

---

# Entity-centric information access for high-end semantic applications

Simone Paolo Ponzetto

(joint work with Michael Schuhmacher,  
Federico Nanni and Laura Dietz)

---

# Hi!

Since 2016 professor of Information Systems at the University of Mannheim

Main research areas

- Knowledge Acquisition
- Natural Language Processing
- Computational Social Science



# Fake News

# Fake News

In an interview with radio host Hugh Hewitt last week, **Ryan said he's run a sub-3:00 marathon** [...] Last night, the vice-presidential candidate gave The New Yorker a statement:



The screenshot shows the Wikipedia article for Paul Ryan. At the top, there is a navigation bar with tabs for "Article" and "Talk". To the right, there are links for "Read", "View source", and "View history", along with a search box labeled "Search Wikipedia". The article title is "Paul Ryan" with a lock icon indicating it is a protected page. Below the title, it says "From Wikipedia, the free encyclopedia". A note reads: "For other people named Paul Ryan, see Paul Ryan (disambiguation)." The main text begins: "Paul Davis Ryan Jr. (born January 29, 1970) is an American politician serving as the 54th Speaker of the United States House of Representatives since 2015. He was the Republican Party nominee for Vice President of the United States, running alongside former Massachusetts Governor Mitt Romney, in the 2012 U.S. presidential election. Ryan also has been the U.S. Representative for Wisconsin's 1st congressional district since 1999. He was previously chair of the House Ways and Means Committee from 2011 to 2015. On October 29, 2015, Ryan was elected to replace John Boehner as Speaker of the U.S. House of Representatives following Boehner's retirement, becoming the first person from Wisconsin to hold this position. On April 11, 2018, Ryan announced that he would not be seeking re-election in 2018." To the right of the text is a portrait of Paul Ryan with the caption "54th Speaker of the United States House of Representatives". On the left side of the page, there is a sidebar with the Wikipedia logo and various navigation links such as "Main page", "Contents", "Featured content", "Current events", "Random article", "Donate to Wikipedia", "Wikipedia store", "Interaction", "Help", "About Wikipedia", "Community portal", "Recent changes", "Contact page", "Tools", "What links here", "Related changes", "Upload file", "Special pages", "Permanent link", and "Page information".

Source: Nicholas Thompson: LIE OR MISTAKE? PAUL RYAN'S MARATHONING PAST. The New Yorker, September 1, 2012

# Fake News

In an interview with radio host Hugh Hewitt last week, Ryan said he's run a sub-3:00 marathon [...] Last night, the vice-presidential candidate gave The New Yorker a statement:

“The race was more than 20 years ago, but my brother Tobin – who ran Boston last year – reminds me that he is the owner of the fastest marathon in the family and has never himself ran a sub-three. If I were to do any rounding, it would certainly be to four hours, not three. He gave me a good ribbing over this at dinner tonight.”

# Fake News

In an interview with radio host Hugh Hewitt last week, Ryan said he's run a sub-3:00 marathon [...] Last night, the vice-presidential candidate gave The New Yorker a statement:

“The race was more than 20 years ago, but my brother Tobin – who ran Boston last year – reminds me that he is the owner of the fastest marathon in the family and has never himself ran a sub-three. If I were to do any rounding, it would certainly be to four hours, not three. He gave me a good ribbing over this at dinner tonight.”

Runners – and Mr. Ryan says he continues to be one – also just don't forget race times. They talk about them with their friends; they think about them when running.

Source: Nicholas Thompson: LIE OR MISTAKE? PAUL RYAN'S MARATHONING PAST. The New Yorker, September 1, 2012

# Fake News

In an interview with radio host Hugh Hewitt last week, Ryan said [...] on [...] Last night, the vice-presidential candidate gave The New Yorker a



“The race was more than 20 years ago, but my brother Tobin – who ran Boston last year – reminds me that he is the owner of the fastest marathon in the family and has never himself ran a sub-three. If I were to do any rounding, it would certainly be to four hours, not three. He gave me a good ribbing over this at dinner tonight.”

Runners – and Mr. Ryan says he continues to be one – also just don’t forget race times. They talk about them with their friends; they think about them when running.

Source: Nicholas Thompson: LIE OR MISTAKE? PAUL RYAN’S MARATHONING PAST. The New Yorker, September 1, 2012

# In Search of Meaning: Entity Disambiguation

In an interview with radio host Hugh Hewitt last week, **Ryan** said he's run a sub-3:00 marathon [...]  
Last night, the **vice-presidential candidate** gave The New Yorker a statement:

“The race was more than 20 years ago, but my brother Tobin – who ran Boston last year – reminds me that he is the owner of the fastest marathon in the family and has never himself ran a sub-three. If I were to do any rounding, it would certainly be to four hours, not three. He gave me a good ribbing over this at dinner tonight.”

Runners – and **Mr. Ryan** says he continues to be one – also just don't forget race times. They talk about them with their friends; they think about them when running.



# In Search of Meaning: Entity Disambiguation

In an interview with radio host Hugh Hewitt last week, **Ryan** said he's run a sub-3:00 marathon [...]  
Last night, the **vice-presidential candidate** gave The New Yorker a story

"The race was more than 20 years ago, but my brother Tobin – who rarely tells me that he is the owner of the fastest marathon in the family and has 100 medals. If I were to do any rounding, it would certainly be to four hours, not three. I'm not ribbing over this at dinner tonight."

Runners – and **Mr. Ryan** says he continues to be one – also just don't forget about them with their friends; they think about them when running.

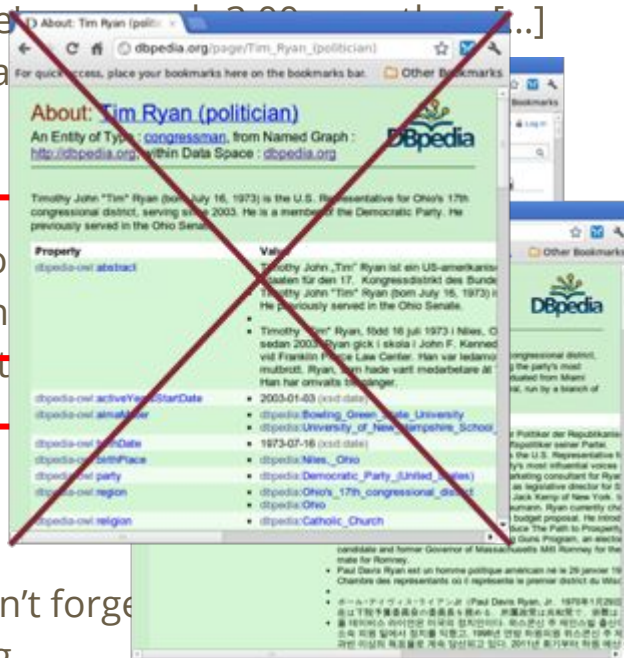


# In Search of Meaning: Entity Disambiguation

In an interview with radio host Hugh Hewitt last week, **Ryan** said he had run a marathon last night. Last night, the **vice-presidential candidate** gave The New Yorker a

“The race was more than 20 years ago, but my brother Tobin – who told me that he is the owner of the fastest marathon in the family and had run it himself. If I were to do any rounding, it would certainly be to four hours, not three. I was ribbing over this at dinner tonight.”

Runners – and **Mr. Ryan** says he continues to be one – also just don’t forget to talk about them with their friends; they think about them when running.



# In Search of Meaning: Metonymy

In an interview with radio host Hugh Hewitt last week, Ryan said he's run a sub-3:00 marathon [...] Last night, the vice-presidential candidate gave The New Yorker a statement:

“The race was more than 20 years ago, but my brother Tobin – who ran **Boston** last year – reminds me that he is the owner of the fastest marathon in the family and has never himself ran a sub-three. If I were to do any rounding, it would certainly be to four hours, not three. He gave me a good ribbing over this at dinner tonight.”

Runners – and Mr. Ryan says he continues to be one – also just don't forget race times. They talk about them with their friends; they think about them when running.

# In Search of Meaning: Metonymy

In an  
Last n



“The m  
me th  
If I we  
ribbin

last week, Ryan said he's run a sub-3:00 marathon [...] gave The New Yorker a statement:

my brother Tobin – who ran **Boston** last year – reminds me of a marathon in the family and has never himself ran a sub-three. It may only be to four hours, not three. He gave me a good

Runners – and Mr. Ryan says he continues to be one – also just don't forget race times. They talk about them with their friends; they think about them when running.

# In Search of Meaning: Metonymy

In an interview with *The New Yorker* last week, Ryan said he's run a sub-3:00 marathon [...]  
Last November, he gave *The New Yorker* a statement:

"The name of the race is **Boston**. It reminds me that I'm a runner. If I were not a runner, I would be ribbing the person who ran **Boston** last year – reminds me that I'm a runner and has never himself ran a sub-three. He gave me a good ribbing, not three. He gave me a good ribbing."



Runners – and Mr. Ryan says that they don't forget race times. They talk about them with their friends; they think about them when running.

# Auf der Suche nach Bedeutung und Sinn...

We have the same problem for virtually any language...



# Who is it?

“I like to drive. I can't drive any more.”

“Hey, I'm a nationalist and a globalist. I'm both.”

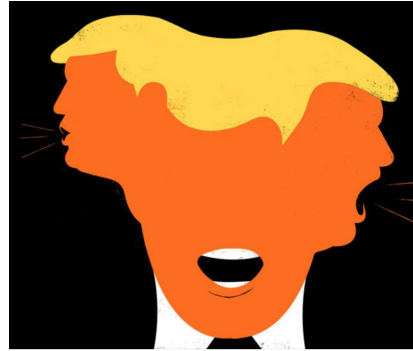
“I've actually been an activist Democrat and Republican”

# Who is it?

"I like to drive. I can't drive any more."

"Hey, I'm a nationalist and a globalist. I'm both."

