



Hasso
Plattner
Institut

IT Systems Engineering | Universität Potsdam

Named Entity Recognition for User-Generated Tags

Nadine Ludwig, Harald Sack
Hasso-Plattner-Institute for IT Systems Engineering
University of Potsdam

8th International Workshop on Text-based Information Retrieval
August 31st, 2011, Toulouse, France

Named Entity Recognition for User-Generated Tags

NER - Purpose and Overview

- Context Creation in Multimedia
- Proposed Method
- Outlook

NER - Purpose & Overview

3

„Tagging is fast becoming one of the primary ways people organize and manage digital information.“

Gene Smith: *Tagging: People-Powered Metadata for the Social Web* (2007)

NER - Purpose & Overview

3

„Tagging is fast becoming one of the primary ways people organize and manage digital information.“

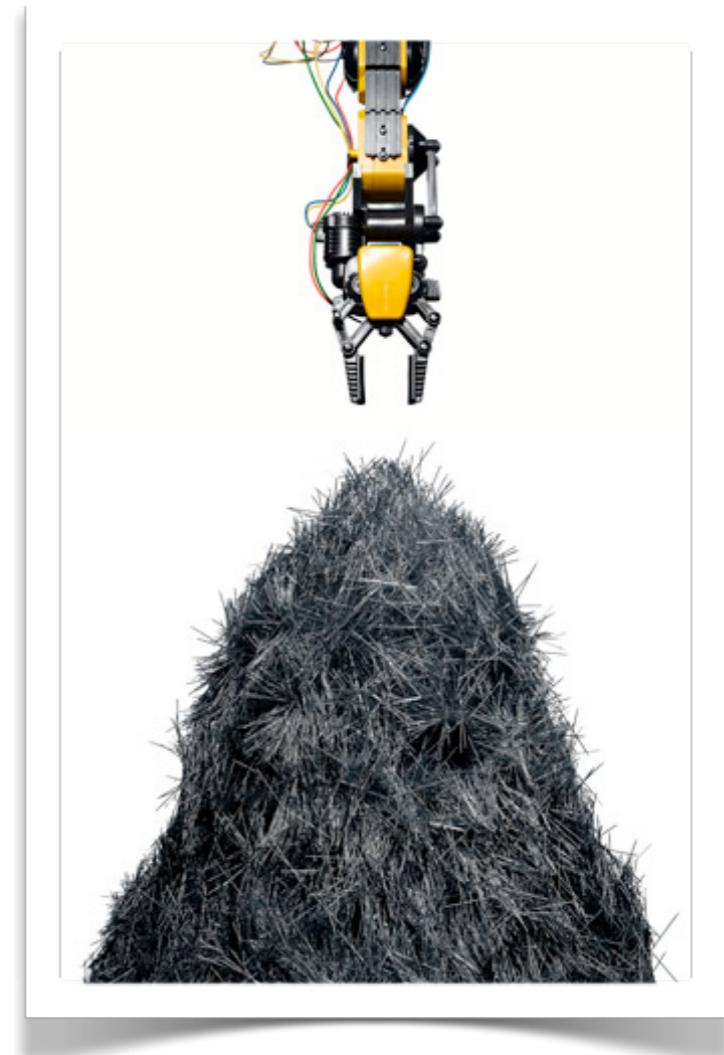
Gene Smith: *Tagging: People-Powered Metadata for the Social Web* (2007)



NER - Purpose & Overview

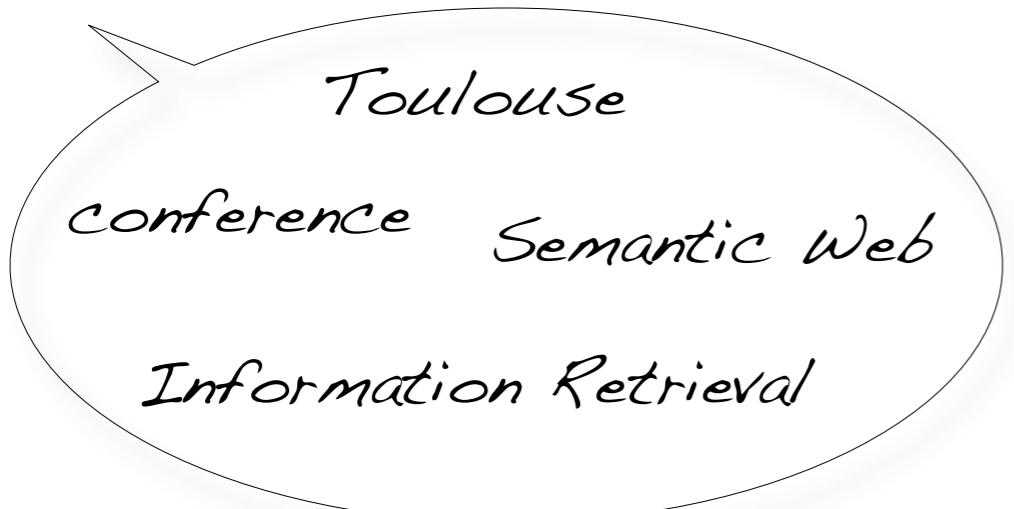
4

- tags = huge amount of user-generated metadata
- metadata is essential for efficient search -
especially on non-textual documents
- next generation of search = semantic search

