Box

An overview of the state of the art

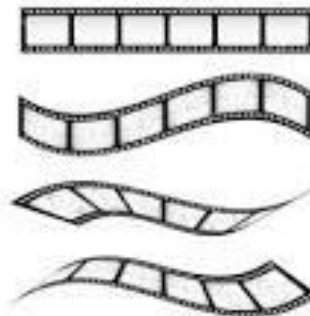
A two-step approach:

1. Segmentation

- Fixed-length segments
- Video shots
- Topic segments
- Utterances



Potential targets

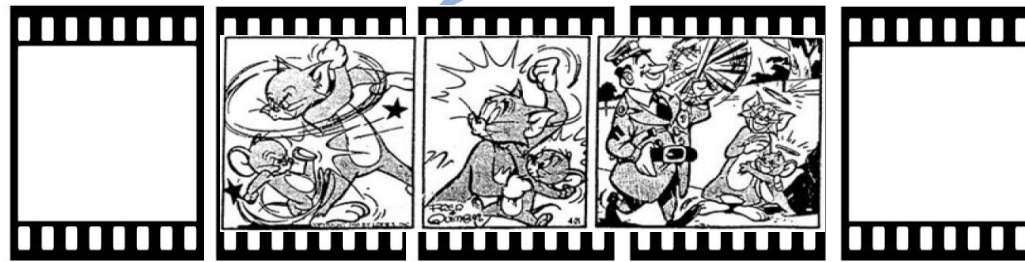


An overview of the state of the art

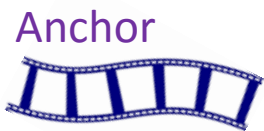
A two-step approach:

1. Segmentation

- Fixed-length segments
- Video shots
- Topic segments
- Utterances

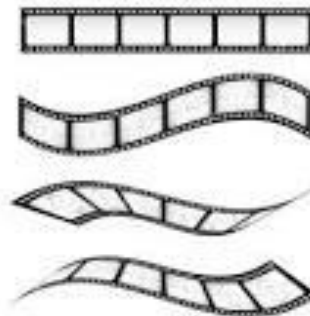


2. Target selection



comparison & selection

Potential targets



- Language via transcripts (entities, prosody)
- Visual content (concepts)
- Metadata

What about diversity?

Direct comparison in vector space with cosine similarity!

Targets very similar to the anchor

- near duplicates
- timeline events
- ... but no diversity and no serendipity

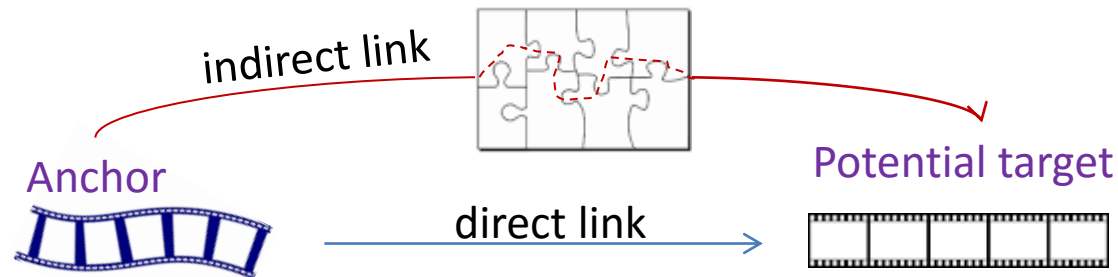
What about diversity?

Direct comparison in vector space with cosine similarity!

Targets very similar to the anchor

- near duplicates
- timeline events
- ... but no diversity and no serendipity

Solution: Indirect comparison



+ link anchor-target pairs with few words in common

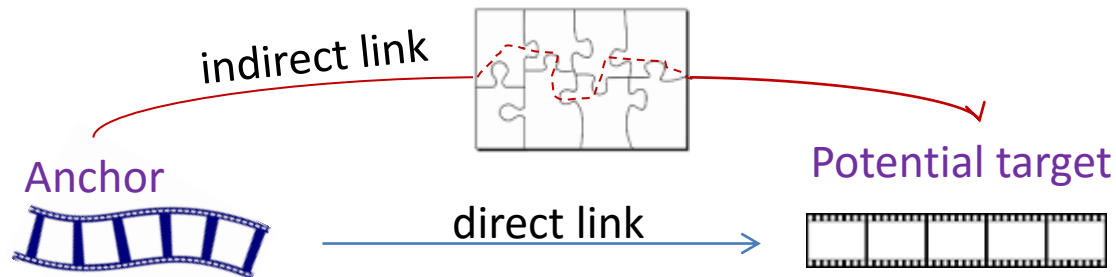
What about diversity?

Direct comparison in vector space with cosine similarity!