"I've actually been an activist Democrat and Republican"



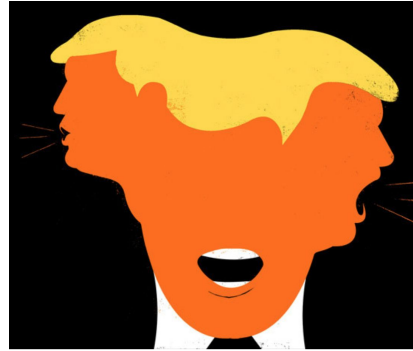


# Who is it?

"I like to drive. I can't drive any more."

"Hey, I'm a nationalist and a globalist. I'm both."

"I've actually been an activist Democrat and Republican"



When I have to build a hotel, we're bombing the hell out of them. Lots of money. To those suffering, I say vote for Donald. #SyriaStrikes

[My next gag order will be on] journalism. They're the problem. I'll educate our country and get rid of politics. @RVAwonk #globalgagrule

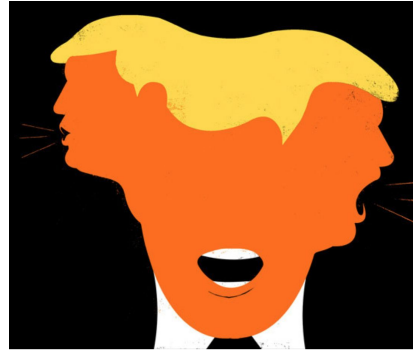
Here's the thing, I horribly abuse women and LGBT citizens. You know that better than anybody. That's my plan to win. @UlrichJvV #ElectionDay

# Who is it?

“I like to drive. I can't drive any more.”

“Hey, I'm a nationalist and a globalist. I'm both.”

“I've actually been an activist Democrat and Republican”



When I have to build a hotel, we're bombing the hell out of them. Lots of money. To those suffering, I say vote for Donald. #SyriaStrikes

[My next gag order will be on] journalism. They're the problem. I'll educate our country and get rid of politics.  
@RVAwonk #globalgagrule

Here's the thing, I horribly abuse women and LGBT citizens. You know that better than anybody. That's my plan to win.@UlrichJvV #ElectionDay

# So *why* Entities?!

- **Entities “make sense” in order for humans to make sense of text** - i.e., they provide useful *anchor points* to build **interpretable** semantic representations of sentences and documents

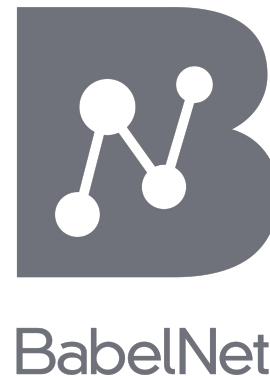
# So *why* Entities?!

- **Entities “make sense” in order for humans to make sense of text** - i.e., they provide useful *anchor points* to build **interpretable** semantic representations of sentences and documents
- Knowledge-rich or entity-based NLP models might not yield the best performance on a wide range of complex tasks (as opposed to distributional methods)...

# So *why* Entities?!

- **Entities “make sense” in order for humans to make sense of text** - i.e., they provide useful *anchor points* to build **interpretable** semantic representations of sentences and documents
- Knowledge-rich or entity-based NLP models might not yield the best performance on a wide range of complex tasks (as opposed to distributional methods)...
- ... BUT they are **quite useful when interacting with users (because of interpretability)**!

# Reference Knowledge Bases



... and many others!

# Entity Linking Tools



... and many others!



# Three Tasks

1. Query-based entity ranking
2. Building entity-centric event collections
3. Entity-aspect linking



# Web Search: What is Running?

Let's start with one of the most popular high-end tasks: Web search

# Web Search: What is Running?

Let's start with one of the most popular high-end tasks: Web search

Query „*running technique*“

# Web Search: What is Running?

Let's start with one of the most popular high-end tasks: Web search

Query „*running technique*“

2016 => **10 blue links**

## [Tips for Proper Running Form - Verywell Fit](https://www.verywellfit.com)

<https://www.verywellfit.com> › Fitness › Running › Beginners ▼

May 14, 2018 - Follow these 10 tips for proper running form to improve your performance. Learn the right posture, foot motion, and arm use to run your best.

## [Natural Running Form - USAF Marathon](https://www.usafmarathon.com/natural-running-form/)

<https://www.usafmarathon.com/natural-running-form/> ▼

Running Form Principle 1: Run Tall Posture. Run tall – imagine your column being stacked under your head. Look straight ahead to the horizon. Ball of foot and heel are level on ground. To move forward lean in like giving a kiss.

## [Learn Pose Method of running technique to run faster! | Pose Method](https://posemethod.com/running/)

<https://posemethod.com/running/> ▼

Running technique in 3 easy steps. This method works for runners of all levels & is used by Physical Therapists to address running injuries.

## [The Five Most Common Running Form Mistakes – Competitor Running](https://running.competitor.com/.../the-five-most-common-running-form-mistakes_48...)

[https://running.competitor.com/.../the-five-most-common-running-form-mistakes\\_48...](https://running.competitor.com/.../the-five-most-common-running-form-mistakes_48...) ▼

Jun 12, 2013 - Good running form is essential for performance and injury prevention. We take a look at the five most common running form mistakes.

## [Proper Running Technique >> Best 3 Ways - Runtastic](https://www.runtastic.com/blog/en/proper-running-technique/)

<https://www.runtastic.com/blog/en/proper-running-technique/> ▼

★★★★★ Rating: 4 - 1,256 reviews

Aug 8, 2017 - Looking to run faster, more efficient and injury-free? Try out these 3 ways to help with proper running technique.

## [Good Running Form for Beginners | ACTIVE](https://www.active.com)

<https://www.active.com> › Running › Articles ▼

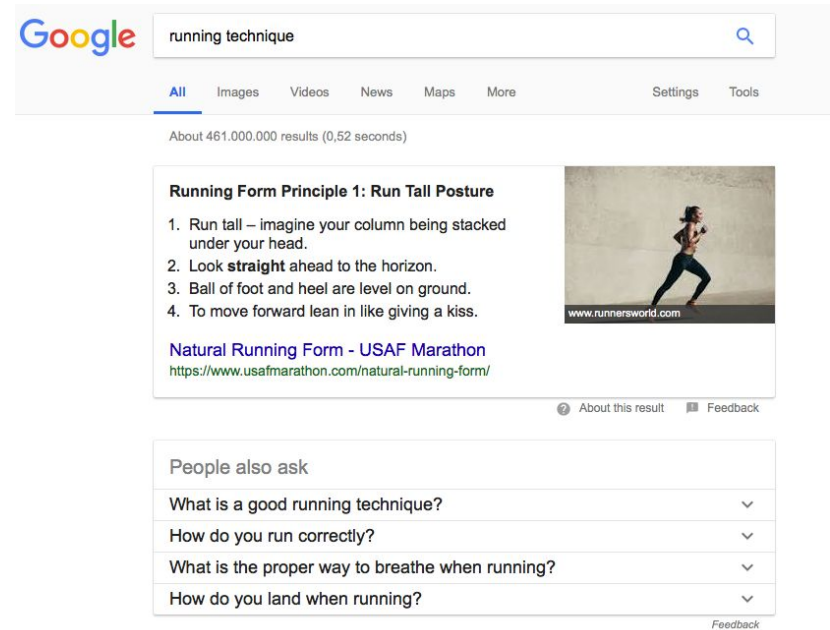
Learn proper running form as a beginner to avoid injuries and slow race times in the future. Use these visual cues from coach Brendan Courmane to identify in...

# Web Search: What is Running?

Let's start with one of the most popular high-end tasks: Web search

Query „running technique“

2018 => **featured snippets**



The image shows a Google search interface for the query "running technique". The search bar contains the text "running technique" and a magnifying glass icon. Below the search bar are navigation tabs for "All", "Images", "Videos", "News", "Maps", and "More", along with "Settings" and "Tools" on the right. The search results indicate "About 461,000,000 results (0,52 seconds)".

The featured snippet is titled "Running Form Principle 1: Run Tall Posture" and lists four points:

1. Run tall – imagine your column being stacked under your head.
2. Look **straight** ahead to the horizon.
3. Ball of foot and heel are level on ground.
4. To move forward lean in like giving a kiss.

To the right of the text is an image of a runner in motion, with the URL "www.runnersworld.com" at the bottom. Below the image are links for "Natural Running Form - USAF Marathon" and "https://www.usafmarathon.com/natural-running-form/".

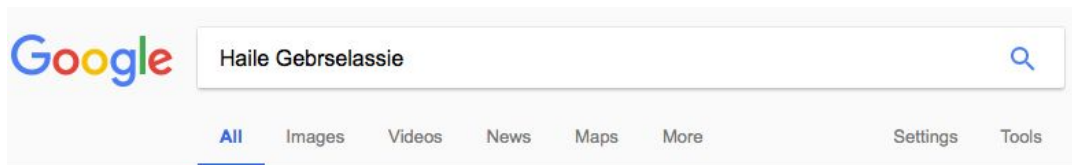
Below the featured snippet are links for "About this result" and "Feedback".

The "People also ask" section contains four questions, each with a dropdown arrow:

- What is a good running technique?
- How do you run correctly?
- What is the proper way to breathe when running?
- How do you land when running?

A "Feedback" link is located at the bottom right of the "People also ask" section.

# Entity-centric Queries



The image shows a Google search interface. The search bar contains the text "Haile Gebrselassie". Below the search bar, there are navigation tabs for "All", "Images", "Videos", "News", "Maps", and "More". The "All" tab is selected and underlined. To the right of the search bar, there are links for "Settings" and "Tools".