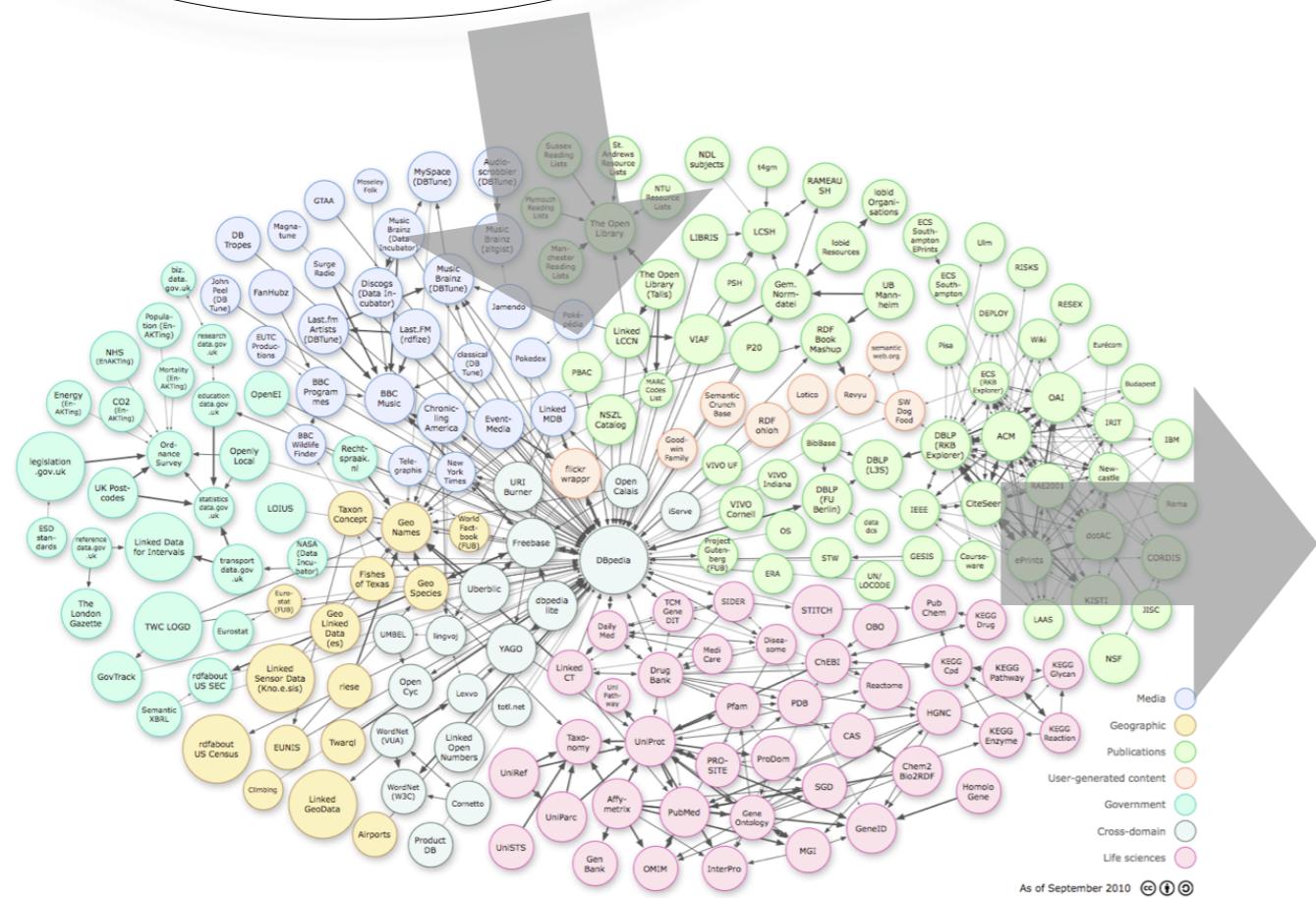


NER - Purpose & Overview

5



*Named Entity Recognition
on User-Generated Metadata*



 .../resource/Toulouse

 .../resource/Academic_conference

 .../resource/Semantic_Web

 .../resource/Information_retrieval

NER - Purpose & Overview

6

Main Problem in NER:
Ambiguity of Terms

NER - Purpose & Overview

6

Main Problem in NER:
Ambiguity of Terms

Example: „Jaguar“ in different contexts

jaguar

NER - Purpose & Overview

6



Main Problem in NER:
Ambiguity of Terms

Example: „Jaguar“ in different contexts



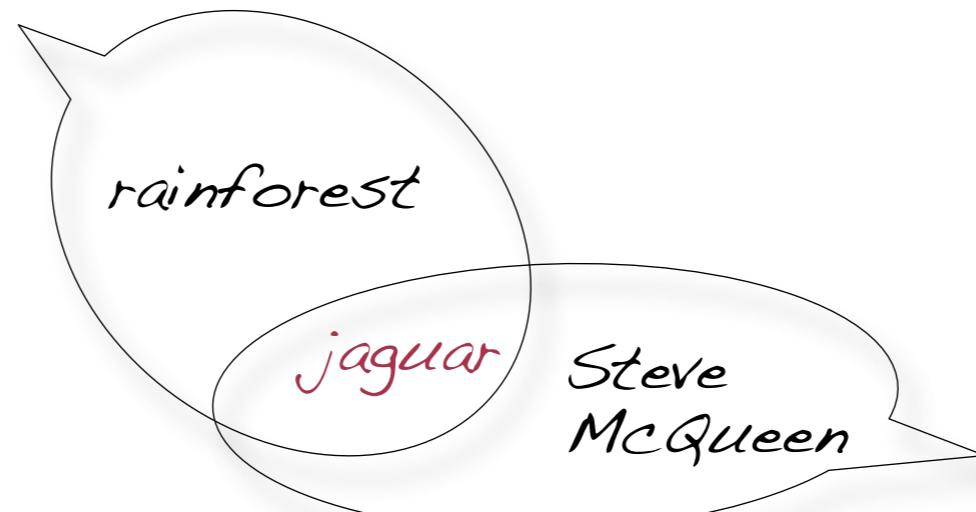
NER - Purpose & Overview

6



Main Problem in NER:
Ambiguity of Terms

Example: „Jaguar“ in different contexts



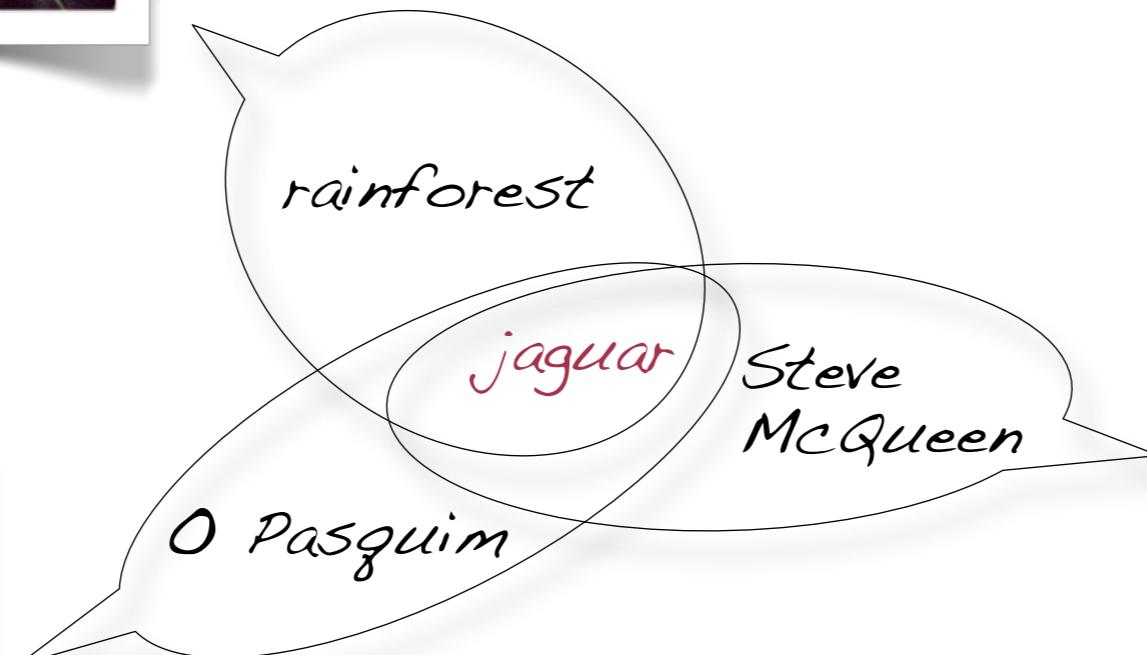
NER - Purpose & Overview

6



Main Problem in NER:
Ambiguity of Terms

Example: „Jaguar“ in different contexts



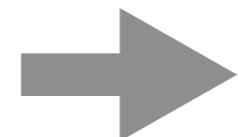
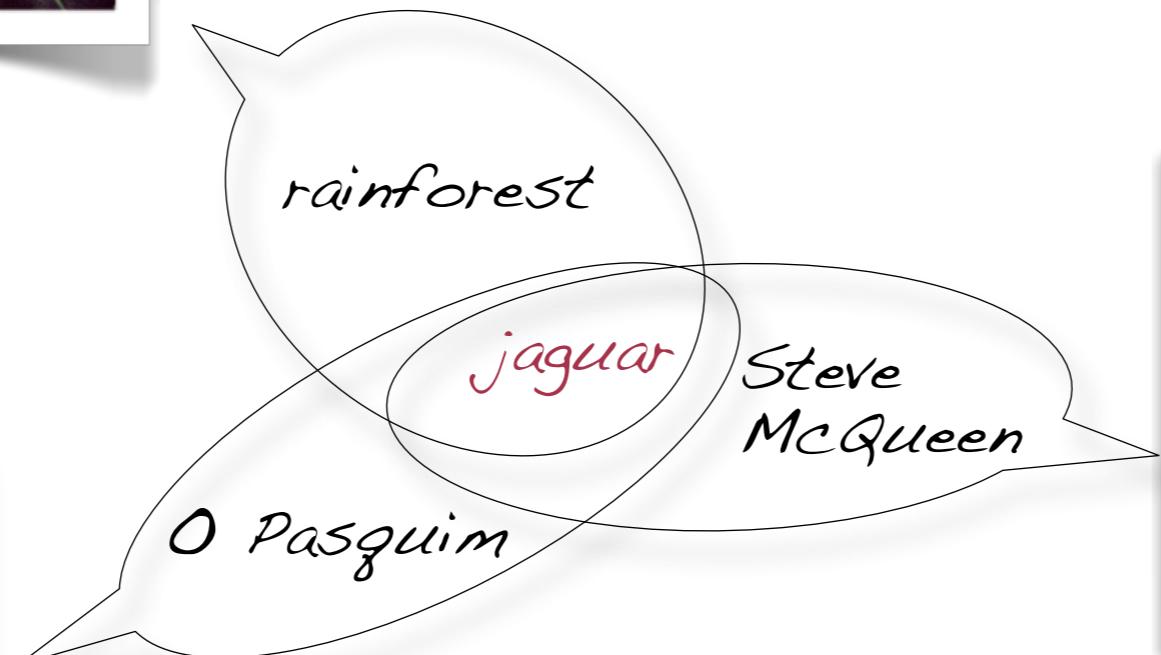
NER - Purpose & Overview

6



Main Problem in NER:
Ambiguity of Terms

Example: „Jaguar“ in different contexts



Context matters!

Named Entity Recognition for User-Generated Tags

- NER - Purpose and Overview

Context Creation in Multimedia

- Proposed Method
- Outlook

Context Creation in Multimedia



- presented approach is based on tagged multimedia esp. videos
- context creation in multimedia is multi-dimensional
- creating context in various combinations is decisive for the quality of the NER process

3 Dimensions of Context Creation in Multimedia

9

*User-centered
Dimension*



*Temporal
Dimension*

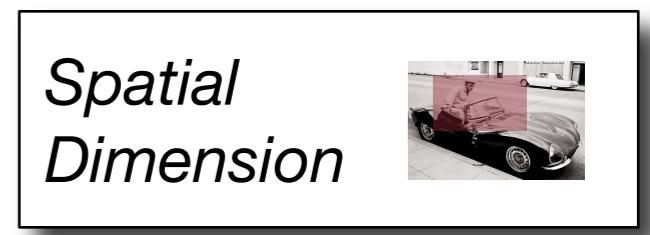
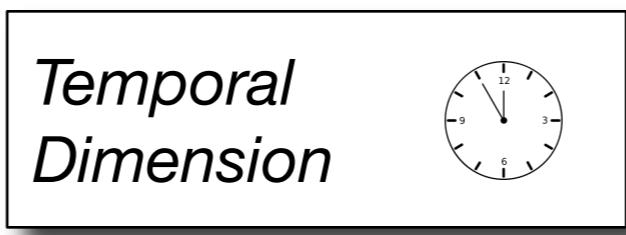