Targets very similar to the anchor

- near duplicates
- timeline events
- ... but no diversity and no serendipity

Solution 1: Indirect comparison via a **hierarchy of topic models**



- + link anchor-target pairs with few words in common
- + control diversity
- + link justification

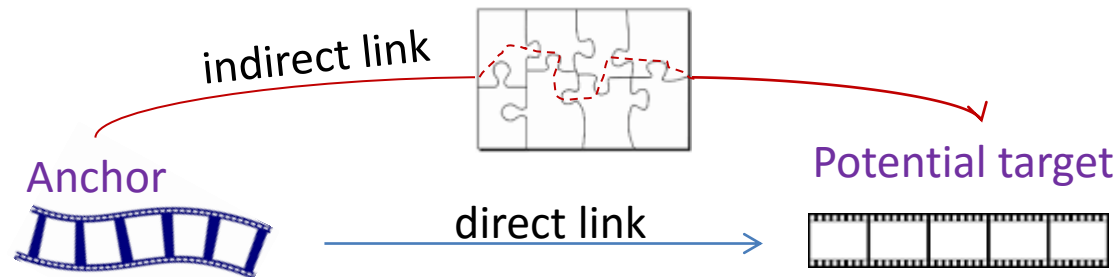
What about diversity?

Direct comparison in vector space with cosine similarity!

Targets very similar to the anchor

- near duplicates
- timeline events
- ... but no diversity and no serendipity

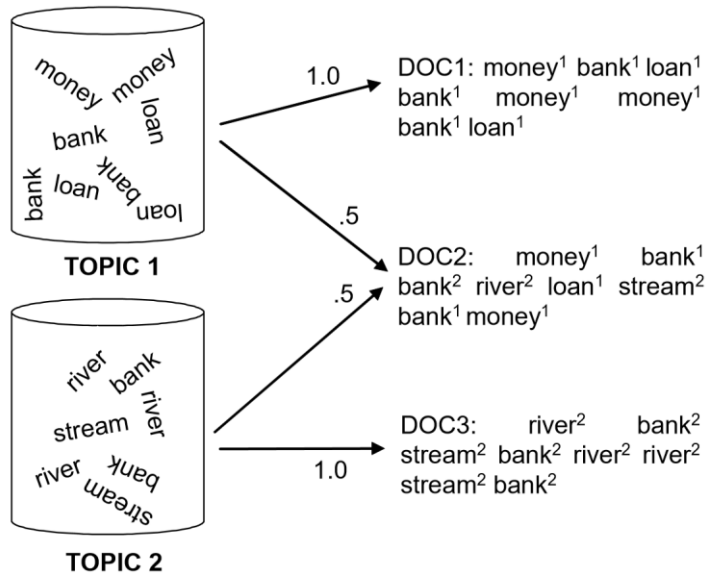
Solution 2: Indirect comparison via a **cross-modal topic models**



- + link anchor-target pairs with few words in common
- + control diversity
- + link justification
- + talk about what is shown or show things that are discussed

LDA model

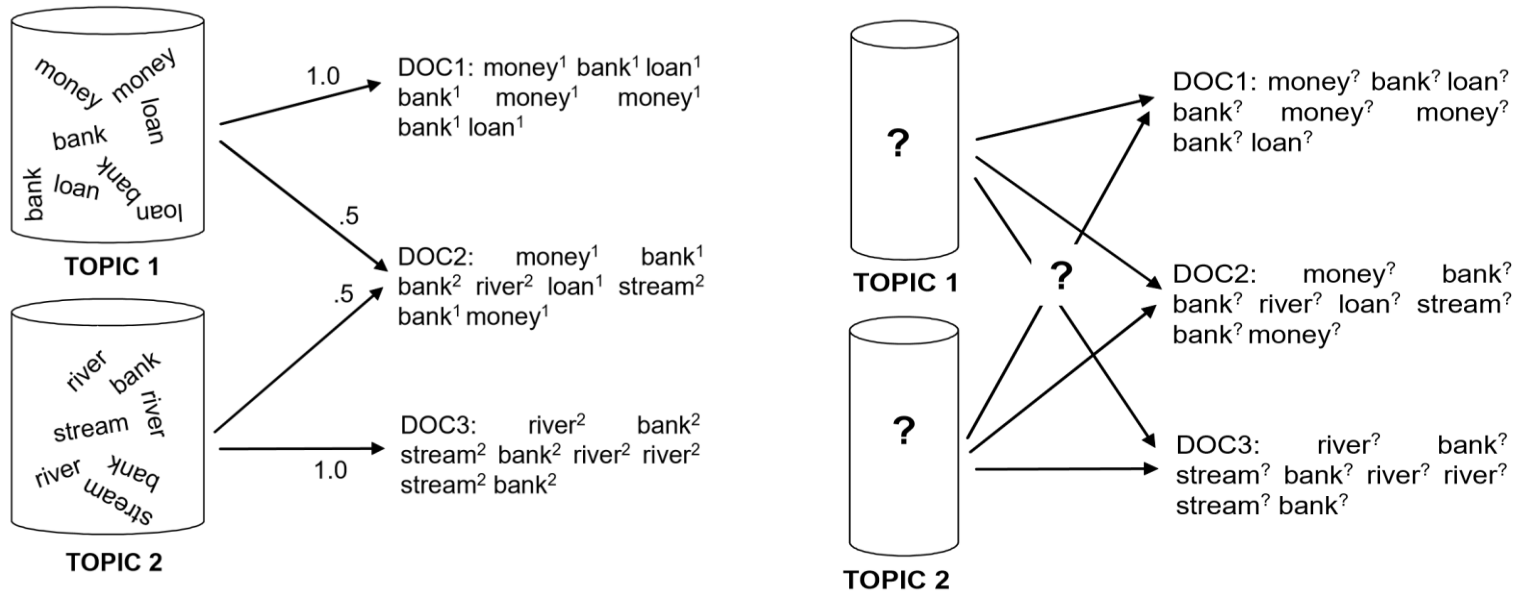
Key idea: there exist latent topics which uncover how words in documents have been generated