About 567.000 results (0,43 seconds)

## Haile Gebrselassie - Wikipedia

[https://en.wikipedia.org/wiki/Haile\\_Gebrselassie](https://en.wikipedia.org/wiki/Haile_Gebrselassie) ▼

Haile Gebrselassie is a retired Ethiopian long-distance track and road running athlete. He won two Olympic gold medals over 10,000 metres and four World ...

[Biography](#) · [Achievements](#) · [World records and best ...](#) · [Personal bests](#)

## Haile Gebrselassie – Wikipedia

[https://de.wikipedia.org/wiki/Haile\\_Gebrselassie](https://de.wikipedia.org/wiki/Haile_Gebrselassie) ▼ [Translate this page](#)

Haile Gebrselassie (amharisch ዓገሌ ገብረ ሥላሴ, \* 18. April 1973 in Assela, Region Oromia) ist ein ehemaliger äthiopischer Langstreckenläufer. Er stellte ...

Disziplin: [Langstreckenlauf](#) Größe: 164 cm

Nation: [Äthiopien](#) Gewicht: 56 kg


[Karriere](#) · [Bestzeiten](#)

## Haile GEBRSELASSIE | Profile | iaaf.org

<https://www.iaaf.org/athletes/athlete=8774> ▼

Ethiopia; DATE OF BIRTH 18 APR 1973 ATHLETE'S IAAF CODE 8774. Haile Gebrselassie (ETH) running in Edmonton World Championships (Getty Images).

# Entity-centric Queries

Google  

[All](#) [Images](#) [Videos](#) [News](#) [Maps](#) [More](#) [Settings](#) [Tools](#)

About 567.000 results (0,43 seconds)

## Haile Gebrselassie - Wikipedia

[https://en.wikipedia.org/wiki/Haile\\_Gebrselassie](https://en.wikipedia.org/wiki/Haile_Gebrselassie) ▼

Haile Gebrselassie is a retired Ethiopian long-distance track and road running athlete. He won two Olympic gold medals over 10,000 metres and four World ...

[Biography](#) · [Achievements](#) · [World records and best ...](#) · [Personal bests](#)

## Haile Gebrselassie – Wikipedia

[https://de.wikipedia.org/wiki/Haile\\_Gebrselassie](https://de.wikipedia.org/wiki/Haile_Gebrselassie) ▼ [Translate this page](#)

Haile Gebrselassie (amharisch ዓዲሌ ገብረ ሥላሴ, \* 18. April 1973 in Assela, Region Oromia) ist ein ehemaliger äthiopischer Langstreckenläufer. Er stellte ...

Disziplin: **Langstreckenlauf** Größe: 164 cm

Nation: Äthiopien Gewicht: 56 kg

[Karriere](#) · [Bestzeiten](#)

## Haile GEBRSELASSIE | Profile | iaaf.org

<https://www.iaaf.org/athletes/athlete=8774> ▼

Ethiopia; DATE OF BIRTH 18 APR 1973 ATHLETE'S IAAF CODE 8774. Haile Gebrselassie (ETH) running in Edmonton World Championships (Getty Images).



## Haile Gebrselassie

Olympiasportler

Haile Gebrselassie ist ein ehemaliger äthiopischer Langstreckenläufer. Er stellte insgesamt 26 Weltrekorde auf, dominierte als mehrfacher Olympiasieger und Weltmeister ein Jahrzehnt lang die Distanzen von 3000 Meter bis 10.000 Meter und war von 2007 bis 2011 Inhaber des Weltrekordes im Marathon. [Wikipedia](#)

**Geboren:** 18. April 1973 (Alter 45 Jahre), [Assela, Äthiopien](#)

**Größe:** 1,64 m

**Gewicht:** 54 kg



**Ehepartnerin:** [Alem Gebrselassie](#) (verh. 1996)

**Olympische Medaillen:** [Olympische Sommerspiele 2000/Leichtathletik – 10.000 m, MEHR](#)

**Filme:** [Der Einzelkämpfer von Atlanta](#)

# Query-based Entity Ranking

# Query-based Entity Ranking


Google   

---

**10 Components of Good Running Technique - Chi Running**  
[www.chirunning.com](http://www.chirunning.com) » [Blog](#) » [Danny Dreyer's Blog](#) ▾ [Diese Seite übersetzen](#)  
Is it possible to improve your **running technique**? Some people, even many doctors, believe that people are born with a given way of moving which they cannot ...

**Perfect Running Form | Runner's World**  
[www.runnersworld.com](http://www.runnersworld.com) » [Training](#) » [Running Tips](#) ▾ [Diese Seite übersetzen](#)  
16.06.2005 - **Head Tilt** How you hold your head is key to overall posture, which determines how efficiently you run. Let your gaze guide you. Look ahead ...

**Haile Gebrselassie Has Perfect Running Form | Runner's ...**  
[www.runnersworld.com](http://www.runnersworld.com) » [Training](#) » [Running Tips](#) ▾ [Diese Seite übersetzen](#)  
07.08.2007 - **GATHER 'ROUND KIDDIES, I'VE GOT SOME VALUABLE ADVICE** for you today. It's the type of thing every runner is interested in. I'm going to ...

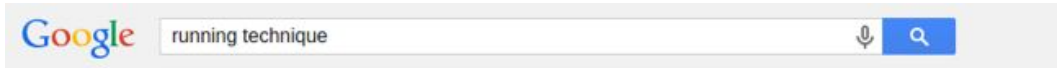
**Proper Running Technique: Running Form Tips and Drills ...**  
 [www.youtube.com/watch?v=wRkeBVMQsgg](http://www.youtube.com/watch?v=wRkeBVMQsgg) ▾  
27.09.2012 - Hochgeladen von Vo2maxProductions  
video by [www.Vo2maxProductions.com](http://www.Vo2maxProductions.com). Tips to keep for a more efficient **running stride** to improve speed ...

**Pose Method® of Running Technique | Pose Method®**  
<https://posemethod.com/running/> ▾ [Diese Seite übersetzen](#)  
**Running** is one of the most popular athletic activities in the world. There are probably more runners than there are participants in any other sport. **Running** is also ...

**How to Run with Proper Form and Technique | SparkPeople**  
[www.sparkpeople.com/resource/fitness\\_articles.asp?...](http://www.sparkpeople.com/resource/fitness_articles.asp?...) ▾ [Diese Seite übersetzen](#)  
Whether you're ready to jump into jogging or have completed dozens of races, this detailed primer on proper **running technique** can help you run better and



# Query-based Entity Ranking



## 10 Components of Good Running Technique - Chi Running

[www.chirunning.com](#) » Blog » Danny Dreyer's Blog » Diese Seite übersetzen  
Is it possible to improve your **running technique**? Some people, even many doctors, believe that people are born with a given way of moving which they cannot ...

## Perfect Running Form | Runner's World

[www.runnersworld.com](#) » Training » Running Tips » Diese Seite übersetzen  
16.06.2005 - Head Tilt How you hold your head is key to overall posture, which determines how efficiently you run. Let your gaze guide you. Look ahead ...

## Haile Gebrselassie Has Perfect Running Form | Runner's ...

[www.runnersworld.com](#) » Training » Running Tips » Diese Seite übersetzen  
07.08.2007 - GATHER 'ROUND KIDDIES, I'VE GOT SOME VALUABLE ADVICE for you today. It's the type of thing every runner is interested in. I'm going to ...

## Proper Running Technique: Running Form Tips and Drills ...



[www.youtube.com/watch?v=wRkeBVMQSGg](#) »  
27.09.2012 - Hochgeladen von Vo2maxProductions  
video by [www.Vo2maxProductions.com](#). Tips to keep for a more efficient **running stride** to improve speed ...

## Pose Method® of Running Technique | Pose Method®

[https://posemethod.com/running/](#) » Diese Seite übersetzen  
**Running** is one of the most popular athletic activities in the world. There are probably more runners than there are participants in any other sport. **Running** is also ...

## How to Run with Proper Form and Technique | SparkPeople

[www.sparkpeople.com/resource/fitness\\_articles.asp?...](#) » Diese Seite übersetzen  
Whether you're ready to jump into jogging or have completed dozens of races, this detailed primer on **proper running technique** can help you run better and

## Running

From Wikipedia, the free encyclopedia

**Running** is a means of **terrestrial locomotion** in athletics terms as a **gait** in which at to **walking**, where one foot is always in vaults over the stance leg or legs in a viewpoint of spring-mass mechanics it



## Haile Gebrselassie

From Wikipedia, the free encyclopedia

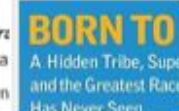
**Haile Gebrselassie** (**Amharic**: ያሌ ክፍለ ወንጌል) is an Ethiopian track and road running athlete. He won four titles in the event. He won the Berlin Marathon. Further to this, he won four



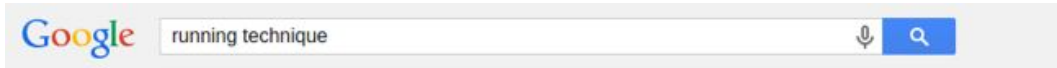
## Born to Run

From Wikipedia, the free encyclopedia

**Born to Run: A Hidden Tribe, Superathletes, and the Greatest Race** is an ethnography written by the American author and journalist Michael Chabon. In *Born to Run*, McDougall tracks down



# Query-based Entity Ranking



## 10 Components of Good Running Technique - Chi Running

[www.chirunning.com](http://www.chirunning.com) › Blog › Danny Dreyer's Blog › Diese Seite übersetzen  
Is it possible to improve your **running technique**? Some people, even many doctors, believe that people are born with a given way of moving which they cannot ...

## Perfect Running Form | Runner's World

[www.runnersworld.com](http://www.runnersworld.com) › Training › Running Tips › Diese Seite übersetzen  
16.06.2005 - Head Tilt How you hold your head is key to overall posture, which determines how efficiently you run. Let your gaze guide you. Look ahead ...