*Spatial
Dimension*



3 Dimensions of Context Creation in Multimedia

9



Tags by

one user

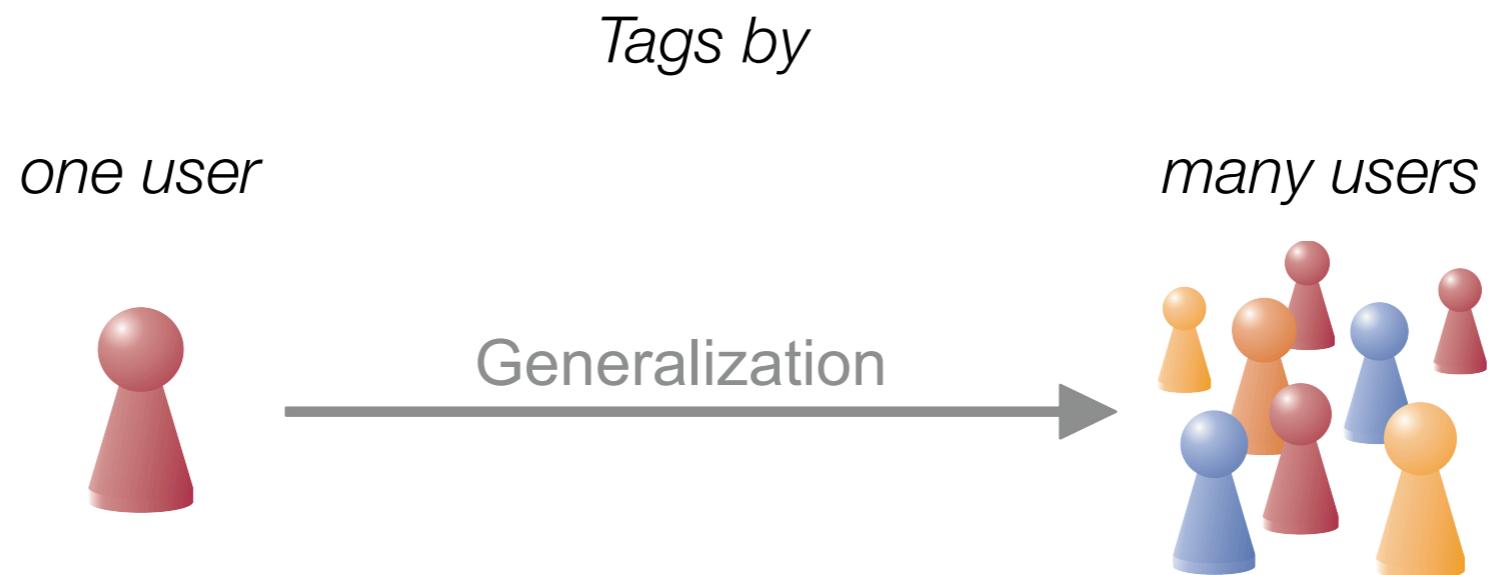
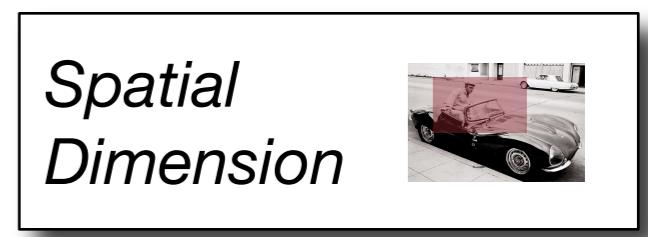
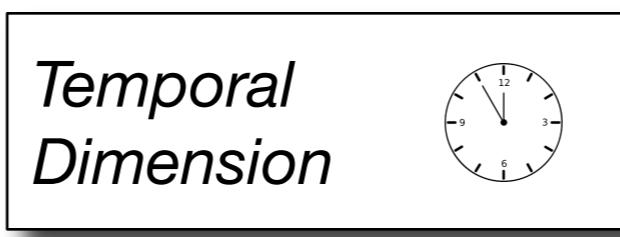


many users



3 Dimensions of Context Creation in Multimedia

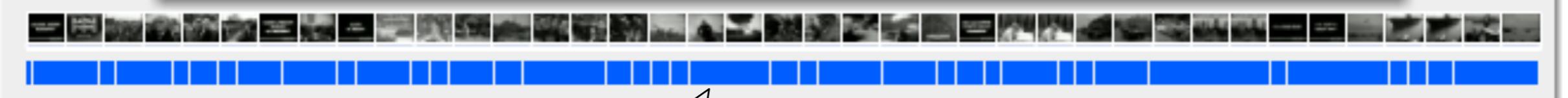
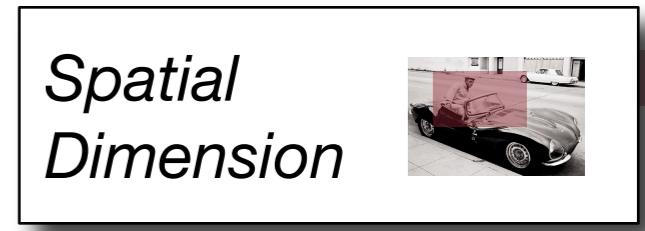
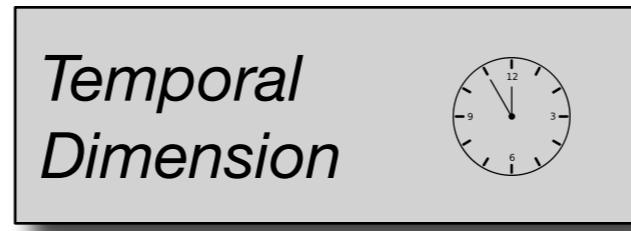
9



3 Dimensions of Context Creation in Multimedia



3 Dimensions of Context Creation in Multimedia



rainforest jaguar

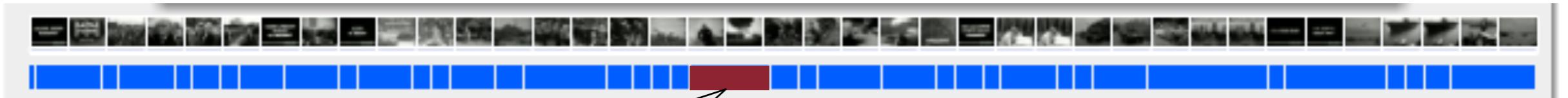
at the same timestamp

3 Dimensions of Context Creation in Multimedia



rainforest jaguar

at the same timestamp



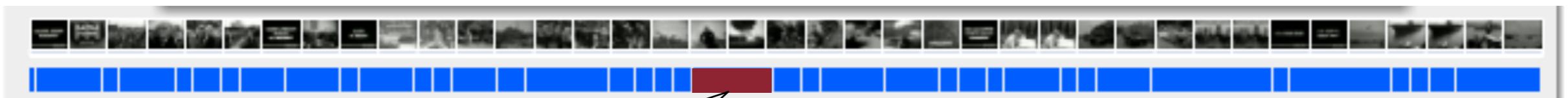
jaguar panthera
rainforest feline

in the same segment

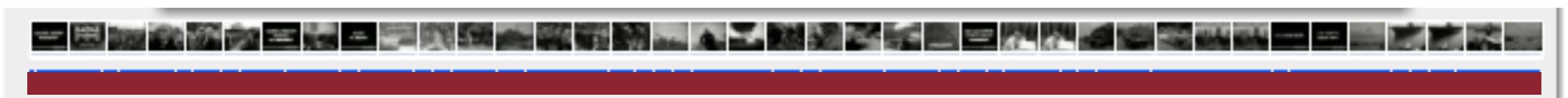
3 Dimensions of Context Creation in Multimedia