Steyvers and Griffiths, 2010

LDA model

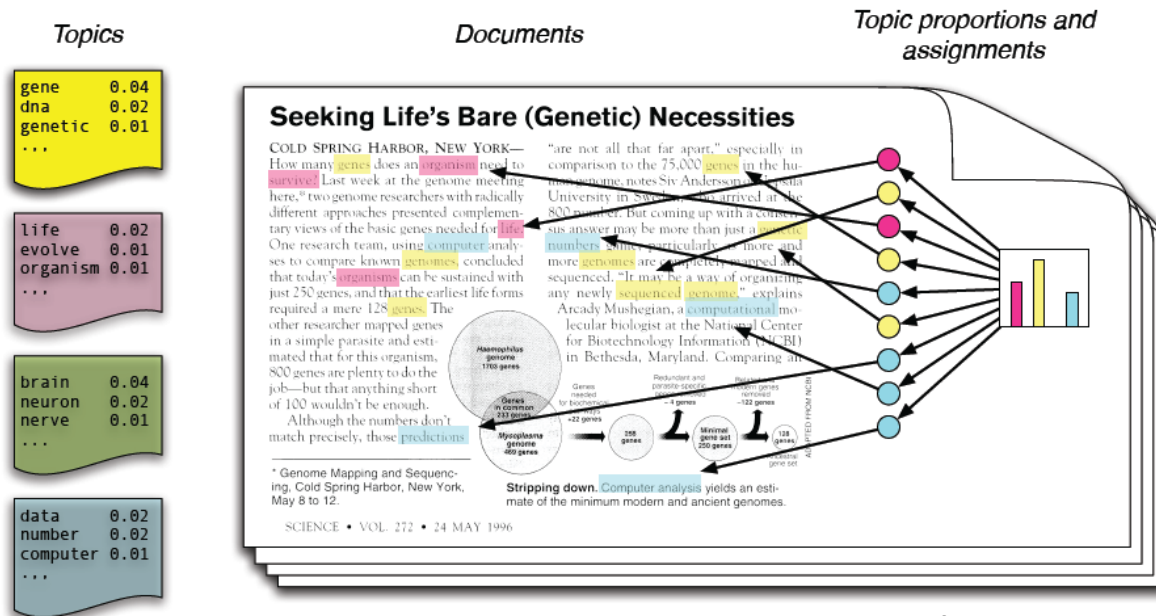
Key idea: there exist latent topics which uncover how words in documents have been generated



Steyvers and Griffiths, 2010

LDA model

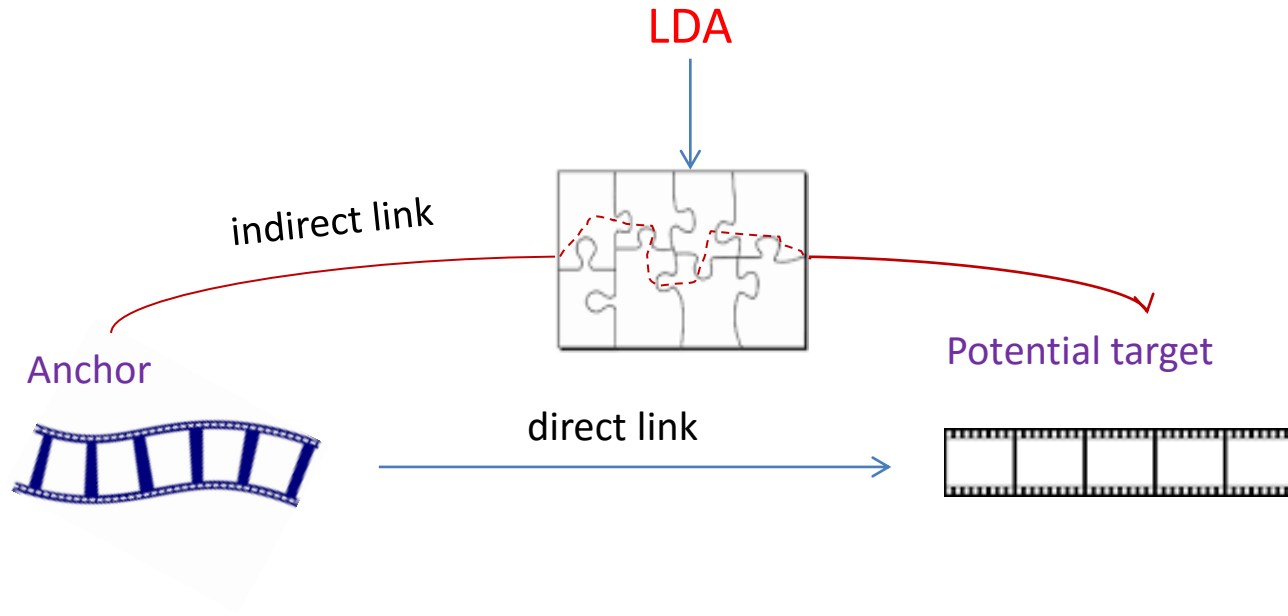
Key idea: there exist latent topics which uncover how words in documents have been generated



Blei, 2012

- Each topic: a probability distribution over words
- Each document: a mixture of topics