## Haile Gebrselassie Has Perfect Running Form | Runner's ...

[www.runnersworld.com](http://www.runnersworld.com) › Training › Running Tips › Diese Seite übersetzen  
07.08.2007 - GATHER 'ROUND KIDDIES, I'VE GOT SOME VALUABLE ADVICE for you today. It's the type of thing every runner is interested in. I'm going to ...

## Proper Running Technique: Running Form Tips and Drills ...



[www.youtube.com/watch?v=wRkeBVMQsgg](http://www.youtube.com/watch?v=wRkeBVMQsgg) ›  
27.09.2012 - Hochgeladen von Vo2maxProductions  
video by [www.Vo2maxProductions.com](http://www.Vo2maxProductions.com). Tips to keep for a more efficient **running stride** to improve speed ...

## Pose Method® of Running Technique | Pose Method®

<https://posemethod.com/running/> › Diese Seite übersetzen  
**Running** is one of the most popular athletic activities in the world. There are probably more runners than there are participants in any other sport. **Running** is also ...

## How to Run with Proper Form and Technique | SparkPeople

[www.sparkpeople.com/resource/fitness\\_articles.asp?...](http://www.sparkpeople.com/resource/fitness_articles.asp?...) › Diese Seite übersetzen  
Whether you're ready to jump into jogging or have completed dozens of races, this detailed primer on **proper running technique** can help you run better and

## Running

From Wikipedia, the free encyclopedia

**Running** is a means of **terrestrial locomotion** in athletics terms as a **gait** in which at to **walking**, where one foot is always in vaults over the stance leg or legs in a viewpoint of spring-mass mechanics it



1

## Haile Gebrselassie

From Wikipedia, the free encyclopedia

**Haile Gebrselassie** (**Amharic**: ያሌ ክፍለ ወንጌል) is a track and road running athlete. He won titles in the event. He won the Berlin Marathon. Further to this, he won four



3

## Born to Run

From Wikipedia, the free encyclopedia

**Born to Run: A Hidden Tribe, Super** ethnography written by the American a In *Born to Run*, McDougall tracks down

**BORN TO**  
A Hidden Tribe, Super  
and the Greatest Race  
Has Never Seen

2

# Entity Retrieval for General Web Queries

Michael Schuhmacher, Laura Dietz, Simone Paolo Ponzetto: **Ranking Entities for Web Queries Through Text and Knowledge**. *CIKM 2015*: 1461-1470

# Entity Retrieval for General Web Queries

Michael Schuhmacher, Laura Dietz, Simone Paolo Ponzetto: **Ranking Entities for Web Queries Through Text and Knowledge**. *CIKM 2015*: 1461-1470

What we want:

- Retrieve entities for general web query  
*"Argentine British relationships"* (TREC Web Track)

# Entity Retrieval for General Web Queries

Michael Schuhmacher, Laura Dietz, Simone Paolo Ponzetto: **Ranking Entities for Web Queries Through Text and Knowledge**. *CIKM 2015*: 1461-1470

What we want:

- Retrieve entities for general web query  
*"Argentine British relationships"* (TREC Web Track)

What we not aim for:

- Answer single entity queries: *"Jonas Salk"*
- Answer entity type queries  
*"European countries where I can pay with Euros"* (INEX)

# Entity Retrieval for General Web Queries

Michael Schuhmacher, Laura Dietz, Simone Paolo Ponzetto: **Ranking Entities for Web Queries Through Text and Knowledge**. *CIKM 2015*: 1461-1470

What we want:

- Retrieve entities for general web query  
*"Argentine British relationships"* (TREC Web Track)

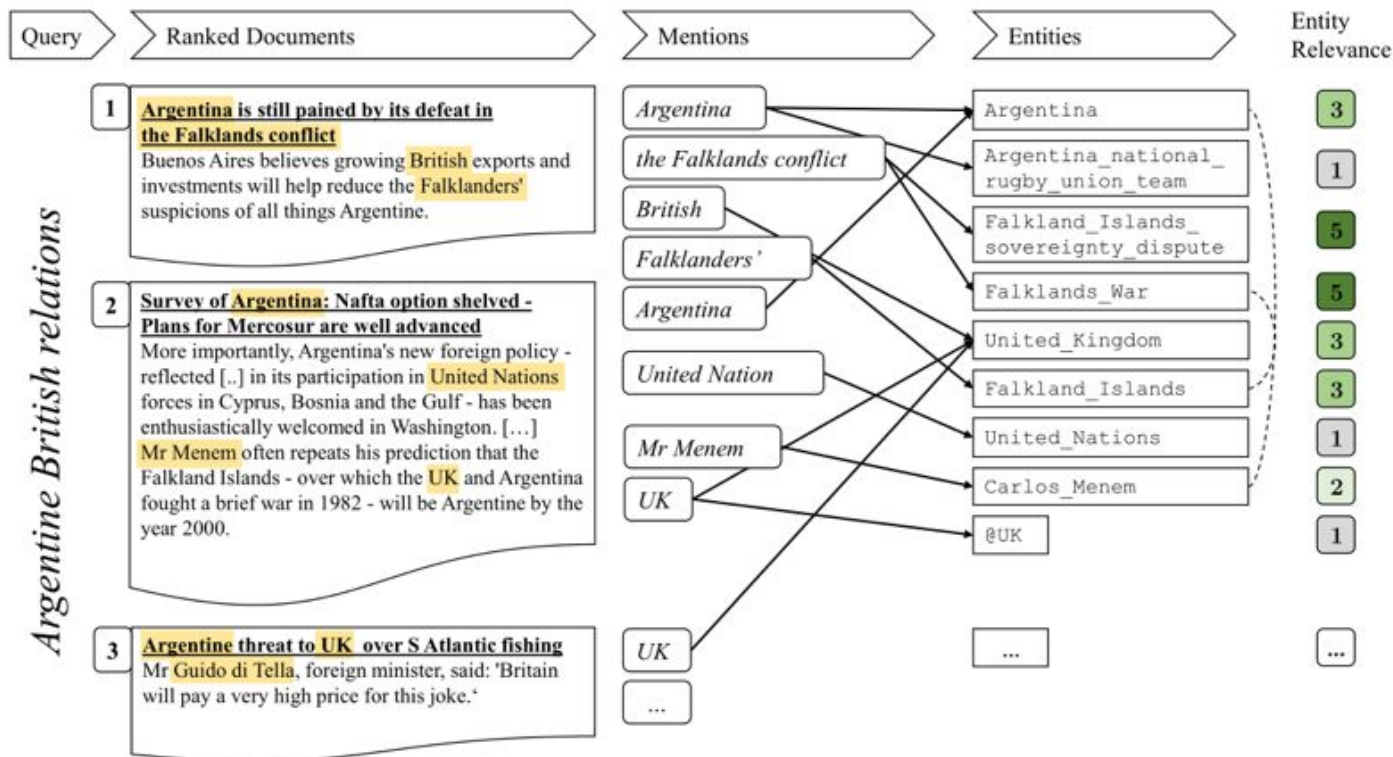
What we not aim for:

- Answer single entity queries: *"Jonas Salk"*
- Answer entity type queries  
*"European countries where I can pay with Euros"* (INEX)

How we do it:

1. Retrieve Documents
2. Extract and Re-rank Entities

# Running Example



# Research Question and Experimental Framework

Given: The extracted entities

Question: Which features are helpful for ranking?



# Research Question and Experimental Framework

Given: The extracted entities

Question: Which features are helpful for ranking?

Method Framework: Learning-to-Rank (LTR)

- Entity Relevance Features
- Labeled Data Set
- LTR Method

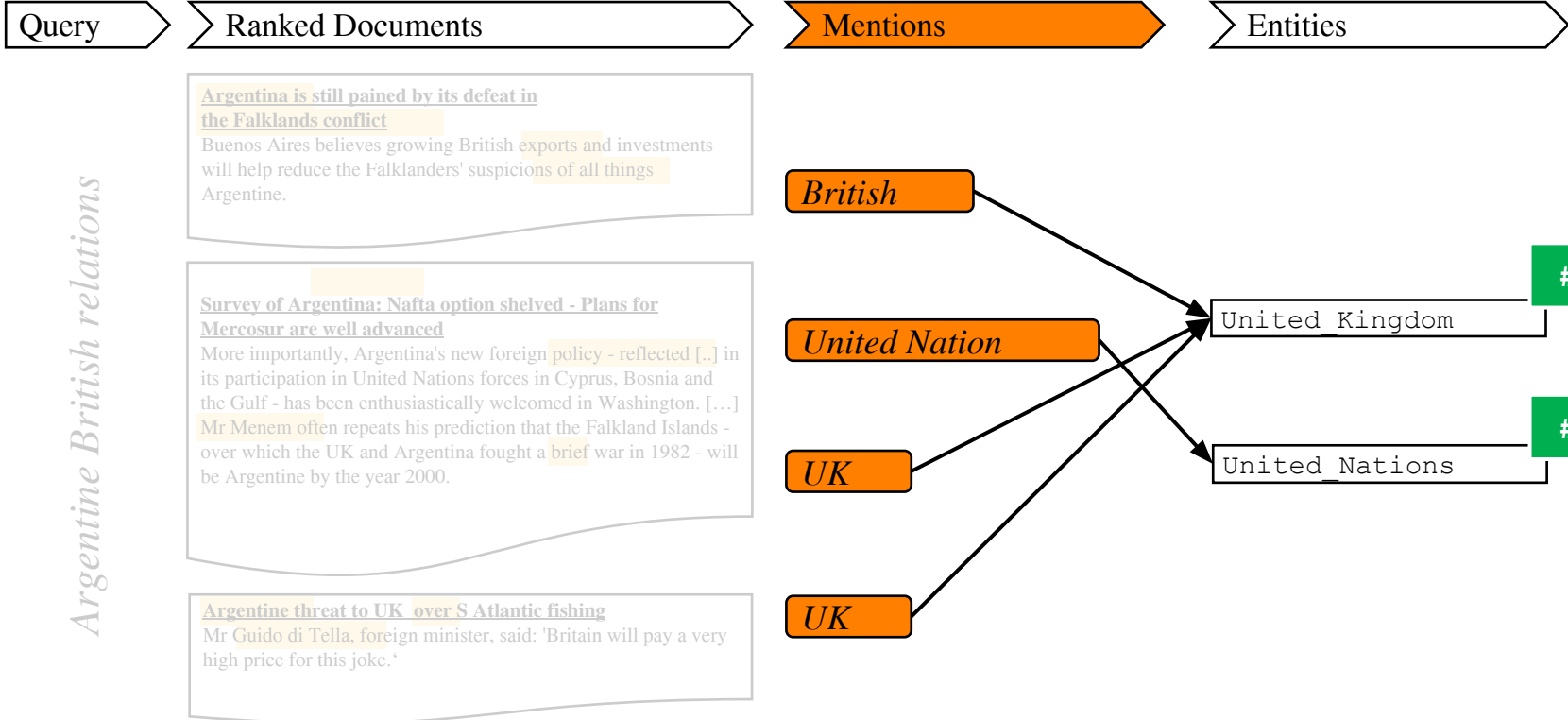
# Entity Ranking Features in our LTR Setting