at the same timestamp



in the same segment



in the entire video

3 Dimensions of Context Creation in Multimedia



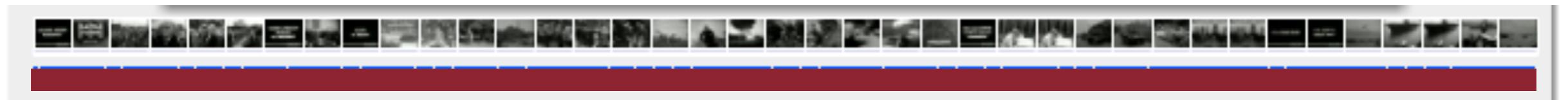
rainforest jaguar

at the same timestamp



jaguar panthera
rainforest feline

in the same segment



paraguay jaguar panthera
aztek rainforest feline

in the entire video

Generalization

3 Dimensions of Context Creation in Multimedia

11

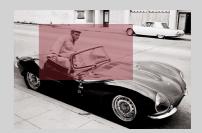
*User-centered
Dimension*



*Temporal
Dimension*



*Spatial
Dimension*



from: Stars & Cars by Tony Nourmand

3 Dimensions of Context Creation in Multimedia

11

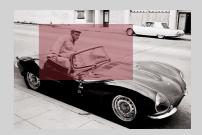
*User-centered
Dimension*



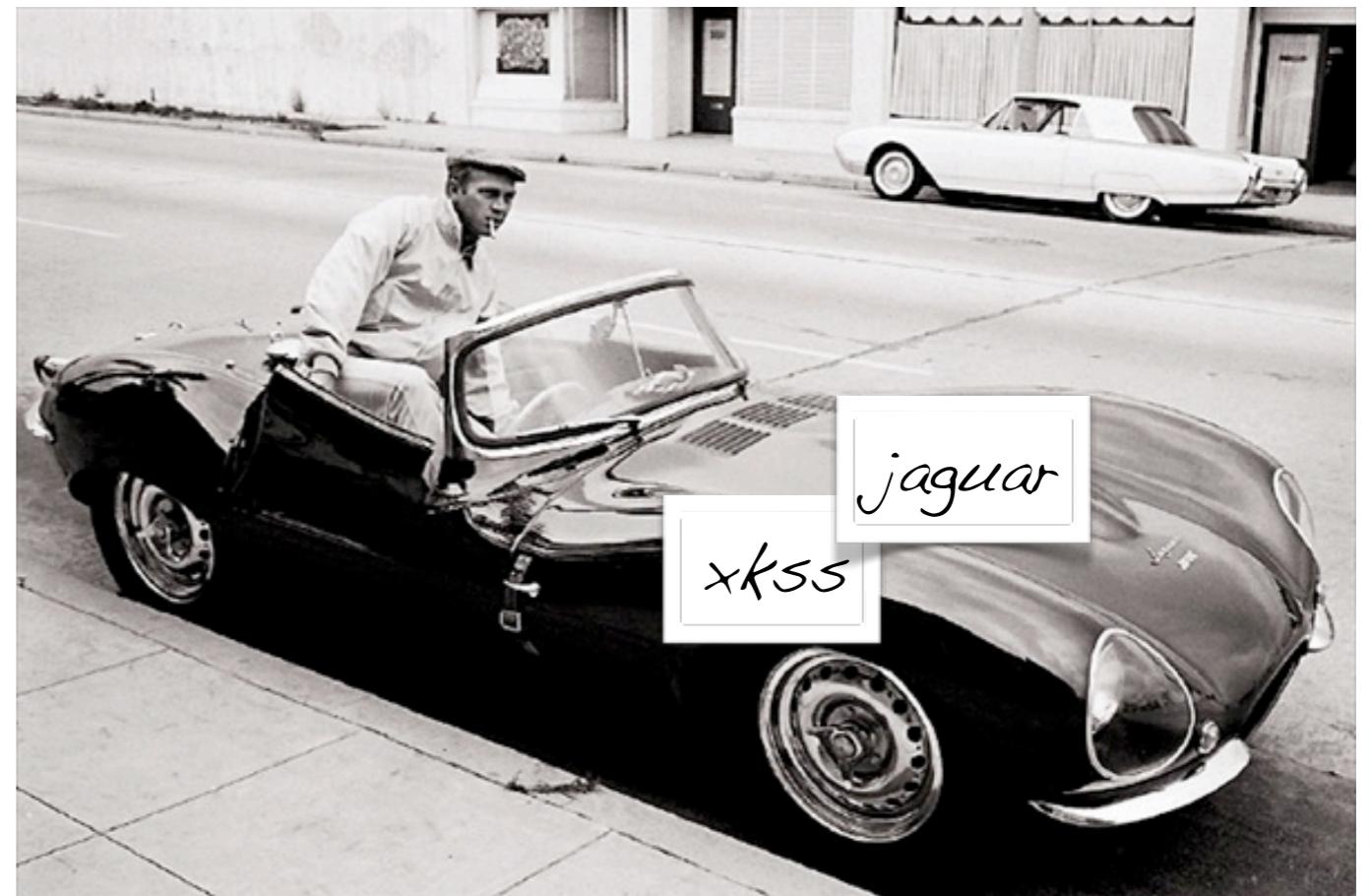
*Temporal
Dimension*



*Spatial
Dimension*



near by tags



from: Stars & Cars by Tony Nourmand

3 Dimensions of Context Creation in Multimedia

11

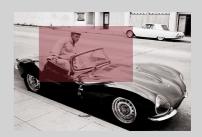
*User-centered
Dimension*



*Temporal
Dimension*



*Spatial
Dimension*



near by tags

tags in same region
(e.g. right lower quarter)



from: Stars & Cars by Tony Nourmand

3 Dimensions of Context Creation in Multimedia

11

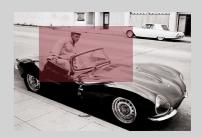
*User-centered
Dimension*



*Temporal
Dimension*



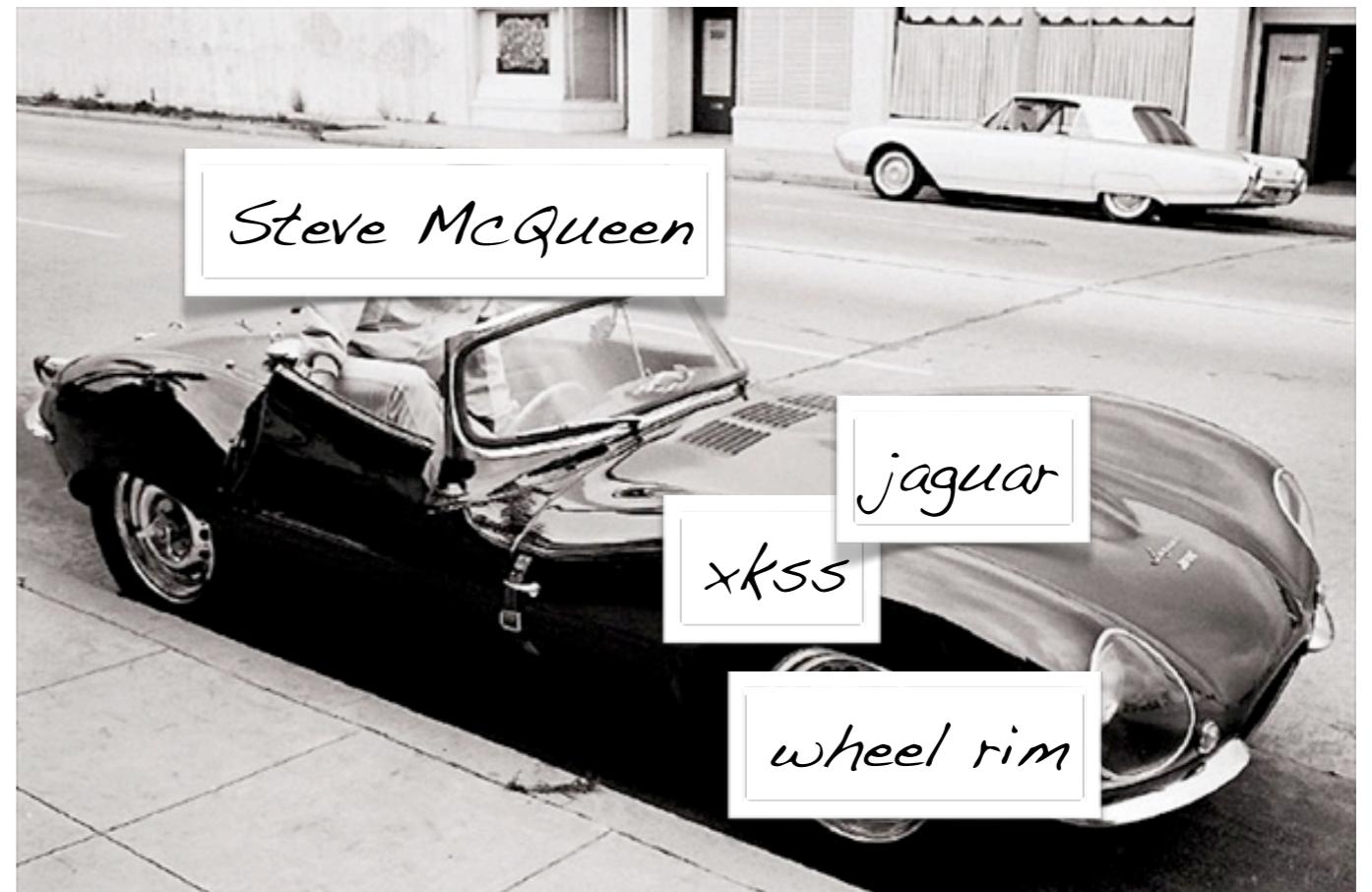
*Spatial
Dimension*



near by tags

tags in same region
(e.g. right lower quarter)

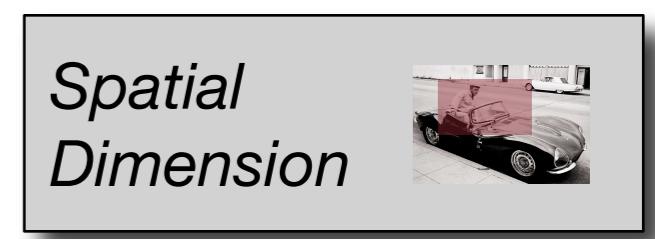
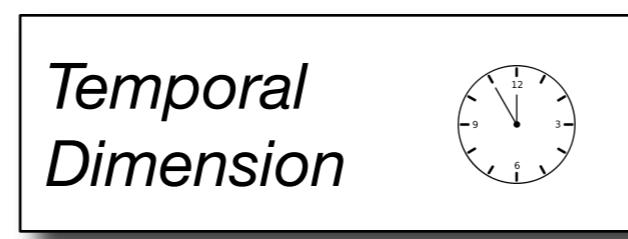
tags in the whole frame



