

IT Systems Engineering | Universität Potsdam



Nadine Steinmetz, Harald Sack

10th Extended Semantic Web Conference Montpellier, France May 30th, 2013





- 1. Motivation Semantic Search on Videos
- 2. Challenges of Video Metadata
- 3. Contextual Model
- 4. Evaluation
- 5. Outlook





#### 1. Motivation - Semantic Search on Videos

- 2. Challenges of Video Metadata
- 3. Contextual Model
- 4. Evaluation
- 5. Outlook





## Semantic Search on Videos

To enable content-based, semantic search on videos:

• authoritative metadata

- non-authoritative metadata (e.g. tags)
- content-related, time-referenced metadata
  - content-based segmentation
  - OCR Optical Character Recognition
  - ASR Automated Speech Recognition
- semantic analysis of textual metadata







By Ratiocination, I mean **computation**. Now to compute, is either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another... and if any man adde Multiplication and Division, I will not be against it, seeing Multiplication is nothing but Addition of equals one to another, and Division nothing but a Substraction of equals one from another, as often as is possible. So that all Ratiocination is comprehended in these two operations of the minde, Addition and Substraction...

- Thomas Hobbes, 1656

(i) 06:28 / 18:37

Nadine Steinmetz - Semantic Multimedia Information Retrieval Using Contextual Descriptions, ESWC 2013, 30-05-2013



#### Title: The birth of the computer Speaker: George Dyson



By Ratiocination, I mean computation. Now to compute, is either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another ... and if any man adde Multiplication and Division, I will not be against it, seeing Multiplication is nothing but Addition of equals one to another, and Division nothing but a Substraction of equals one from another, as often as is possible. So that all Ratiocination is comprehended in these two operations of the minde, Addition and Substraction ...

- Thomas Hobbes, 1656

(18:37) 06:28 / 18:37





#### Title: The birth of the computer Speaker: George Dyson



5

hobbes computation thomas 1651

By Ratiocination, I mean **computation**. Now to compute, is either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another... and if any man adde Multiplication and Division, I will not be against it, seeing Multiplication is nothing but Addition of equals one to another, and Division nothing but a Substraction of equals one from another, as often as is possible. So that all Ratiocination is comprehended in these two operations of the minde, Addition and Substraction...

Thomas Hobbes, 1656

Authoritative Metadata Non-Authoritative Metadata

æ



#### Title: The birth of the computer Speaker: George Dyson



Authoritative Metadata Non-Authoritative Metadata OCR Metadata



#### Title: The birth of the computer Speaker: George Dyson



... was Thomas Hobbes who in 1651 explained how arithmetic and logic are the same thing ...



#### Title: The birth of the computer Speaker: George Dyson



... was Thomas Hobbes who in 1651 explained how arithmetic and logic are the same thing ...







- 1. Motivation Semantic Search on Videos
- 2. Challenges of Video Metadata
- 3. Contextual Model
- 4. Evaluation
- 5. Outlook





minimum of manual effort





# Challenges of Analyzing Video Metadata

Automated Text Extraction

Advantage

- time-referenced metadata
- minimum of manual effort

By Ratiocination, I mean **computation**. Now to compute, is either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another... and if any man adde Multiplication and Division, I will not be against it, seeing Multiplication is nothing but Addition of equals one to another, and Division nothing but a Substraction of equals one from another, as often as is possible. So that all Ratiocination is comprehended in these two operations of the minde, Addition and Substraction...

- Thomas Hobbes, 1656

#### Disadvantage

error rates!



# Challenges of Analyzing Video Metadata

Automated Text Extraction

Advantage

- time-referenced metadata
- minimum of manual effort



Disadvantage

error rates!



# Challenges of Analyzing Video Metadata





Challenges:

- context-driven disambiguation of textual information
- heterogenous context
  - different sources different reliabilities
  - different text types
  - dragging of error rates





- 1. Motivation Semantic Search on Videos
- 2. Challenges of Video Metadata
- 3. Contextual Model
- 4. Evaluation
- 5. Outlook





# Disambiguation within a Context

#### Title: The birth of the computer Speaker: George Dyson

By Ratiocination, I mean **computation**. Now to compute, is either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another... and if any man adde Multiplication and Division, I will not be against it, seeing Multiplication is nothing but Addition of equals one to another, and Division nothing but a Substraction of equals one from another, as often as is possible. So that all Ratiocination is comprehended in these two operations of the minde, Addition and Substraction...

- Thomas Hobbes, 1656

#### 🕪) 06:28 / 18:37

... was Thomas Hobbes who in 1651 explained how arithmetic and logic are the same thing ...