Learning-to-Rank (LTR)

- Entity Relevance Features
  1. Mention Features
  2. Query-Mention Features
  3. Query-Entity Features
  4. Entity-Entity Features
- Labeled Data Set
- LTR Method

# Mention Features

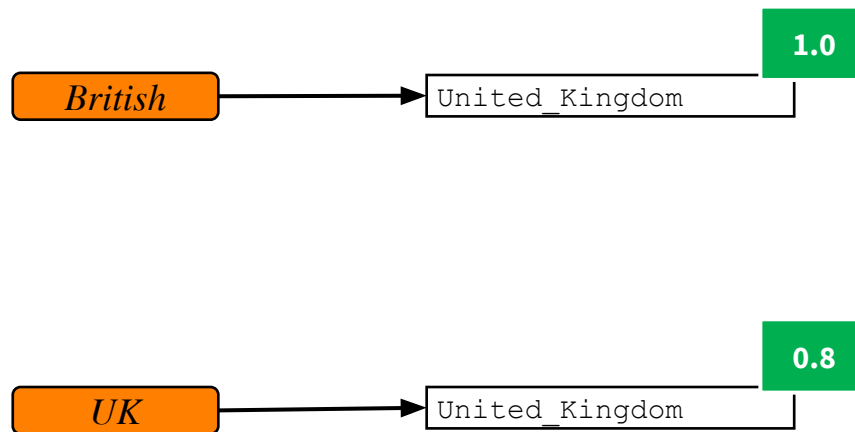
Argentine British relations



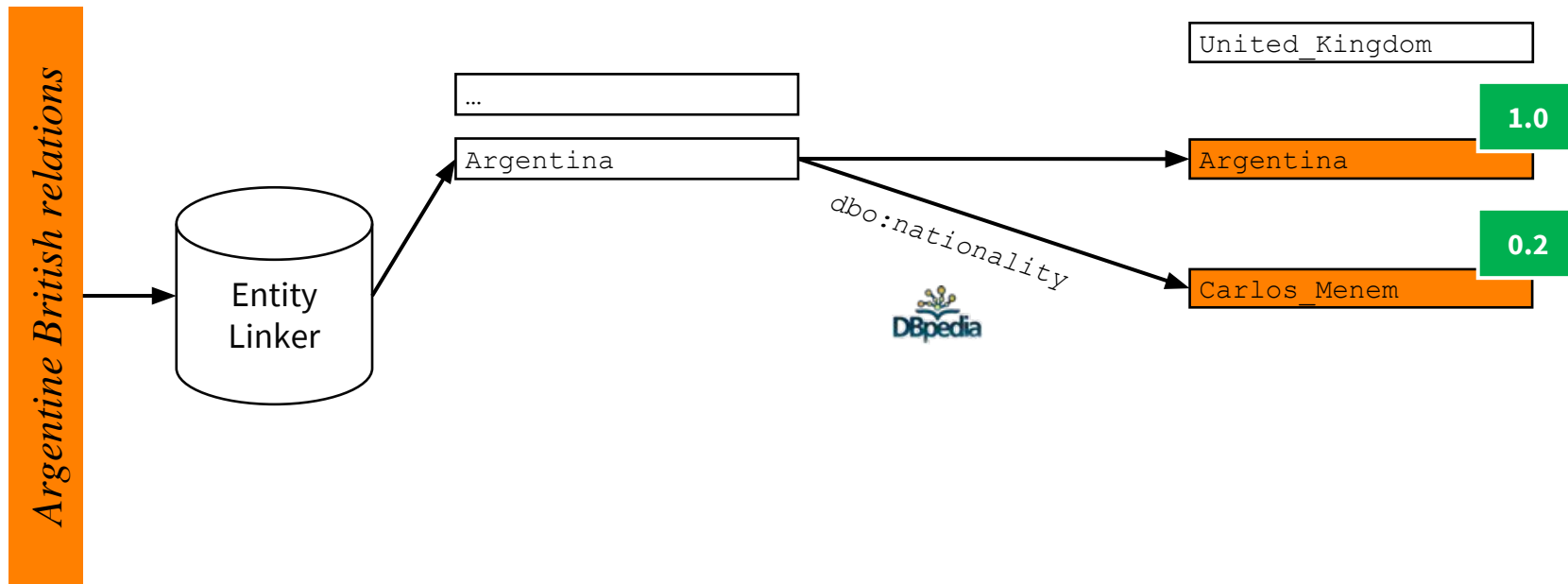
# Query-Mention Features



*Argentine British relations*



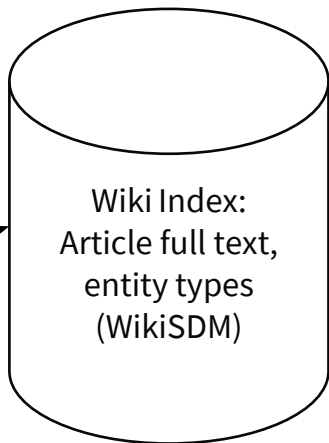
# Query-Entity Features (Query -> Entity Linker -> Entities)



# Query-Entity Features (Query -> Wikipedia Index)



Argentine British relations

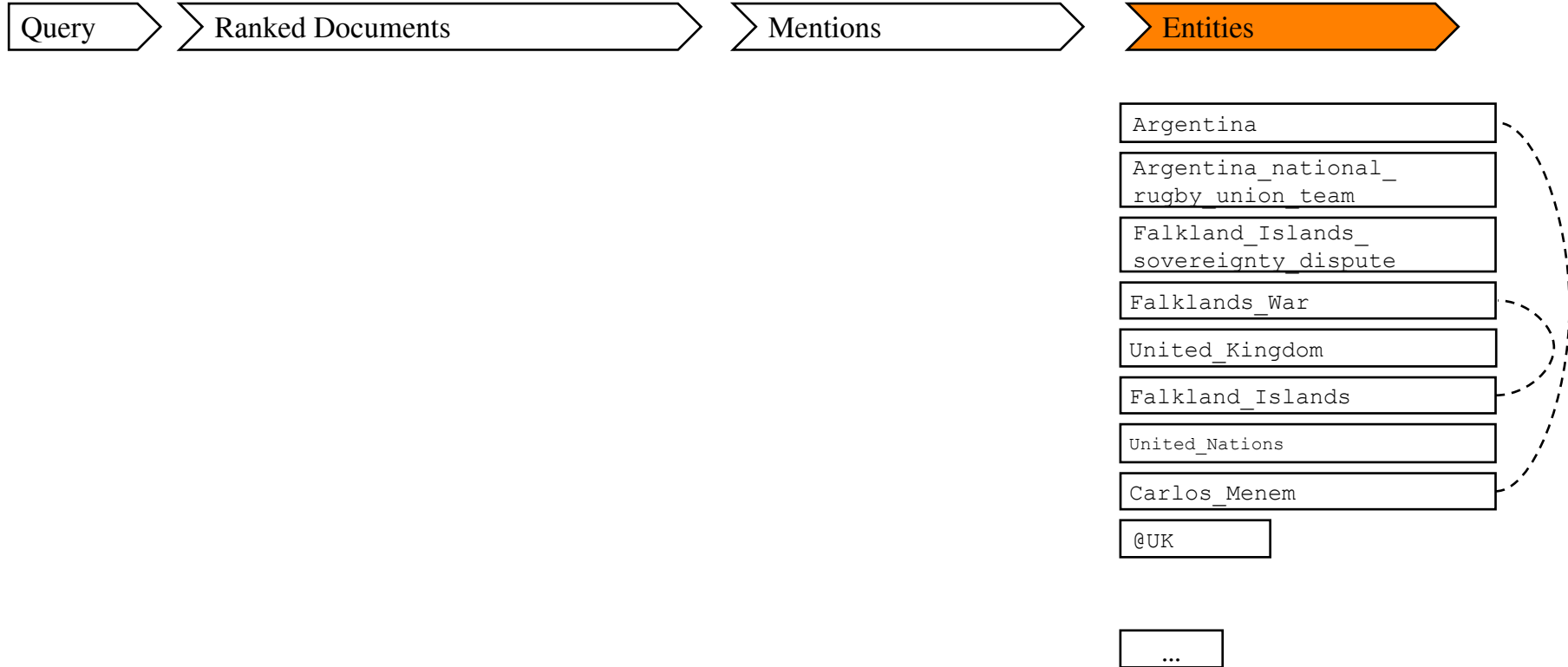


United Kingdom

2.6

Carlos Menem

# Entity-Entity Features (using Dbpedia to compare entities)



# Entity-Entity Features (using Dbpedia to compare entities)

## Entities

Argentina

Carlos\_Menem

Falklands\_War

Falkland\_Islands



Carlos\_Menem dbo:nationality Argentina



Falklands\_War dbo:place Falkland\_Islands



# REWQ Datasets

## Learning-to-Rank (LTR)

- Entity Relevance Features

- Labeled Data Set

	REWQ Robust04	REWQ Clueweb12
Doc collection	TREC Disk4+5	ClueWeb12
Queries	TREC Robust Track '04	TREC Web Track '13/14
# queries	25	22
Doc retrieval	EQFE (w/ entities)	SDM (text-only)
Top-k docs	19	20
Top-k entities	50	all
Entity linker	KB Bridge	FACC1
GS annotations	graded, 1-5	binary 0/1