from: Stars & Cars by Tony Nourmand

3 Dimensions of Context Creation in Multimedia

11

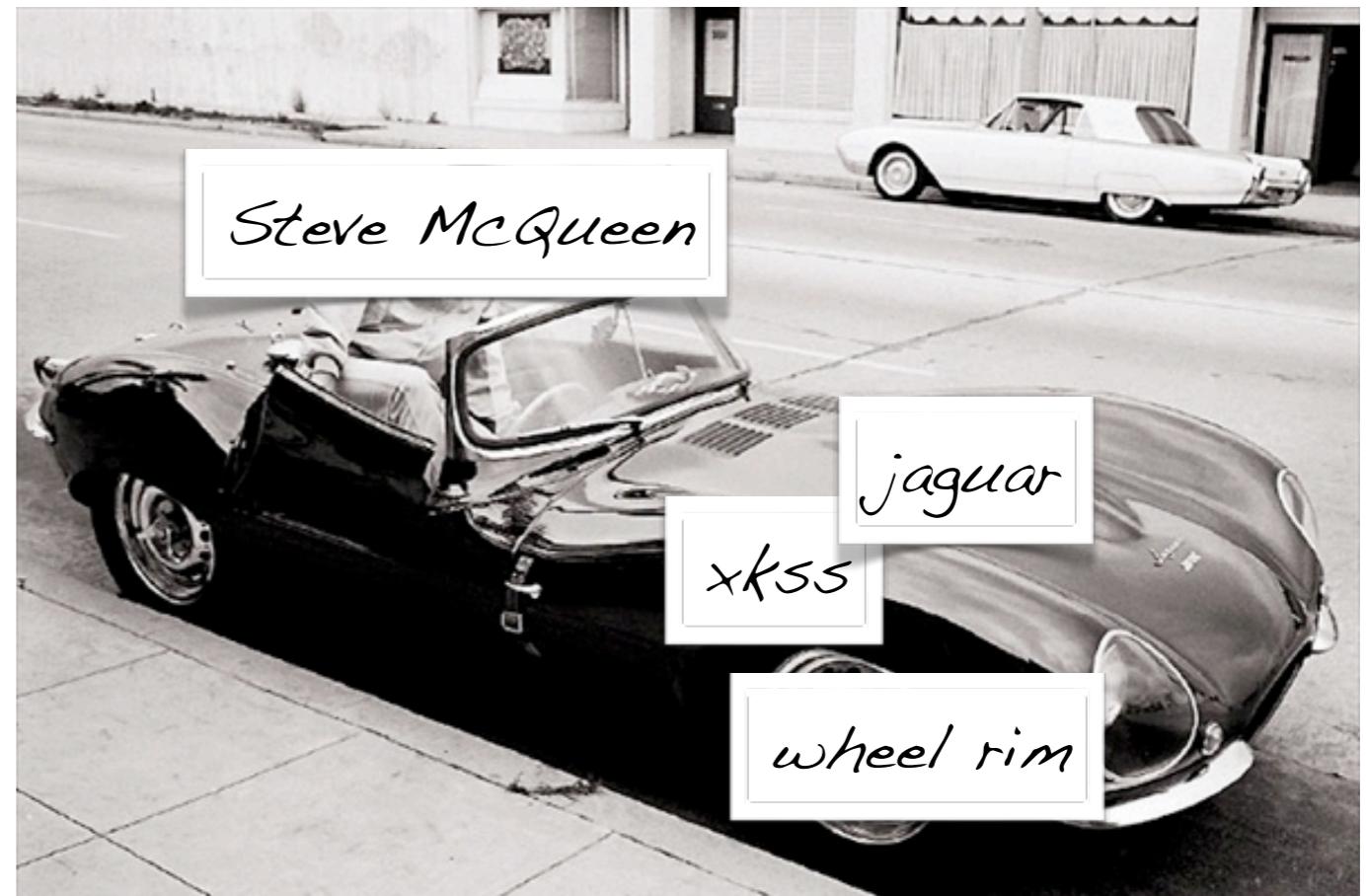


Generalization

near by tags

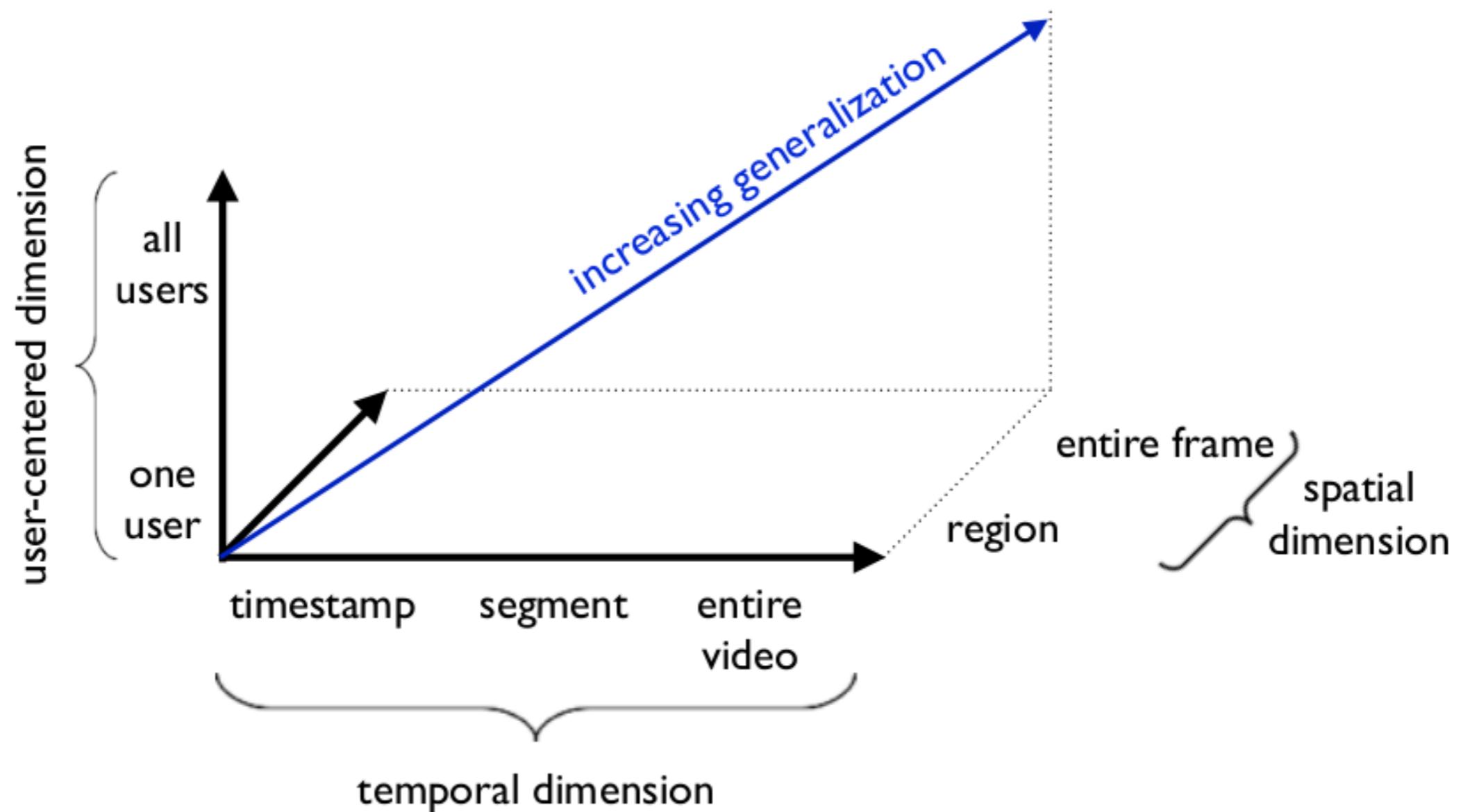
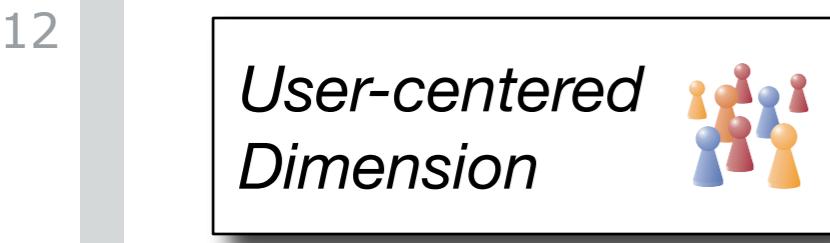
tags in same region
(e.g. right lower quarter)

tags in the whole frame



from: Stars & Cars by Tony Nourmand

3 Dimensions of Context Creation in Multimedia



Named Entity Recognition for User-Generated Tags

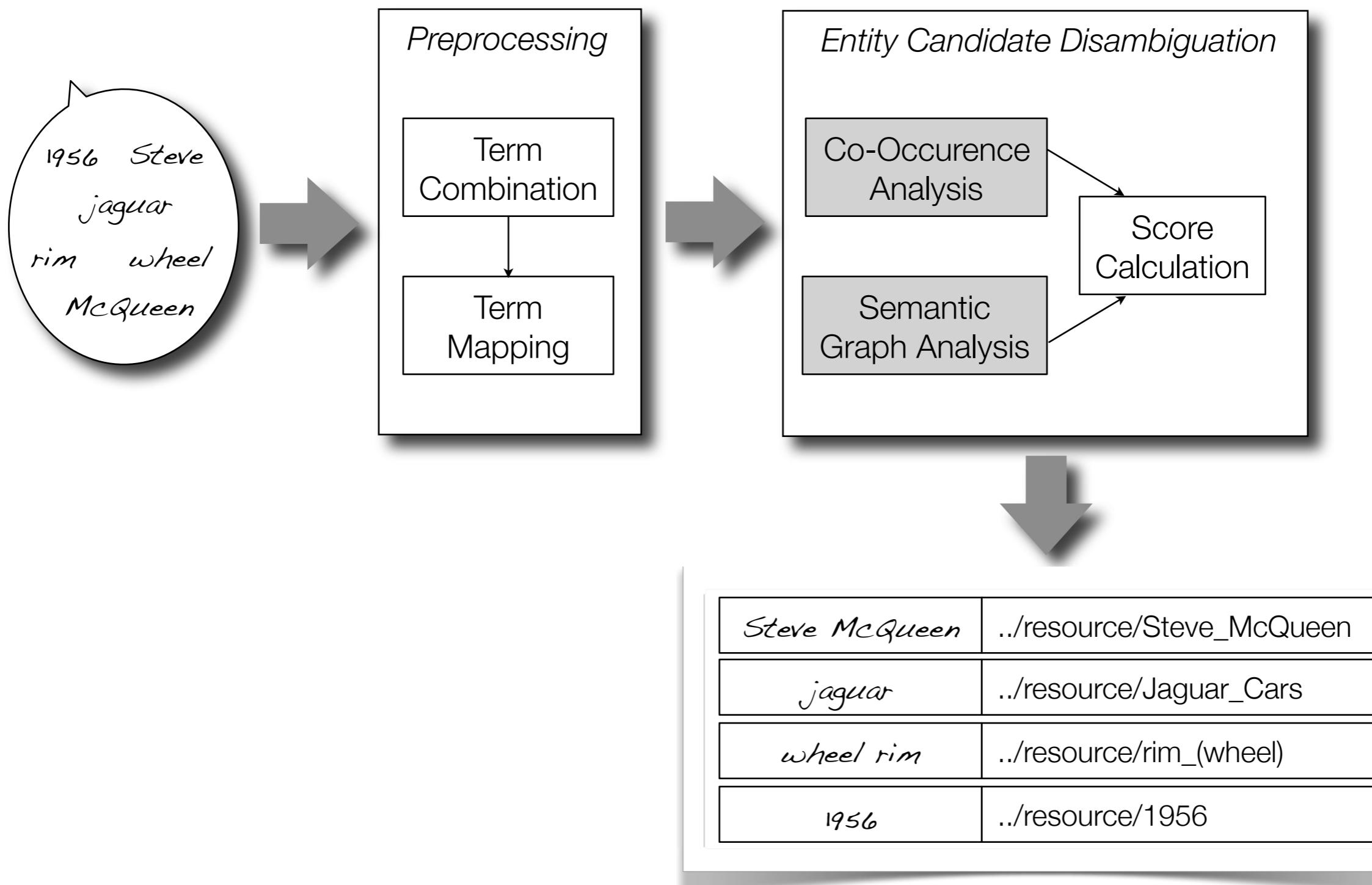
- NER - Purpose and Overview
- Context Creation in Multimedia