# Disambiguation within a Context

#### Title: The birth of the computer Speaker: George Dyson

By Ratiocination, I mean **computation**. Now to compute, is either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another... and if any man adde Multiplication and Division, I will not be against it, seeing Multiplication is nothing but Addition of equals one to another, and Division nothing but a Substraction of equals one from another, as often as is possible. So that all Ratiocination is comprehended in these two operations of the minde, Addition and Substraction...

- Thomas Hobbes, 1656

... was Thomas Hobbes who in 1651 explained he arithmetic and logic are the same thing ...





# Disambiguation within a Context

#### Title: The birth of the computer Speaker: George Dyson

By Ratiocination, I mean **computation**. Now to compute, is either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another... and if any man adde Multiplication and Division, I will not be against it, seeing Multiplication is nothing but Addition of equals one to another, and Division nothing but a Substraction of equals one from another, as often as is possible. So that all Ratiocination is comprehended in these two operations of the minde, Addition and Substraction...

- Thomas Hobbes, 1656

(1) 06:28 / 18:37

... was Thomas Hobbes who in 1651 explained he arithmetic and logic are the same thing ...





# **Contextual Descriptions - Approach**

- 12
- Scoring of Metadata Items
- Calculation of a **Confidence Value (0.0...1.0)**, consisting of:
  - Source Reliability
  - Source Diversity
  - Text Type
  - Class Cardinality
- Disambiguating the Metadata items in order according to confidence value



**Context Matters** 



#### Source Reliability

13

(Textual) Metadata Sources for Video Documents:

- Authoritative Sources: Person, who uploaded or created video
- Non-Authoritative Sources: Human, but not Authors, e.g. tags
- ASR: Automated Speech Recognition
- OCR: Optical Character Recognition

### Source Reliability



HPI

- 13
  - (Textual) Metadata Sources for Video Documents:
    - Authoritative Sources: Person, who uploaded or created video
    - Non-Authoritative Sources: Human, but not Authors, e.g. tags
    - **ASR: Automated Speech Recognition**
    - **OCR: Optical Character Recognition**



Hasso



Subsurface

#### **Source Diversity**



By Ratiocination, I mean **computation**. Now to compute, is either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another... and if any man adde Multiplication and Division, I will not be against it, seeing Multiplication is nothing but Addition of equals one to another, and Division nothing but a Substraction of equals one from another, as often as is possible. So that all Ratiocination is comprehended in these two operations of the minde, Addition and Substraction...

- Thomas Hobbes, 1656

Nadine Steinmetz - Semantic Multimedia Information Retrieval Using Contextual Descriptions, ESWC 2013, 30-05-2013

OCR

Substraction ...























- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR
- tags



- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR
- tags

15

Text processing goal: Identify key terms within the texts



3 different text types within video metadata:

- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR
- tags

Text processing goal: Identify key terms within the texts





3 different text types within video metadata:

- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR
- tags

Text processing goal: Identify key terms within the texts





3 different text types within video metadata:

- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR
- tags

Text processing goal: Identify key terms within the texts



natural language text



- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR
- tags





- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR
- tags





- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR
- tags

   Text processing goal: Identify key terms within the texts

   keyterms

   1.0

   natural language text
   POS tagging accuracy rate: 0.56



- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR
- tags

   Text processing goal: Identify key terms within the texts

   keyterms

   1.0

   natural language text
   POS tagging accuracy rate: 0.56



- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR
  - tags Text processing goal: Identify key terms within the texts keyterms 1.0 0.56 POS tagging - accuracy rate: 0.56 natural language text hobbes tags computation thomas 1651



- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR



- keyterms speaker information, production place, keywords
- natural language text descriptive texts, title, ASR, OCR





16

Some metadata items can be determined

to be an instance of a distinct ontology class, e.g.:

- Speaker information Person
- Film location Place
- NER tagger on natural language text: Person, Place or Organisation





16

Some metadata items can be determined

to be an instance of a distinct ontology class, e.g.:

- Speaker information Person
- Film location Plac
- NER tagger on natu



root class of ontology / used knowledge base





16

Some metadata items can be determined

to be an instance of a distinct ontology class, e.g.:

- Speaker information Person
- Film locatio
- ... was/O Thomas/PERSON Hobbes/PERSON who/O in/O
- NER tagger 1651/DATE explained/O how/O arithmetic/O and/O logic/ O are/O the/O same/O thing/O ...



root class of ontology / used knowledge base

assigned class





16

Some metadata items can be determined

to be an instance of a distinct ontology class, e.g.:

- Speaker information Person
- Film location Place
- NER tagger on natural language text: Person, Place or Organisation





### **Confidence** calculation



17



### **Confidence calculation**

	birth	Authoritative: Title	0.36
	computer	Authoritative: Title	0.46
	George Dyson	Authoritative: Speaker	0.72
	machine	OCR	0.33
Ö	Subsurface	OCR	0.22
	Thomas Hobbes	ASR	0.58



### **Confidence** calculation





# **Disambiguation process**

#### confidence value

CONTEXT	George Dyson	0.72
	Thomas Hobbes	0.58
	computer	0.46
	birth	0.36
	machine	0.33
	Subsurface	0.22



		confidence value	Entity with highest Disambiguation scor	e
	George Dyson	0.72	dbp:George_Dyson	1.0
2	Thomas Hobbes	0.58	dbp:Thomas_Hobbes	1.0
XI.	computer	0.46	dbp:Personal_computer	1.0
LNO	birth	0.36	dbp:Birth	0.1
Ö	machine	0.33	dbp:Machine	0.2
	Subsurface	0.22	dbp:Subsurface	0.0





- 1. Motivation Semantic Search on Videos
- 2. Challenges of Video Metadata
- 3. Contextual Model
- 4. Evaluation
- 5. Outlook





# **Evaluation Data Set**

Video Metadata of 5 TED talks

- 211 content-based segments
- 822 Metadata Items
- 1 1000 words per Metadata item
- 2550 entities identified





#### Download @ http://bit.ly/15a1YC0



# **Evaluation Data Set**

Video Metadata of 5 TED talks

- 211 content-based segments
- 822 Metadata Items

158

155

3447

12360

63205

• 1 - 1000 words per Metadata item

authorative/Speaker

authorative/Category

authorative/Title

• 2550 entities identified

asrtext

ocrtext





#### Download @ http://bit.ly/15a1YC0



# **Evaluation Data Set**

Video Metadata of 5 TED talks

- 211 content-based segments
- 822 Metadata Items

158

155

3447

12360

63205

• 1 - 1000 words per Metadata item

authorative/Speaker

authorative/Category

authorative/Title

• 2550 entities identified

asrtext

ocrtext





5851 0 http://dbpedia.org/resource/Psychology authorative/Category

Download @ http://bit.ly/15a1YC0



	conTagger			Simple NER			Wiki Machine			Spotlight		
	R	Р	F1	R	Р	F1	R	Р	F1	R	Р	F1
Authoritative	60,0	54,5	57,0	52,0	46,0	49,0	59,5	56,5	58,0	50,0	44,0	47,0
Tags	71,0	69,5	70,0	61,0	60,0	60,5	44,0	62,0	51,5	60,0	59,0	59,5
ASR	55,0	61,0	58,0	56,5	38,0	45,5	61,5	50,0	55,0	56,0	34,0	42,5
OCR	56,0	24,0	34,0	44,0	17,5	25,0	24,5	18,0	21,0	47,0	18,0	26,0
Segments	54,0	58,0	56,0	57,0	39,0	46,5	57,0	49,0	52,5	59,0	35,0	43,5
Video	56,0	48,0	52,0	57,0	30,0	39,5	58,0	43,0	49,5	54,0	31,0	39,5



	conTagger			Simple NER			Wiki Machine			Spotlight		
	R	Р	F1	R	Р	F1	R	Р	F1	R	Р	F1
Authoritative	60,0	54,5	57,0	52,0	46,0	49,0	59,5	56,5	58,0	50,0	44,0	47,0
Tags	71,0	69,5	70,0	61,0	60,0	60,5	44,0	62,0	51,5	60,0	59,0	59,5
ASR	55,0	61,0	58,0	56,5	38,0	45,5	61,5	50,0	55,0	56,0	34,0	42,5
OCR	56,0	24,0	34,0	44,0	17,5	25,0	24,5	18,0	21,0	47,0	18,0	26,0
Segments	54,0	58,0	56,0	57,0	39,0	46,5	57,0	49,0	52,5	59,0	35,0	43,5
Video	56,0	48,0	52,0	57,0	30,0	39,5	58,0	43,0	49,5	54,0	31,0	39,5



	conTagger		Simple NER			Wiki Machine			Spotlight			
	R	Р	F1	R	Р	F1	R	Р	F1	R	Р	F1
Authoritative	60,0	54,5	57,0	52,0	46,0	49,0	59,5	56 -		50,0	44,0	47,0
Tags	71,0	69,5	70,0	61,0	60,0	60,5	41	c to	iano!	60,0	59,0	59,5
ASR	55,0	61,0	58,0	56,5	38,0	45,	Thank	dio Gil	JIIO	56,0	34,0	42,5
OCR	56,0	24,0	34,0	44,0	17,5	25,0	Clau	0,0	21,0	47,0	18,0	26,0
Segments	54,0	58,0	56,0	57,0	39,0	46,5	51,0	49,0	52,5	59,0	35,0	43,5
Video	56,0	48,0	52,0	57,0	30,0	39,5	58,0	43,0	49,5	54,0	31,0	39,5