Indirect link



Leverage LDA for hyperlinking

Create a hierarchy of topics:

$$K \in \{50, 100, 150, 200, 300, 500, 700, 1000, 1500, 1700\}$$

- Level 1, $K_1 = 50$, broad topics $z_i^1, i \in [1, K_1]$
- Level 10, $K_{10} = 1700$, fine-grained topics $z_i^{10}, i \in [1, K_{10}]$

Leverage LDA for hyperlinking

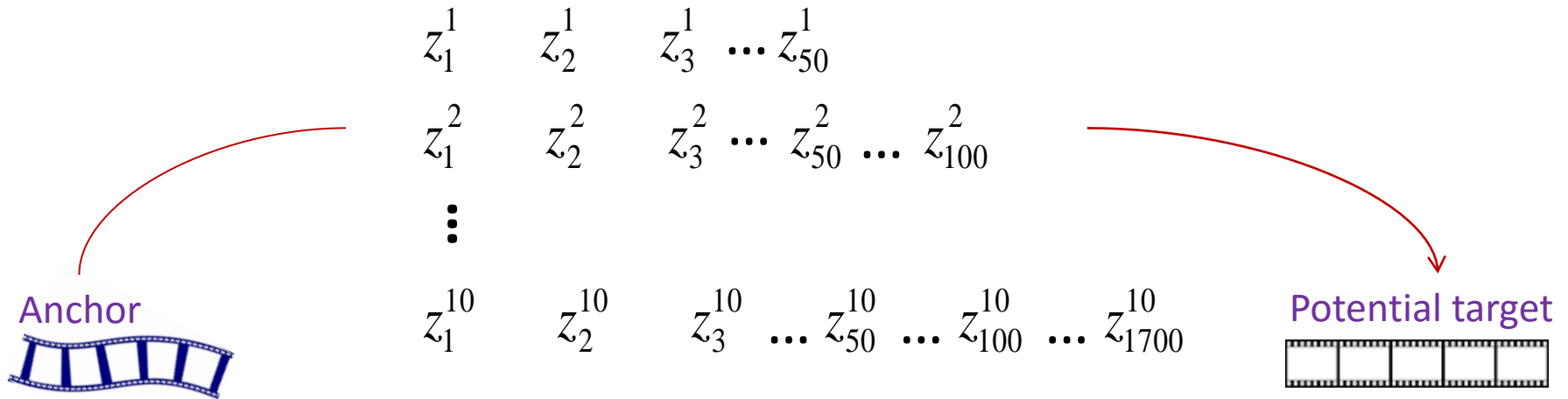
Create a hierarchy of topics:

$$K \in \{50, 100, 150, 200, 300, 500, 700, 1000, 1500, 1700\}$$

- Level 1, $K_1 = 50$, broad topics $z_i^1, i \in [1, K_1]$
- Level 10, $K_{10} = 1700$, fine-grained topics $z_i^{10}, i \in [1, K_{10}]$

broad	fine-grained						
$z_3^1, K_1=50$	$z_{50}^{10}, K_{10}=1700$	z_1^1	z_2^1	z_3^1	\dots	z_{50}^1	
People	Referendum	z_1^2	z_2^2	z_3^2	\dots	z_{50}^2	\dots
Government	Minister	z_1^2	z_2^2	z_3^2	\dots	z_{50}^2	\dots
Tax	Scotland	\vdots					
Minister	Independence						
Party	Alexander	z_1^{10}	z_2^{10}	z_3^{10}	\dots	z_{50}^{10}	\dots
							\dots
							z_{1700}^{10}