Proposed Method

- Outlook

NER Workflow

14

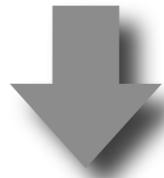
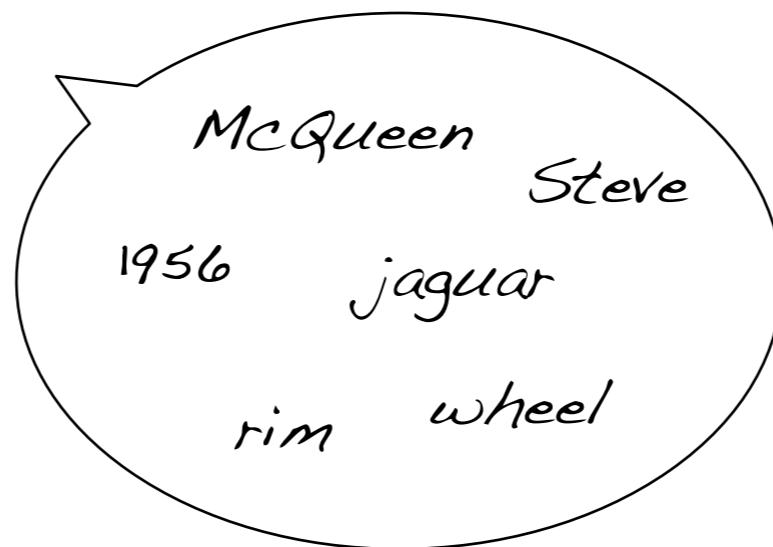


Preprocessing

15

Term Combination

$$c = \sum_{k=1}^j \frac{n!}{(n-k)!}$$



max.
156 combinations

Preprocessing

15

Term Combination

$$c = \sum_{k=1}^j \frac{n!}{(n-k)!}$$

Steve McQueen

wheel rim

jaguar

1956

Preprocessing

15

Term Combination

$$c = \sum_{k=1}^j \frac{n!}{(n-k)!}$$

Minimizing amount of combinations
by only using combined terms and discard single
words that are part of a combined term !

Steve McQueen

wheel rim

jaguar

1956

Preprocessing

16

Assigning Entity Candidates

 7 entity candidates

Steve McQueen

 1 entity candidate

wheel rim

 36 entity candidates

jaguar

 2 entity candidates

1956

Co-Occurrence Analysis

17

„jaguar“
http://dbpedia.org/resource/Jaguar_Cars

Co-Occurrence Analysis

17

„jaguar“

http://dbpedia.org/resource/Jaguar_Cars



context tags:

1956 wheel rim
steve mcqueen

Jaguar Cars

From Wikipedia, the free encyclopedia
(Redirected from [Jaguar cars](#))

Jaguar Cars Ltd., known simply as *Jaguar* (/dʒægjuːər/), is a British luxury car manufacturer, headquartered in Whitley, [Coventry](#), England. A wholly owned subsidiary of the Indian company [Tata Motors Ltd.](#), it is operated as part of the [Jaguar Land Rover](#) business.^[2]

Jaguar was founded as the [Swallow Sidecar Company](#) by [Sir William Lyons](#) in 1922, originally making motorcycle [sidecars](#) before evolving into passenger cars. The name was changed to Jaguar after WWII to avoid the unfavourable connotations of the SS initials.^[3] Following a merger with the British Motor Corporation in 1968, subsequently subsumed by Leyland, which itself was later nationalised as [British Leyland](#), Jaguar was listed on the [London Stock Exchange](#) in 1984, and became a constituent of the [FTSE 100 Index](#) until it was acquired by [Ford](#) in 1989.^[4] Jaguar has, in recent years, manufactured cars for the [Prime Minister](#), the most recent delivery being of a [XJ](#) model on 11 May 2010.^[5] The company also holds [Royal Warrants](#) from [HM Queen Elizabeth II](#) and [HRH Prince Charles](#).^[6]

Jaguar cars today are designed in Jaguar Land Rover's engineering centres at the [Whitley plant](#) in Coventry and at their [Gaydon site](#) in [Warwickshire](#), and are manufactured in Jaguar's [Castle Bromwich assembly plant](#) near Birmingham.

Contents [hide]

- 1 History
 - 1.1 Birth of the company
 - 1.2 British Leyland
 - 1.3 Ford Motor Company era
 - 1.4 Tata Motors era
 - 1.5 Assembly plant
- 2 Current car models
 - 2.1 XJ
 - 2.2 XF
 - 2.3 XK

Jaguar Cars Limited



Type	Private limited company
Industry	Automotive
Founded	11 September 1922 (as Swallow Sidecar Company), changed to Jaguar Cars Limited on 9 April 1945
Founder(s)	Sir William Lyons and William Walmsley
Headquarters	Whitley, Coventry , United Kingdom
Key people	Ratan Tata (Chairman) Dr Ralf Speth (CEO) Adrian Hallmark (Global Brand Director)
Products	Automobiles
Owner(s)	Tata Motors
Employees	10,000 ^[1]
Parent	Jaguar Land Rover
Website	jaguar.com

Co-Occurrence Analysis

17

„jaguar“

http://dbpedia.org/resource/Jaguar_Cars



context tags:

1956 wheel rim
steve mcqueen

$$score_{CA_i} = counter_{cooc_i} \cdot \frac{|W(uri(t)_i) \cap C(t)|}{|C(t)|}$$

Jaguar Cars

From Wikipedia, the free encyclopedia
(Redirected from [Jaguar cars](#))

Jaguar Cars Ltd., known simply as *Jaguar* (/dʒægju:ər/), is a British luxury car manufacturer, headquartered in Whitley, [Coventry](#), England. A wholly owned subsidiary of the Indian company [Tata Motors Ltd.](#), it is operated as part of the [Jaguar Land Rover](#) business.^[2]

Jaguar was founded as the [Swallow Sidecar Company](#) by [Sir William Lyons](#) in 1922, originally making motorcycle [sidecars](#) before evolving into passenger cars. The name was changed to Jaguar after WWII to avoid the unfavourable connotations of the SS initials.^[3] Following a merger with the British Motor Corporation in 1968, subsequently subsumed by Leyland, which itself was later nationalised as [British Leyland](#), Jaguar was listed on the [London Stock Exchange](#) in 1984, and became a constituent of the [FTSE 100 Index](#) until it was acquired by [Ford](#) in 1989.^[4] Jaguar has, in recent years, manufactured cars for the [Prime Minister](#), the most recent delivery being of a [XJ](#) model on 11 May 2010.^[5] The company also holds [Royal Warrants](#) from [HM Queen Elizabeth II](#) and [HRH Prince Charles](#).^[6]

Jaguar cars today are designed in Jaguar Land Rover's engineering centres at the [Whitley plant](#) in Coventry and at their [Gaydon site](#) in [Warwickshire](#), and are manufactured in Jaguar's [Castle Bromwich assembly plant](#) near Birmingham.

Contents [hide]

- 1 History
 - 1.1 Birth of the company
 - 1.2 British Leyland
 - 1.3 Ford Motor Company era
 - 1.4 Tata Motors era
 - 1.5 Assembly plant
- 2 Current car models
 - 2.1 XJ
 - 2.2 XF
 - 2.3 XK

Jaguar Cars Limited



Type	Private limited company
Industry	Automotive
Founded	11 September 1922 (as Swallow Sidecar Company), changed to Jaguar Cars Limited on 9 April 1945
Founder(s)	Sir William Lyons and William Walmsley
Headquarters	Whitley, Coventry , United Kingdom
Key people	Ratan Tata (Chairman) Dr Ralf Speth (CEO) Adrian Hallmark (Global Brand Director)
Products	Automobiles
Owner(s)	Tata Motors
Employees	10,000 ^[1]
Parent	Jaguar Land Rover
Website	jaguar.com

Co-Occurrence Analysis

17

„jaguar“

http://dbpedia.org/resource/Jaguar_Cars



context tags:

1956 wheel rim
steve mcqueen

$$score_{CA_i} = counter_{cooc_i} \cdot \frac{|W(uri(t)_i) \cap C(t)|}{|C(t)|}$$



score: 2.0

Jaguar Cars

From Wikipedia, the free encyclopedia
(Redirected from [Jaguar cars](#))

Jaguar Cars Ltd., known simply as *Jaguar* (/dʒægjuːər/), is a British luxury car manufacturer, headquartered in Whitley, [Coventry](#), England. A wholly owned subsidiary of the Indian company [Tata Motors Ltd.](#), it is operated as part of the [Jaguar Land Rover](#) business.^[2]

Jaguar was founded as the [Swallow Sidecar Company](#) by [Sir William Lyons](#) in 1922, originally making motorcycle [sidecars](#) before evolving into passenger cars. The name was changed to Jaguar after WWII to avoid the unfavourable connotations of the SS initials.^[3] Following a merger with the British Motor Corporation in 1968, subsequently subsumed by Leyland, which itself was later nationalised as [British Leyland](#), Jaguar was listed on the [London Stock Exchange](#) in 1984, and became a constituent of the [FTSE 100 Index](#) until it was acquired by [Ford](#) in 1989.^[4] Jaguar has, in recent years, manufactured cars for the [Prime Minister](#), the most recent delivery being of a [XJ](#) model on 11 May 2010.^[5] The company also holds [Royal Warrants](#) from [HM Queen Elizabeth II](#) and [HRH Prince Charles](#).^[6]

Jaguar cars today are designed in Jaguar Land Rover's engineering centres at the [Whitley plant](#) in Coventry and at their [Gaydon site](#) in [Warwickshire](#), and are manufactured in Jaguar's [Castle Bromwich assembly plant](#) near Birmingham.