- LTR Method

# Learning Algorithms

## Learning-to-Rank (LTR)

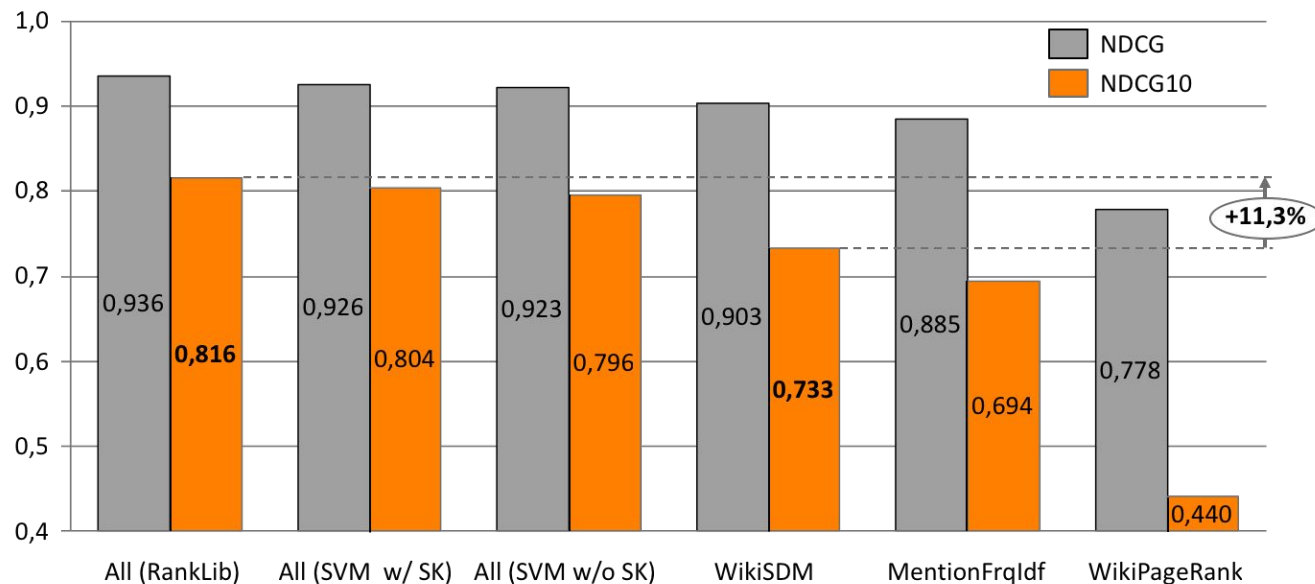
- Entity Relevance Features
- Labeled Data Set
- LTR Method

	<b>Ranking SVM</b>	<b>RankLib</b>
LTR type	Pairwise	Listwise
Method/Impl.	Joachims Ranking SVM	RankLib Coordinate Asc
Optimized metric	Disordered pairs	MAP

# Retrieved Ranking for REWQ Rob04 (selected examples)

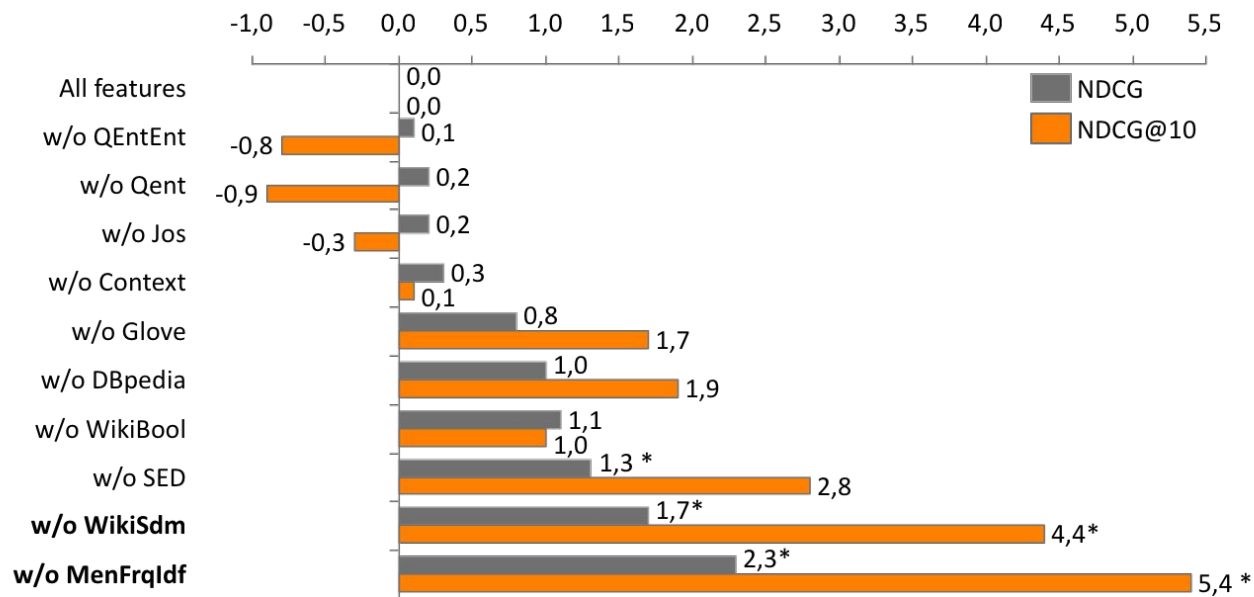
gt	ndcg	query	top-1 entity	top-2 entity	top-3 entity
5.0	.895	schengen agreement	Schengen_Agreement	Schengen_Area	Schengen_Information_System
..	..	...	...	...	...
4.3	.879	poliomyelitis and post polio	Poliomyelitis	Polio_vaccine	Jonas_Salk
..	..	...	...	...	...
3.3	.748	argentine british relations	Foreign_relations_of_Argentina	Argentina_national_rugby_team	Falklands_War
..	..	...	...	...	...
1.9	.966	agoraphobia	Charles_M._Schulz	Snoopy	UGM-27_Polaris

# Overall Results on REWQ Robust04



Feature combination sign. better than best single feature (+11.3% NDCG@10)

# Feature Ablation Study (leave one-out) on REWQ Robust04



Doc Mention Frq and Wikipedia KB most important performance drivers

# The Story so Far...

Michael Schuhmacher, Laura Dietz, Simone Paolo Ponzetto: **Ranking Entities for Web Queries Through Text and Knowledge**. *CIKM 2015*: 1461-1470

# The Story so Far...

Michael Schuhmacher, Laura Dietz, Simone Paolo Ponzetto: **Ranking Entities for Web Queries Through Text and Knowledge**. *CIKM 2015*: 1461-1470

We contributed an in-depth study of doc-based entity ranking

# The Story so Far...

Michael Schuhmacher, Laura Dietz, Simone Paolo Ponzetto: **Ranking Entities for Web Queries Through Text and Knowledge**. *CIKM 2015*: 1461-1470

## We contributed an in-depth study of doc-based entity ranking

- Combing document with entity retrieval gives overall high NDCG/MAP
- Two simple, but complementary feature are already helpful
  - **Document-based**: Mention frequency in query-specific documents (MenFrqIdf)
  - **Knowledge-based**: Full-text SDM on Wikipedia (WikiSDM)



# The Story so Far...

Michael Schuhmacher, Laura Dietz, Simone Paolo Ponzetto: **Ranking Entities for Web Queries Through Text and Knowledge**. *CIKM 2015*: 1461-1470

## We contributed an in-depth study of doc-based entity ranking

- Combing document with entity retrieval gives overall high NDCG/MAP
- Two simple, but complementary feature are already helpful
  - **Document-based**: Mention frequency in query-specific documents (MenFrqIdf)
  - **Knowledge-based**: Full-text SDM on Wikipedia (WikiSDM)
- We created two new gold standards for query-specific entity ranking

**REWQ datasets available at <http://rewq.dwslab.de>**

# Building Entity-Centric Event Collections

Federico Nanni, Simone Paolo Ponzetto, Laura Dietz:

**Building Entity-Centric Event Collections.** *JCDL 2017*: 199-208

# Building Entity-Centric Event Collections

Federico Nanni, Simone Paolo Ponzetto, Laura Dietz:  
**Building Entity-Centric Event Collections.** *JCDL 2017*: 199-208

Part of a larger research project in collaboration with:

- Federico Nanni (DWS Group Mannheim)
- Laura Dietz (University of New Hampshire)
- Nikolay Marinov (University of Houston)

# Building Entity-Centric Event Collections

Federico Nanni, Simone Paolo Ponzetto, Laura Dietz:  
**Building Entity-Centric Event Collections.** *JCDL 2017*: 199-208

Part of a larger research project in collaboration with:

- Federico Nanni (DWS Group Mannheim)
- Laura Dietz (University of New Hampshire)
- Nikolay Marinov (University of Houston)

**Overall focus:** using information from text and knowledge graphs to support research in **international relations**.

# Political Scientists Study Events



# Named Events



Identified by a common name: the **Orange Revolution**.

# Named Events



Identified by a common name: the **Orange Revolution**.