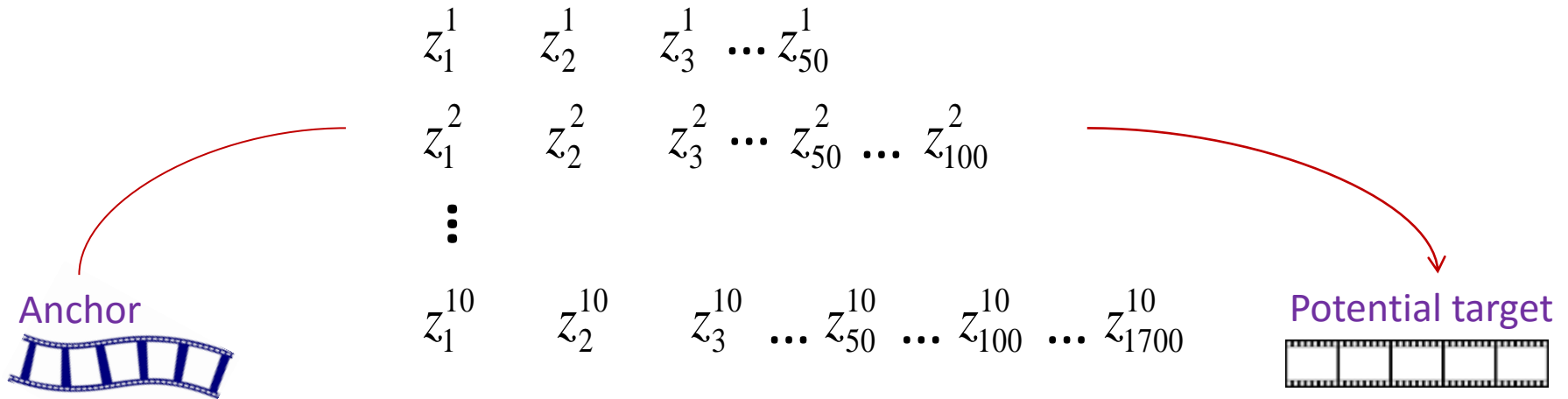
Changing the representation space



➤ New representation of an anchor/target segment

$$x_l = (p(x | z_1^l) \dots p(x | z_{K_l}^l))$$

Changing the representation space



➤ New representation of an anchor/target segment

$$x_l = (p(x | z_1^l) \dots p(x | z_{K_l}^l))$$

➤ **1st strategy: independent topic levels (IT)**

➤ **2nd strategy: hard and soft links between topics**

Independent levels

➤ Anchor segment x $x_l = (p(x | z_1^l) \dots p(x | z_{K_l}^l))$

➤ Target segment y $y_l = (p(y | z_1^l) \dots p(y | z_{K_l}^l))$

$$\textit{Similarity}(x, y) = \sum_l \alpha_l \log(x_l \cdot y_l)$$

IT_k only level k $\alpha_k = 1, \alpha_{i \neq k} = 0$

$\text{IT}_=$ equal weights $\alpha_k = 0.2, \forall k \in \{1, 3, 5, 7, 9\}$

$\text{IT}_<$ general < specific $\alpha_1 = 0.1, \alpha_3 = 0.15, \alpha_5 = 0.2, \alpha_7 = 0.25, \alpha_9 = 0.3$

$\text{IT}_>$ specific < general $\alpha_1 = 0.3, \alpha_3 = 0.25, \alpha_5 = 0.2, \alpha_7 = 0.15, \alpha_9 = 0.1$

Data

2013 & 2014 Search & Hyperlinking data

- BBC broadcast videos
- automatic speech transcripts (LIMSI)

Task considered: reranking targets

- Targets proposed by all the participants!
- Relevance judgments provided by turkers (AMT)

year	#hours of video	#anchors	avg. anchor duration (95% interval)	#targets (% relevant)	avg. target duration (95%interval)
2013	1,335	30	32.2 [13.4,51]	9,973 (29.9%)	83.38 sec. [82.58,84.18]
2014	2,686	30	22.9 [11.1,34.8]	12,340 (15.3%)	58.85 sec. [58.1,59.58]

Watch 2 video segments and say whether the second video is related to the first one according to the given description

Please first follow the instructions on the left and then answer the questions on the right side of the screen.

1) Please watch the *first video clip* shown below.



2) Imagine a person watched this *first video clip* on a site like YouTube and wishes to see more video clips with the following *description*:

I would like to watch more mafia clips; or something about links between mafia and other singers/famous people.

3) Please watch the following *second video clip* to see whether it satisfies the wish of the person.



4) Based on the *description*, would the person be satisfied watching the *second video clip* after having watched the *first video clip*?

Yes No

5) Please write 1-3 sentences in the box below that explain your decision.

6) Please write 3-5 meaningful words spoken in each of the video clip.

first video clip

second video clip

NOTES: Please note that in doing this HIT you are taking part in an academic research study. Our review process involves many manual steps. We are also a small team. For this reason, there might be a delay in the approval of your work. We do our best to keep this delay to 2-3 days at the very maximum.

NOTES: It is important that before you submit the HIT you take one more look at the answer that you provided. We ask you to double check that you have written 2-3 complete sentences and that your grammar is OK. We also ask you to check to make sure that the relationship between your sentences and the videos themselves is very clear.

When you are finished with answering the questions, don't forget to click the "Submit" button at the bottom of the page.

Thank you very much for your work!

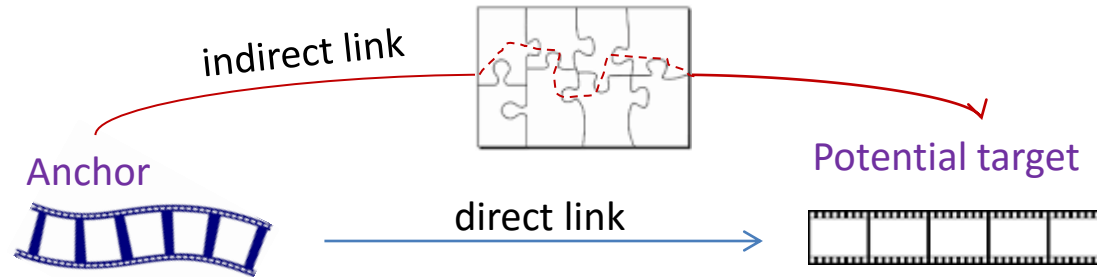
Relevance assessment

- Baseline: direct cos-similarity (DirectH)
- Measures: relevance (P@10);
tolerance to irrelevance (P@10_tol)

	2013		2014	
method	P@10	P@10_tol	P@10	P@10_tol
DirectH	0.61	0.25	0.41	0.19
IT_{50}	0.65	0.44*	0.26	0.18
IT_{150}	0.57	0.34*	0.37	0.25*
IT_{300}	0.61	0.35*	0.34	0.26*
IT_{700}	0.64	0.34*	0.31	0.21
IT_{1500}	0.59	0.32*	0.32	0.24
$IT_{Comb=}$	0.66	0.35*	0.27	0.22
$IT_{Comb<}$	0.67	0.37*	0.27	0.21
$IT_{Comb>}$	0.65	0.35*	0.29	0.22