	C	onTagge	er	Simple NER			Wiki Machine			Spotlight		
	R	Р	F1	R	Р	F1	R	Р	F1	R	Р	F1
Authoritative	60,0	54,5	57,0	52,0	46,0	49,0	59,5	56,5	58,0	50,0	44,0	47,0
Tags	71,0	69,5	70,0	61,0	60,0	60,5	44,0	62,0	51,5	60,0	59,0	59,5
ASR	55,0	61,0	58,0	56,5	38,0	45,5	61,5	50,0	55,0	56,0	34,0	42,5
OCR	56,0	24,0	34,0	44,0	17,5	25,0	24,5	18,0	21,0	47,0	18,0	26,0
Segments	54,0	58,0	56,0	57,0	39,0	46,5	57,0	49,0	52,5	59,0	35,0	43,5
Video	56,0	48,0	52,0	57,0	30,0	39,5	58,0	43,0	49,5	54,0	31,0	39,5



	conTagger			Si	Simple NER			Wiki Machine			Spotlight		
	R	Р	F1	R	Р	F1	R	Р	F1	R	Р	F1	
Authoritative	60,0	54,5	57,0	52,0	46,0	49,0	59,5	56,5	58,0	50,0	44,0	47,0	
Tags	71,0	69,5	70,0	61,0	60,0	60,5	44,0	62,0	51,5	60,0	59,0	59,5	
ASR	55,0	61,0	58,0	56,5	38,0	45,5	61,5	50,0	55,0	56,0	34,0	42,5	
OCR	56,0	24,0	34,0	44,0	17,5	25,0	24,5	18,0	21,0	47,0	18,0	26,0	
Segments	54,0	58,0	56,0	57,0	39,0	46,5	57,0	49,0	52,5	59,0	35,0	43,5	
Video	56,0	48,0	52,0	57,0	30,0	39,5	58,0	43,0	49,5	54,0	31,0	39,5	

# Further Evaluation Findings -Dynamic Context Creation





#### computer dbp:Computer

confidence: 0.46

disambiguation score: 1.0

Dynamic Context Generation										
Source	Confidence	Disambiguation Score								
ASR	0,7	0,2								
OCR	0,7	0,2								
Authoritative	0,25	0,0								
Tags	0,4	0,0								

22

# Further Evaluation Findings -Dynamic Context Creation

22





Dynamic Context Generation									
Source	Confidence	Disambiguation Score							
ASR	0,7	0,2							
OCR	0,7	0,2							
Authoritative	0,25	0,0							
Tags	0,4	0,0							

## Further Evaluation Findings -Context Boundaries





Context Boundary											
	AS	5R	00	CR	Tags						
	Recall	Precision	Recall	Precision	Recall	Precision					
segment-based	55,0	61,0	56,0	24,0	71,0	69,5					
video-based	53,0	46,0	51,0	21,0	69,0	68,0					

23



- 1. Motivation Semantic Search on Videos
- 2. Challenges of Video Metadata
- 3. Contextual Model
- 4. Evaluation
- 5. Outlook



## Contributions



- Semantic analysis of videos exploiting metadata extracted from various sources
- Introduction of contextual descriptions model:
  - scoring of video metadata
  - order metadata items for subsequent disambiguation
- Evaluation on video dataset
  - improve precision and/or recall especially on text from sources with low confidence

25



### Ongoing & Future Work

- Weighting of single confidence value parts for total calculation
- Additional confidence "scorers"
- Negative context scoring
  - Penalties for links to negative context entities
  - comprehension of topic clusters
- Use of different knowledge bases DBpedia, GND, etc.
- Comprehension of User context





Thank you for your attention!

#### **Contact:**

Nadine Steinmetz Hasso-Plattner-Institut für Softwaresystemtechnik Universität Potsdam Prof.-Dr.-Helmert-Str. 2-3 D-14482 Potsdam

E-Mail: <u>nadine.steinmetz@hpi.uni-potsdam.de</u>