Different from Event Extraction!

*In 2004, as part of the Orange Revolution, the constitution was changed.*

# Named Events



Identified by a common name: the **Orange Revolution**.

Different from Event Extraction!

*In 2004,* <sup>When</sup> *as part of the Orange Revolution,* <sup>What</sup> *the constitution was changed.* <sup>Event</sup>



# Named Events



Identified by a common name: the **Orange Revolution**.

Different from Event Extraction!

*In 2004, as part of the Orange Revolution, the constitution was changed.*

# Retrospective Analyses

Understanding how/whether **international events** have been discussed and perceived by **politicians, media, society**.



# Retrospective Analyses

Understanding how/whether **international events** have been discussed and perceived by **politicians, media, society**.



The New York Times



CONGRESS.GOV

# Current Solution

Document filtering: collect all documents that contain **the event-name**.

# Current Solution

Document filtering: collect all documents that contain the event-name.



Orange Revolution  
(Ukraine - 2004)

# Our Assumption

If you focus on the name of the event, you risk missing **its early stages**.



Orange Revolution  
(Ukraine - 2004)

# Our Assumption

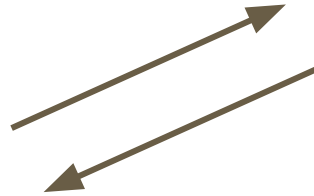
If you focus on the name of the event, you risk missing **its early stages**.

Ukrainian presidential election, 2004



← 1999    October 31, November 21, December 26, 2004    2010 →

		
Nominee	Viktor Yushchenko	Viktor Yanukovich
Party	Independent	Party of Regions
Popular vote	15,115,712	12,848,528
Percentage	51.99%	44.20%



Orange Revolution  
(Ukraine - 2004)

# Our Approach

Using **related concepts** and **entities** to retrieve a comprehensive set of relevant documents.



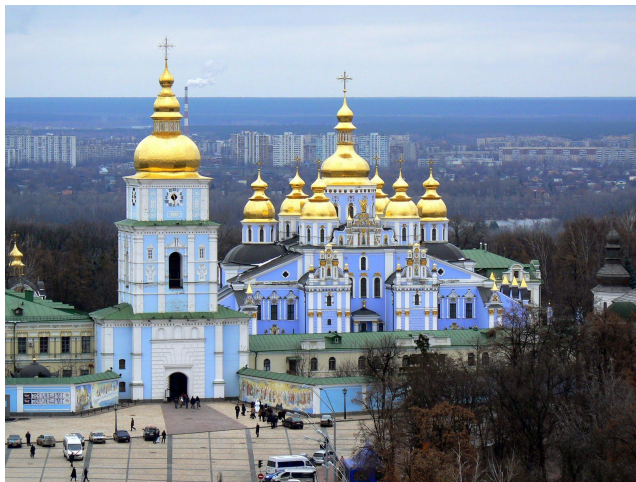
# Our Approach

Using **related concepts** and **entities** to retrieve a comprehensive set of relevant documents.

Yulia Tymoshenko



Kiev

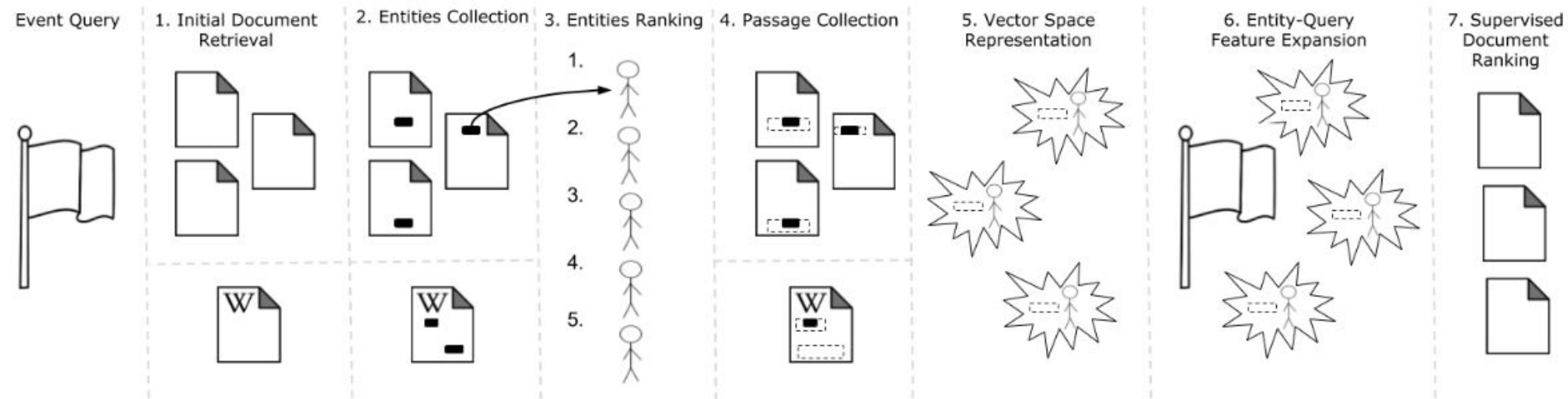


Russia



# Our Approach

Using **related concepts** and **entities** to retrieve a comprehensive set of relevant documents.



# Case Studies

## Types of event:

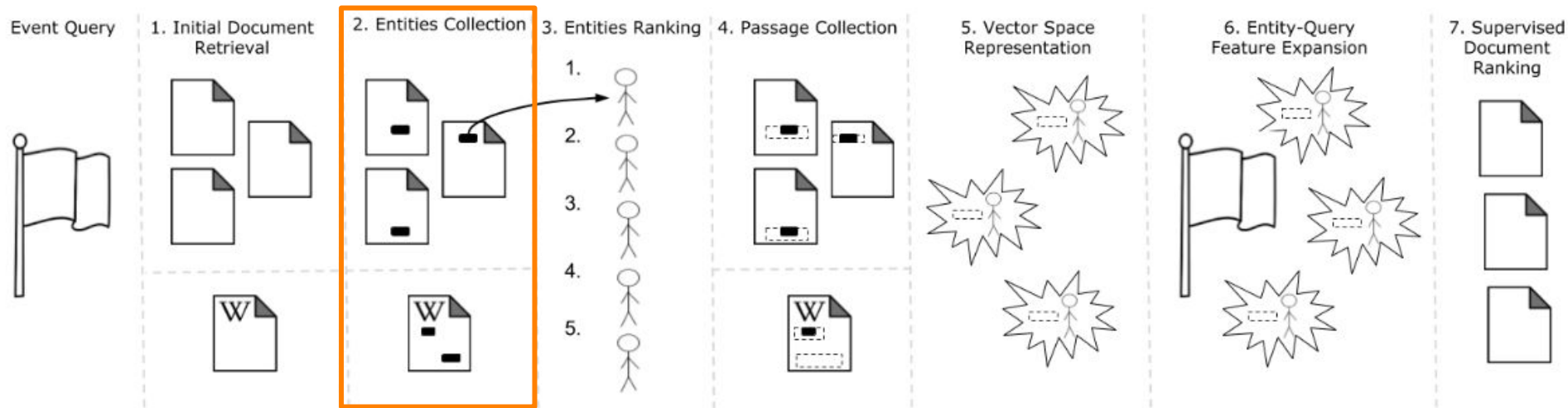
- 1) 15 unexpected elections
- 2) 15 political crises
- 3) 15 civil wars

## Datasets:

- 1) New York Times Corpus
- 2) US Congressional Record
- 3) TREC KBA Stream Corpus

# Our Approach

Using **related concepts** and **entities** to retrieve a comprehensive set of relevant documents.



# Collecting Entities

Collect relevant entities from:

- a) An initial pool of relevant documents (all entities mentioned in the **context** of the event-name).
- b) Wikipedia page of the event, as **outlinks**.

# Collecting Entities

Collect relevant entities from:

- An initial pool of relevant documents (all entities mentioned in the **context** of the event-name).
- Wikipedia page of the event, as **outlinks**.

<b>Method</b>	<b>Prec</b>	<b>Rec</b>	<b>F1</b>
NELDA			
Info-Box			
Context			
Outlinks			
<b>Cont+Out</b>			

# Collecting Entities

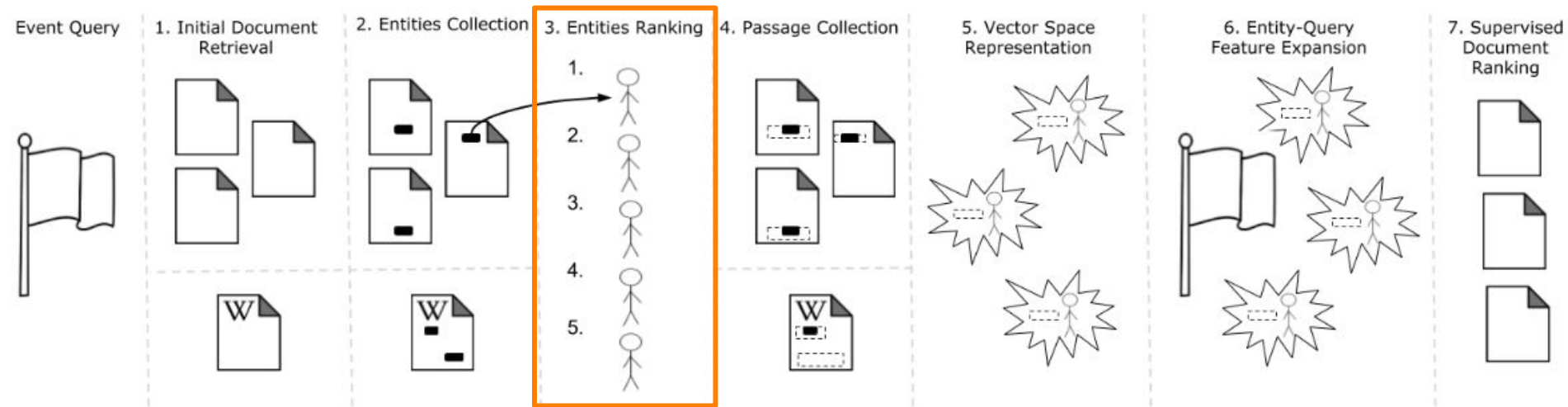
Collect relevant entities from:

- An initial pool of relevant documents (all entities mentioned in the **context** of the event-name).
- Wikipedia page of the event, as **outlinks**.