* Statistical significant values (paired t-test, $p < 0.05$)

Indirect linking



Solution 1: Indirect comparison via a **hierarchy of topic models**

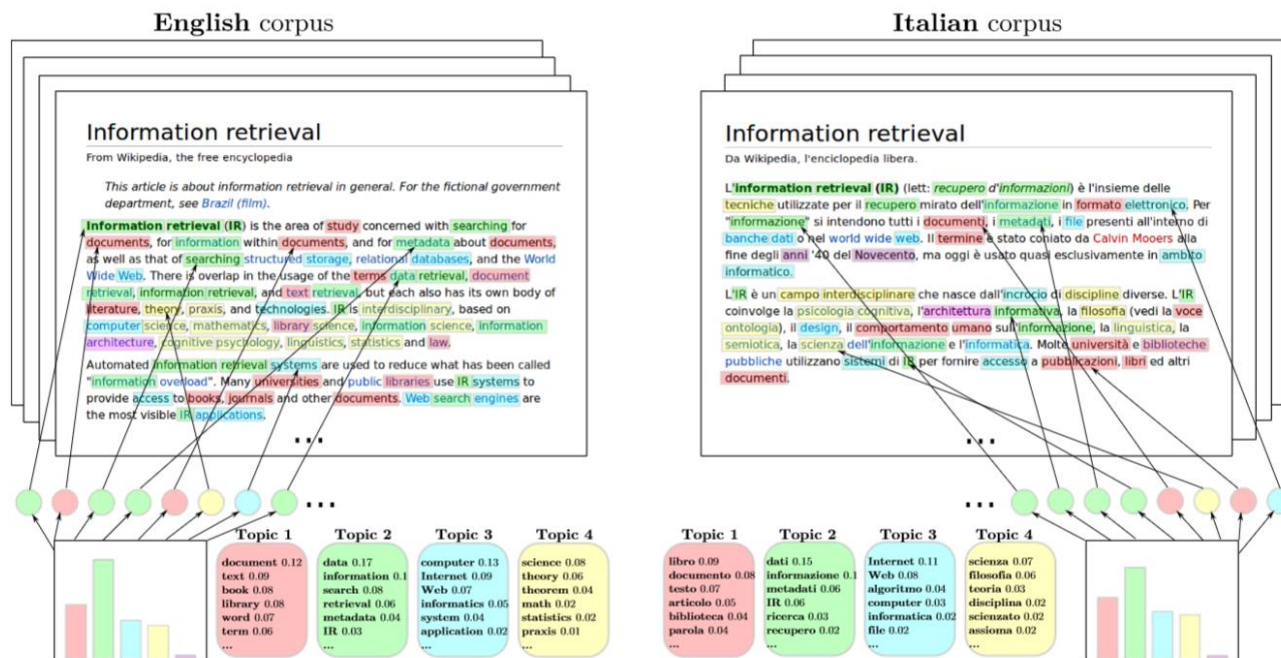
- + link anchor-target pairs with few words in common
- + control diversity
- + link justification

Solution 2: Indirect comparison via a **cross-modal topic models**

- + link anchor-target pairs with few words in common
- + control diversity
- + link justification
- + talk about what is shown or show things that are discussed

Bilingual LDA model

Key idea: discover the latent cross-lingual topics that describe a given bilingual document collection

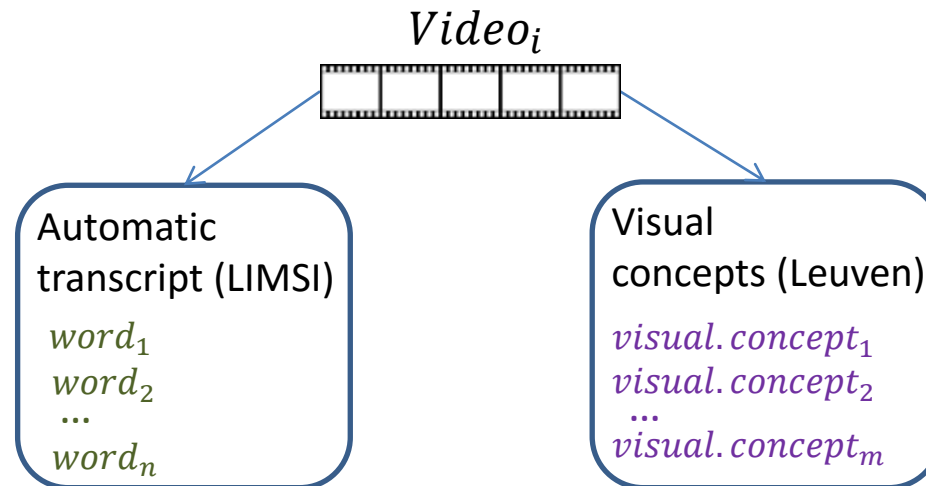


Vulić et al., 2014

- Each pair of comparable documents share the same distribution of topics
- Each topic is modeled as a distribution over vocabulary words in each language