Contents [hide]

- 1 History
 - 1.1 Birth of the company
 - 1.2 British Leyland
 - 1.3 Ford Motor Company era
 - 1.4 Tata Motors era
 - 1.5 Assembly plant
- 2 Current car models
 - 2.1 XJ
 - 2.2 XF
 - 2.3 XK

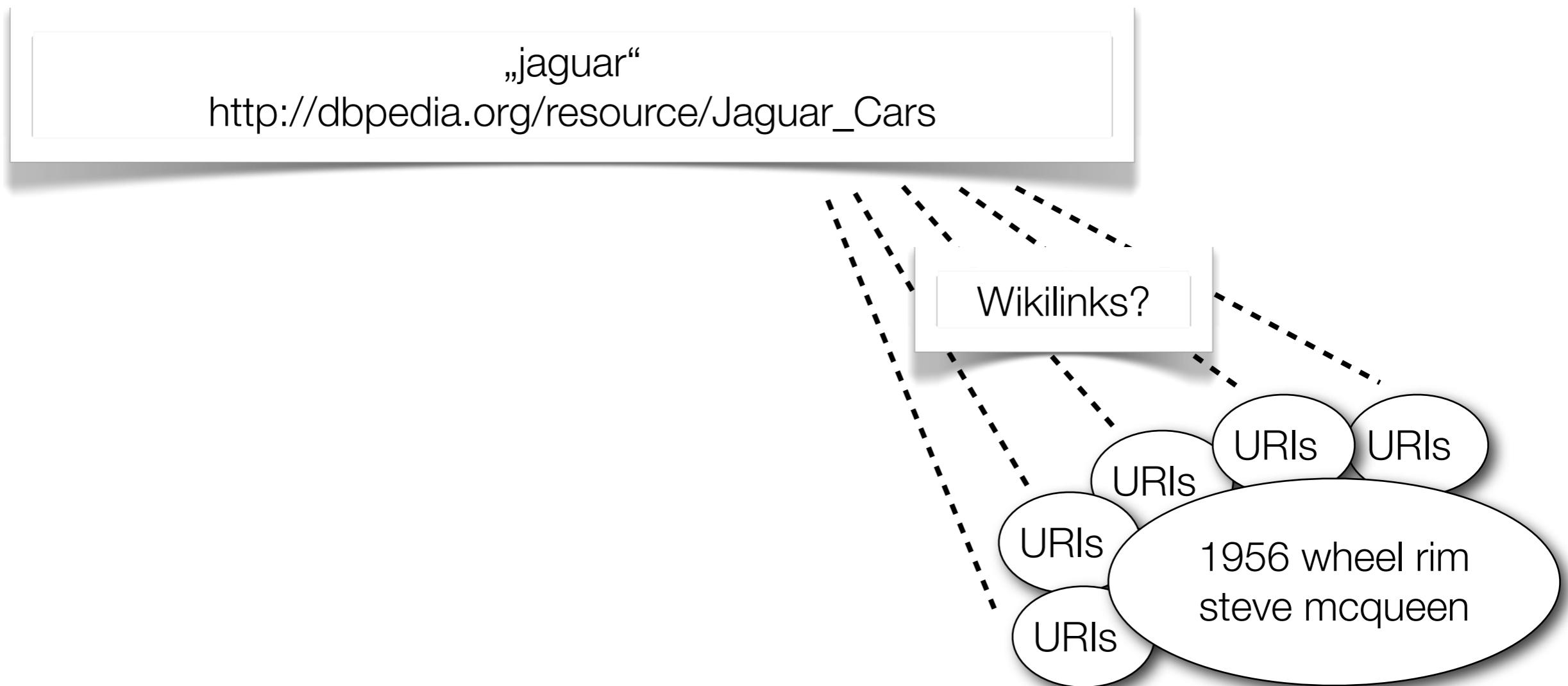
Jaguar Cars Limited



Type	Private limited company
Industry	Automotive
Founded	11 September 1922 (as Swallow Sidecar Company), changed to Jaguar Cars Limited on 9 April 1945
Founder(s)	Sir William Lyons and William Walmsley
Headquarters	Whitley, Coventry , United Kingdom
Key people	Ratan Tata (Chairman) Dr Ralf Speth (CEO) Adrian Hallmark (Global Brand Director)
Products	Automobiles
Owner(s)	Tata Motors
Employees	10,000 ^[1]
Parent	Jaguar Land Rover
Website	jaguar.com

Semantic Graph Analysis

18



Semantic Graph Analysis

19

„jaguar“
http://dbpedia.org/resource/Jaguar_Cars

Three different Types of Links:

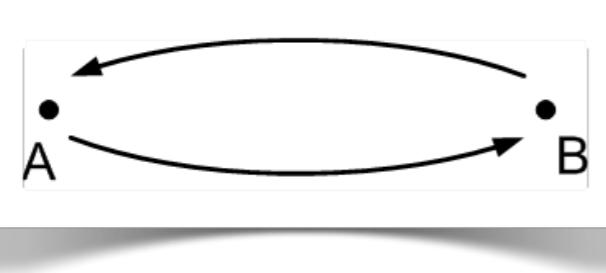
Semantic Graph Analysis

19

„jaguar“

http://dbpedia.org/resource/Jaguar_Cars

Three different Types of Links:



1. Direct Links

Weight: 0.45

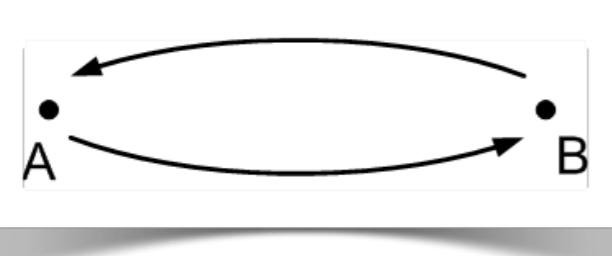
Semantic Graph Analysis

19

„jaguar“

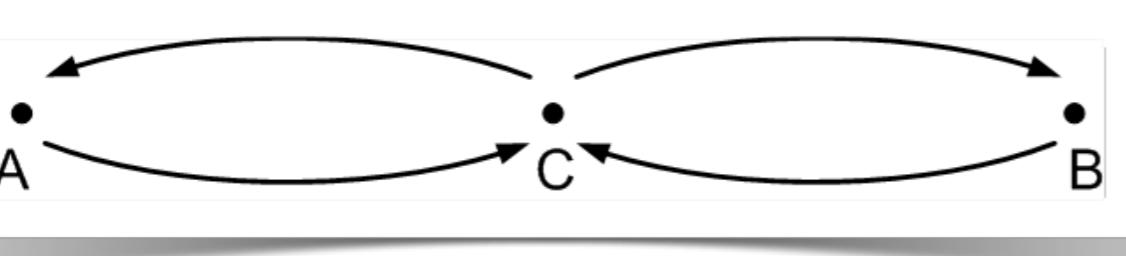
http://dbpedia.org/resource/Jaguar_Cars

Three different Types of Links:



1. Direct Links

Weight: 0.45



2. Symmetric Links
over a node

Weight: 0.30

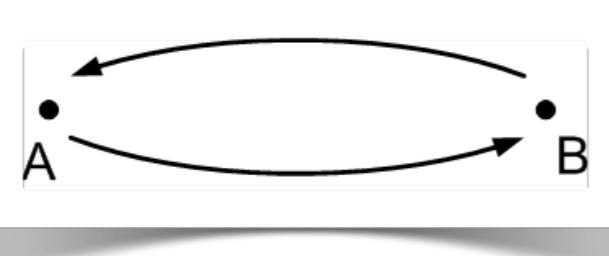
Semantic Graph Analysis

19

„jaguar“

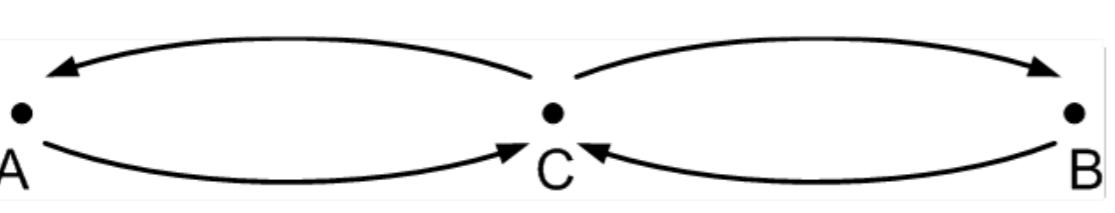
http://dbpedia.org/resource/Jaguar_Cars

Three different Types of Links:



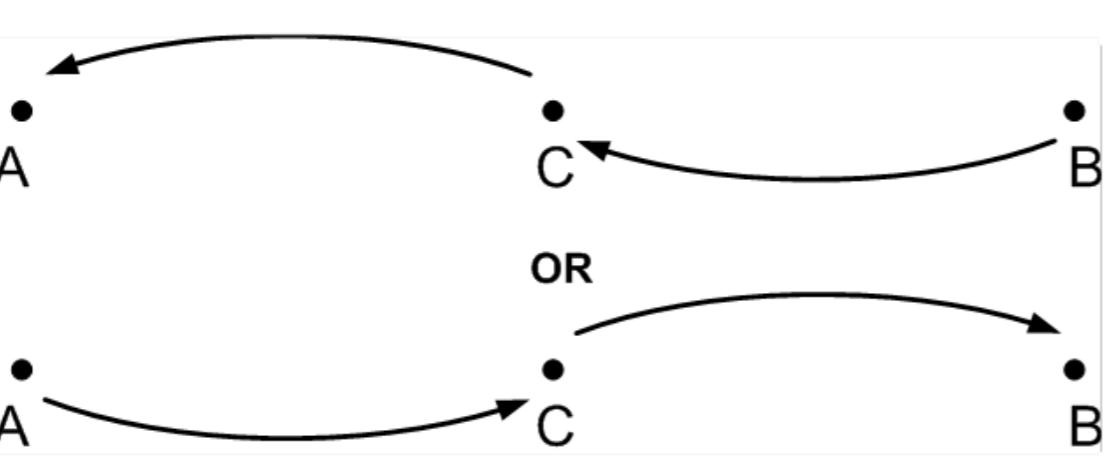
1. Direct Links

Weight: 0.45



2. Symmetric Links
over a node

Weight: 0.30



3. Links through a node
(not symmetric)

Weight: 0.25

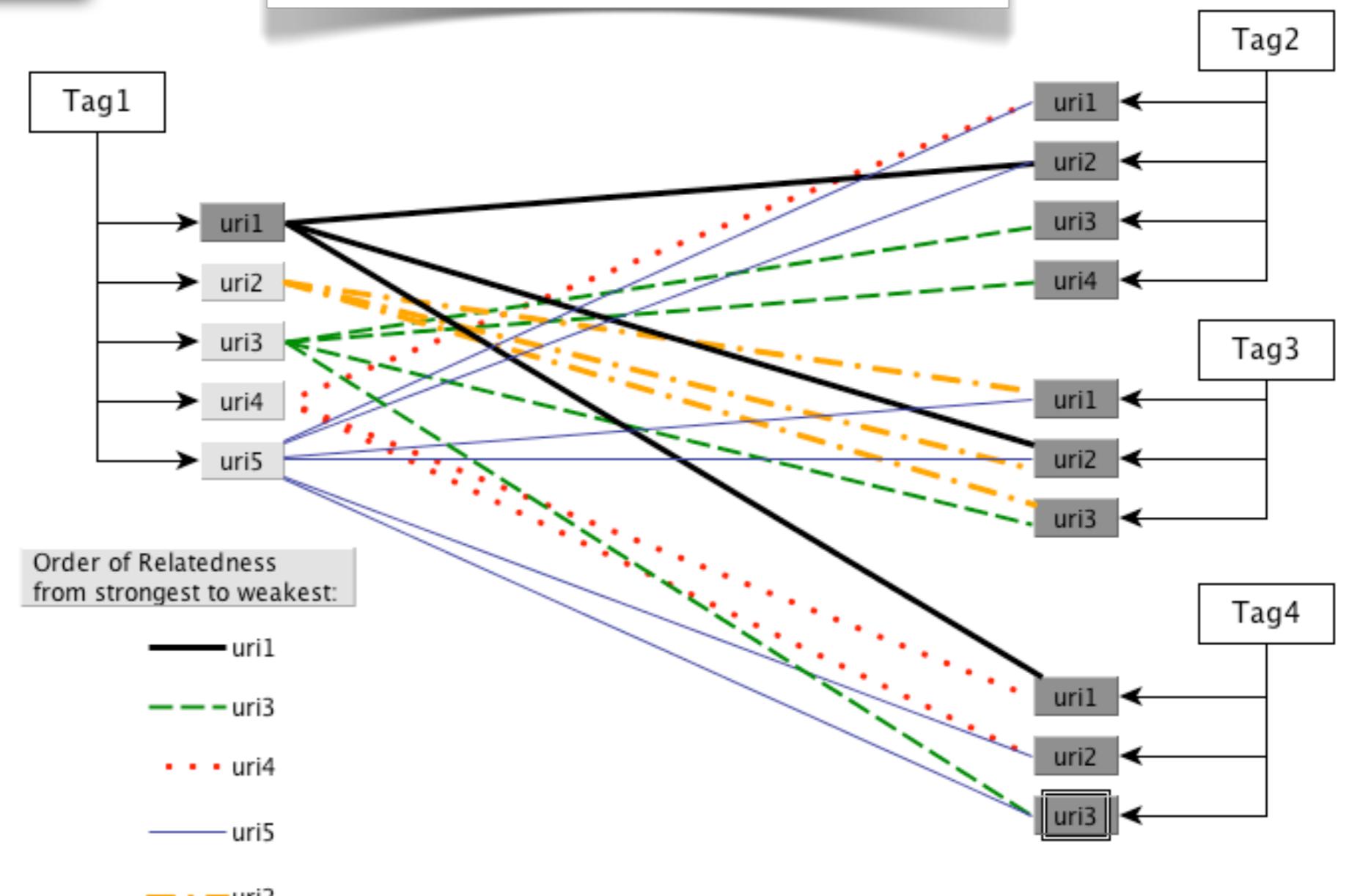
Semantic Graph Analysis

20

Score Calculation

Example: Direct Links

$$score_{dlinks_i} = \frac{|t \rightarrow t_k|}{|C(t)| \cdot counter_{dlinks_i}}$$



Score Calculation

21

- Result of score analysis is a list of all URIs for a tag with corresponding score in that context

$$\text{score}_{\text{total}} = \alpha \cdot \text{score}_{\text{cooc}} + \beta \cdot \text{score}_{\text{wikilinks}}$$



Normalized to [0.0 ... 1.0]



tag	URI	score
jaguar	http://dbpedia.org/resource/Jaguar_Cars	1,0
jaguar	http://dbpedia.org/resource/Jaguar_(cartoonist)	0,94
jaguar	http://dbpedia.org/resource/Jaguar	0,90

Results

22

2 test sets of tags from



	50 Segments (256 Tags)	50 Timestamps (315 Tags)
Original Mappings	11794 entity candidates (9-1224 candidates per context)	7562 entity candidates (13 - 1282 candidates per context)
Assignments	300 Entities	485 Entities

Results:

	Spotlight	HPI
Recall	39% (42%)	78% (81%)
Precision	34% (39%)	64% (41%)
F_1 -measure	36% (40%)	69% (54%)



Test sets available at: <http://yovisto.com/labs/ner>

Named Entity Recognition for User-Generated Tags

- NER - Purpose and Overview
- Context Creation in Multimedia
- Proposed Method & Results

Ongoing Work

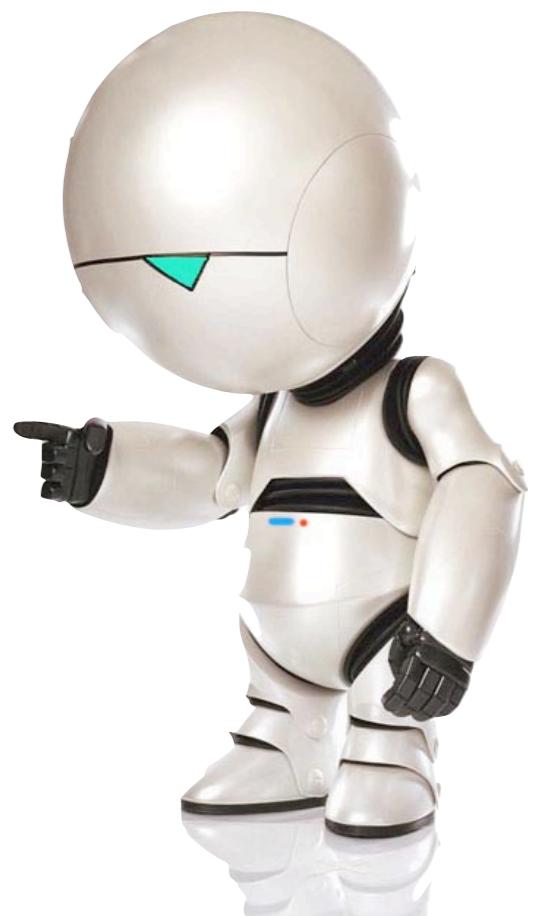
Conclusion

24

presented an Approach of Named Entity
Recognition for User-Generated Tags

introduced multi-dimensional Problem of Context
Creation in Multimedia

Results are promising, but ongoing work is
focusing on improving the algorithm



Ongoing Work

25

Enrichment of Wikipedia articles with paragraphs from directly linked articles (Combination of Graph and Co-Occurrence Analysis)

Enhancement of Term Combination and Mapping

Context Combination with different types of metadata:

- authorative metadata
- data from OCR and ASR



Named Entity Recognition for User-Generated Tags

Thank You!