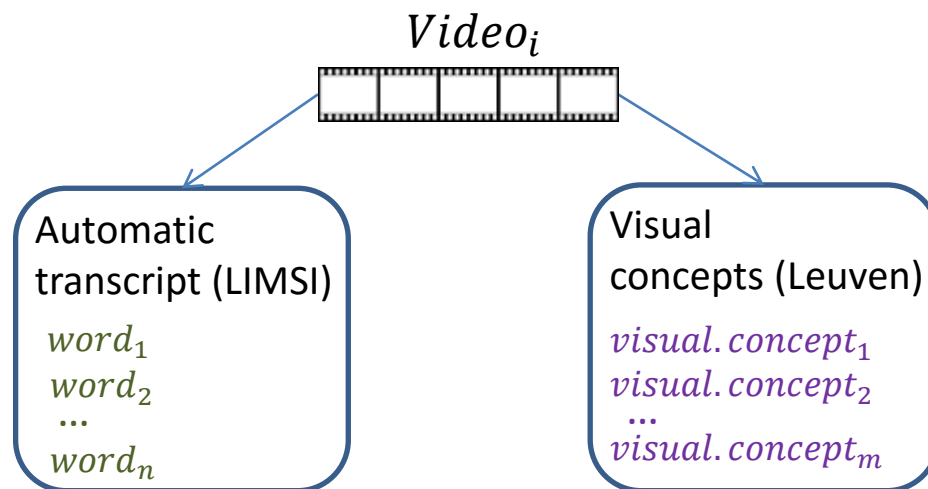
Leverage bimodal LDA for video hyperlinking

- ✓ We use **audio** and **visual information** as two different languages and build cross-modal topics



Leverage bimodal LDA for video hyperlinking

✓ We use **audio** and **visual information** as two different languages and build cross-modal topics

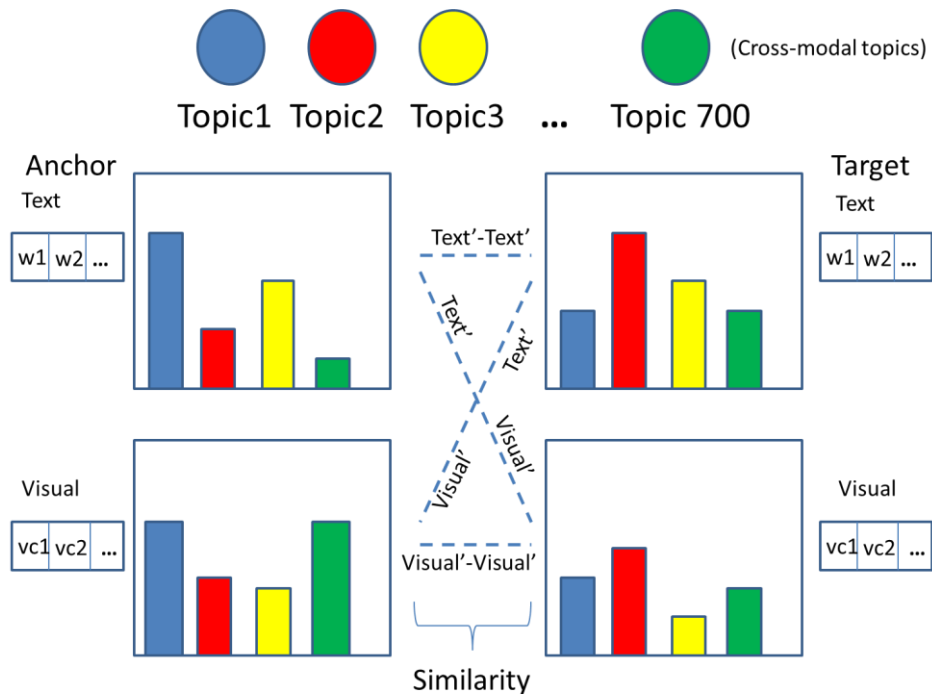


$K=700$	Audio	Visual
Topic 3	love, home, feel, life, baby	singer, microphone, sax, concert, flute
Topic 7	food, bit, chef, cook, kitchen	fig, acorn, pumpkin, guava, zucchini
Topic 25	years, technology, computer key, future	tape-player, computer, equipment, machine, appliance

Leverage bimodal LDA for video hyperlinking

- By learning the cross-modal topics, we can apply
 - ✓ the usual topic similarities (i.e. audio → audio or visual → visual)
 - ✓ cross-modality similarities (i.e. audio → visual or visual → audio):

seeing more about what is said and hearing more about what is shown



Relevance assessment

1) Reranking targets proposed by all participants in 2014

Method	Audio->Audio	Audio->Visual	Visual->Visual	Visual->Audio
P@10	25.3	21	30	24

2) TRECVID results in 2015

(100 anchors)	Minimum	25%	50%	75%	Maximum
P@10	0.017	0.198	0.275	0.524	0.608
Direct Visual similarity	0.207				
Visual->Audio	0.224				

Diversity assessment

Success of a hyperlinking system:

cover potential (idiosyncratic) user interest & enable serendipity

Solution 1

Links differ between systems

System 1	System 2	% difference	
		2013	2014
IT_{700}	<i>DirectH</i>	93	86
IT_{700}	$IT_{Comb>}$	82	90
IT_{700}	<i>Hierarchy</i>	98	93
$IT_{Comb=}$	<i>Hierarchy</i>	94	95

Solution 2

Cross-modal topics

- Share <7.4% of top 10 targets

Direct Visual Vs. with Visual->Visual

- Share 30.3% of top 10 targets

Diversity assessment

Success of a hyperlinking system:

cover potential (idiosyncratic) user interest & enable serendipity

Solution 1

Links differ between systems

System 1	System 2	% difference	
		2013	2014
IT_{700}	<i>DirectH</i>	93	86
IT_{700}	$IT_{Comb>}$	82	90
IT_{700}	<i>Hierarchy</i>	98	93
$IT_{Comb=}$	<i>Hierarchy</i>	94	95

Solution 2

Cross-modal topics

- Share <7.4% of top 10 targets

Direct Visual Vs. with Visual->Visual

- Share 30.3% of top 10 targets

AMT evaluation scenario at MediaEval

- 1 judgement/anchor-target pair
- yes/no relevance assessment
- description of potential targets

Diversity in the links

Design a new evaluation scenario:

- At least 3 assessments per anchor-target pair
- Each participant should do 5 tests
- Test for: **relevance** (same topic, related topic, same show);
unexpectedness;
interestingness;

Clip A

Anchor:



Two video clips (B and C) that could be linked to video A are recommended to you that should encourage this further exploration. Please watch the two videos and answer the questions.

Clip B

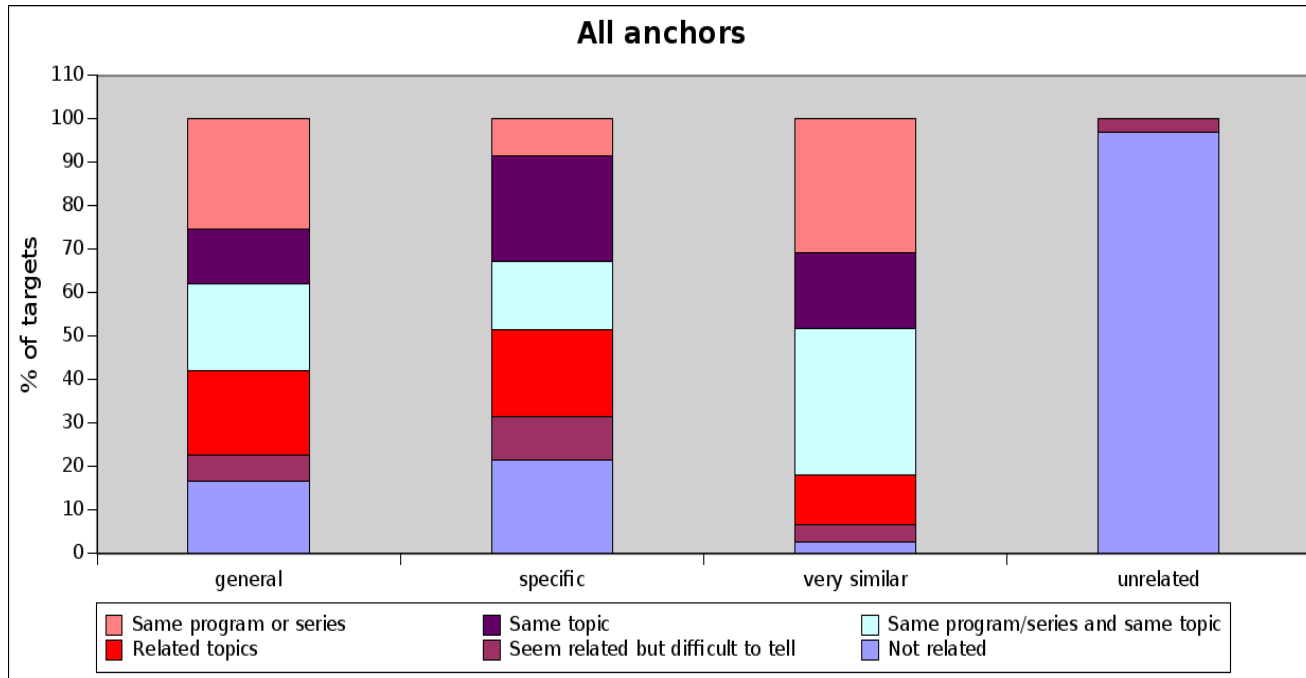


Clip C



Targets:

Results for the new scenario



➤ Very similar targets:

- same program/series and same topic (91% expected; 9% possibly)
- most expected

➤ Specific topics:

- same topic (47% expected; 53% possibly)
- less expected

Lessons learned

➤ From taking part in the challenges:

- ✓ Evaluation is challenging
(resource constraints; subjectivity of the task)
- ✓ Easy to score points with very similar targets (near duplicates)
- ✓ Yes/No relevance assessment is not enough
- ✓ One judgment per anchor-target pair is not enough
- ✓ Each year it improves based on the feedback from participants

➤ From the survey evaluation:

- ✓ Large disagreement between participants
- ✓ The task should not take a lot of time
- ✓ Difficult to define questions about the topical relations

➤ From using topic models:

- ✓ Increase diversity
- ✓ Offer more control over link creation and justification
- ✓ Cross-modal topics don't work on some anchors

Perspectives

✓ User-centric evaluation

- ✓ Diverse targets to evaluate for the same anchor
 - > user can choose the type of target to follow on
- ✓ Add link justification
 - > this link is proposed because...

- ✓ Improve/refine the models proposed
- ✓ Use a hierarchy of cross-modal topics
- ✓ Design a survey that evaluates the translation between modalities