<b>Method</b>	<b>Prec</b>	<b>Rec</b>	<b>F1</b>
NELDA	<b>1.00 ± 0.00</b>	0.13 ± 0.02	0.23 ± 0.02
Info-Box	0.88 ± 0.03	0.27 ± 0.05	0.41 ± 0.05
Context	0.52 ± 0.04	0.60 ± 0.05	0.55 ± 0.05
Outlinks	0.89 ± 0.03	0.53 ± 0.05	0.66 ± 0.05
<b>Cont+Out</b>	0.74 ± 0.04	<b>1.00 ± 0.00</b>	<b>0.85 ± 0.05</b>

# Our Approach

Using **related concepts** and **entities** to retrieve a comprehensive set of relevant documents.





# Ranking Entities

We use **RDF2Vec** (Ristoski et al., 2016) and rank by the cosine similarity of vector representation for each entity and the event.

# Ranking Entities

We use **RDF2Vec** (Ristoski et al., 2016) and rank by the cosine similarity of vector representation for each entity and the event.

<b>Method</b>	<b>MAP</b>	<b>P@5</b>	<b>P@10</b>
ContFreq			
CheapEntRel			
<b>RDF2Vec</b>			

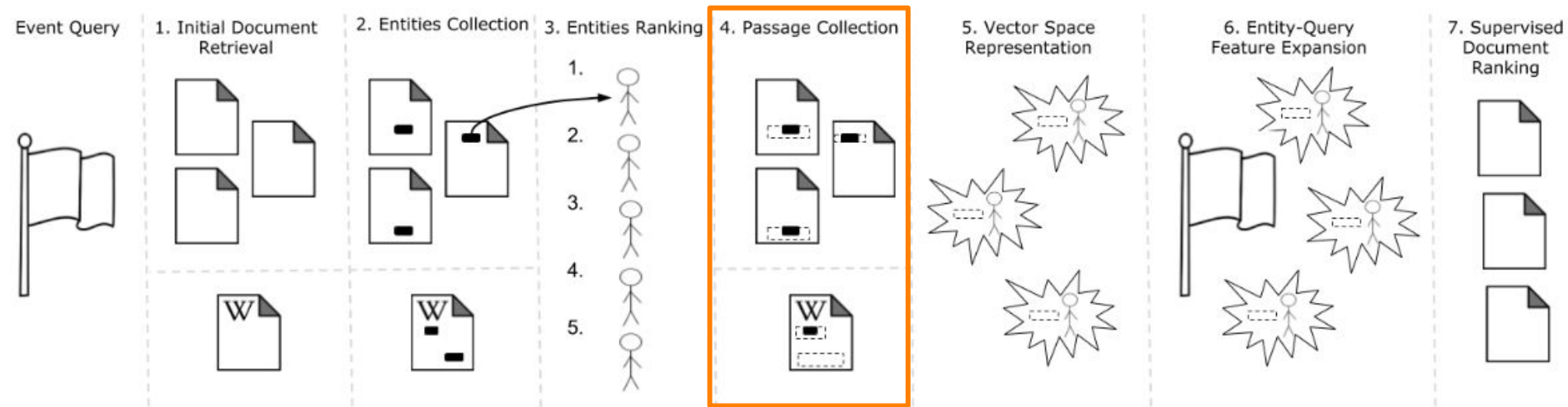
# Ranking Entities

We use **RDF2Vec** (Ristoski et al., 2016) and rank by the cosine similarity of vector representation for each entity and the event.

<b>Method</b>	<b>MAP</b>	<b>P@5</b>	<b>P@10</b>
ContFreq	0.22 ± 0.03	0.55 ± 0.06	0.48 ± 0.05
CheapEntRel	0.51 ± 0.05	0.70 ± 0.05	0.62 ± 0.05
<b>RDF2Vec</b>	<b>0.65 ± 0.05</b>	<b>0.80 ± 0.06</b>	<b>0.74 ± 0.05</b>

# Our Approach

Using **related concepts** and **entities** to retrieve a comprehensive set of relevant documents.



# Getting Contextual Passages

For each relevant entity, we collect a **text snippet** from the **Wikipedia page of the event** by retrieving the first passage (i.e., three sentences) that contains a mention of the entity.

# Getting Contextual Passages

For each relevant entity, we collect a **text snippet** from the **Wikipedia page of the event** by retrieving the first passage (i.e., three sentences) that contains a mention of the entity.

<b>Method</b>	<b>Prec</b>	<b>Rec</b>	<b>F1</b>
Wiki-Intro			
NYT-Pass			
USC-Pass			
<b>Wiki-Pass</b>			

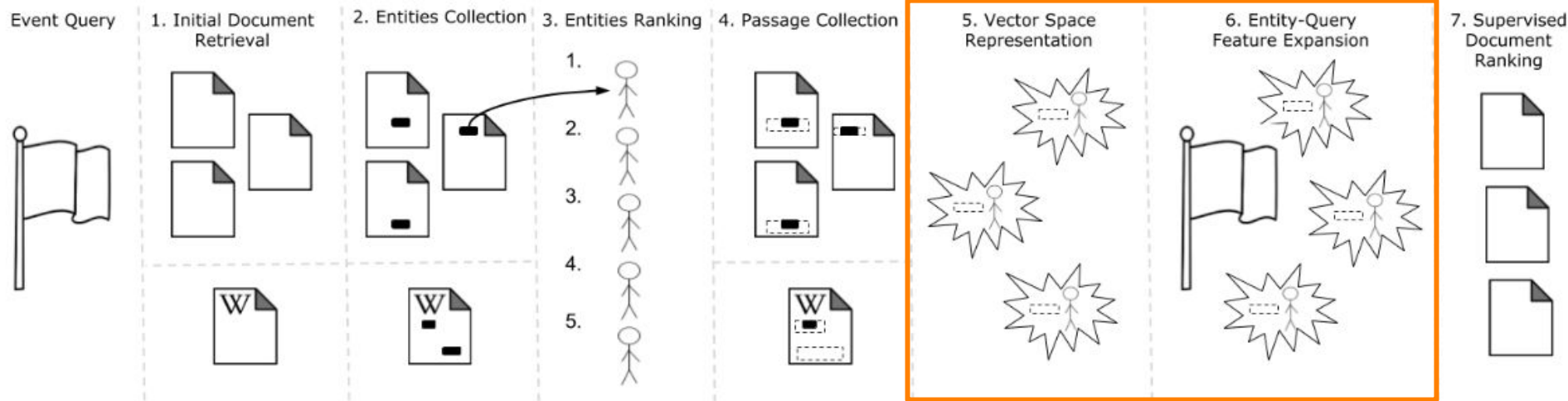
# Getting Contextual Passages

For each relevant entity, we collect a **text snippet** from the **Wikipedia page of the event** by retrieving the first passage (i.e., three sentences) that contains a mention of the entity.

<b>Method</b>	<b>Prec</b>	<b>Rec</b>	<b>F1</b>
Wiki-Intro	0.45 ± 0.03	<b>1.00 ± 0.00</b>	0.62 ± 0.03
NYT-Pass	<b>0.99 ± 0.03</b>	0.36 ± 0.03	0.53 ± 0.03
USC-Pass	0.92 ± 0.03	0.19 ± 0.03	0.31 ± 0.03
<b>Wiki-Pass</b>	<b>0.99 ± 0.03</b>	0.83 ± 0.04	<b>0.89 ± 0.03</b>

# Our Approach

Using **related concepts** and **entities** to retrieve a comprehensive set of relevant documents.





# Query Expansion Approaches

Place

Entities

Passages (i.e., entities in context)

GloVe vector representation of Entities

GloVe vector representation of Passages (**our-light**)

# Query Expansion Approaches

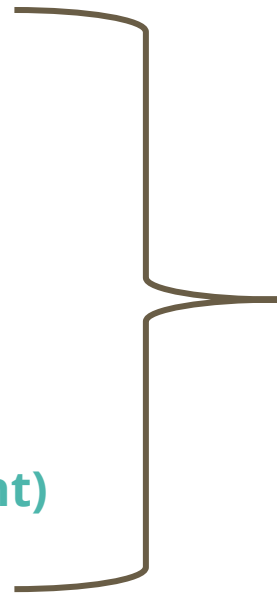
Place

Entities

Passages (i.e., entities in context)

GloVe vector representation of Entities

GloVe vector representation of Passages (**our-light**)



**Cosine similarity**  
between  
expanded query  
and document

# Query Expansion Approaches

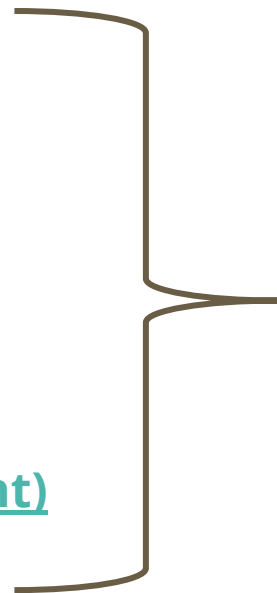
Place

Entities

Passages (i.e., entities in context)

GloVe vector representation of Entities

GloVe vector representation of Passages (our-light)



**Cosine similarity**  
between  
expanded query  
and document



All combined with Learning to Rank (our-full)

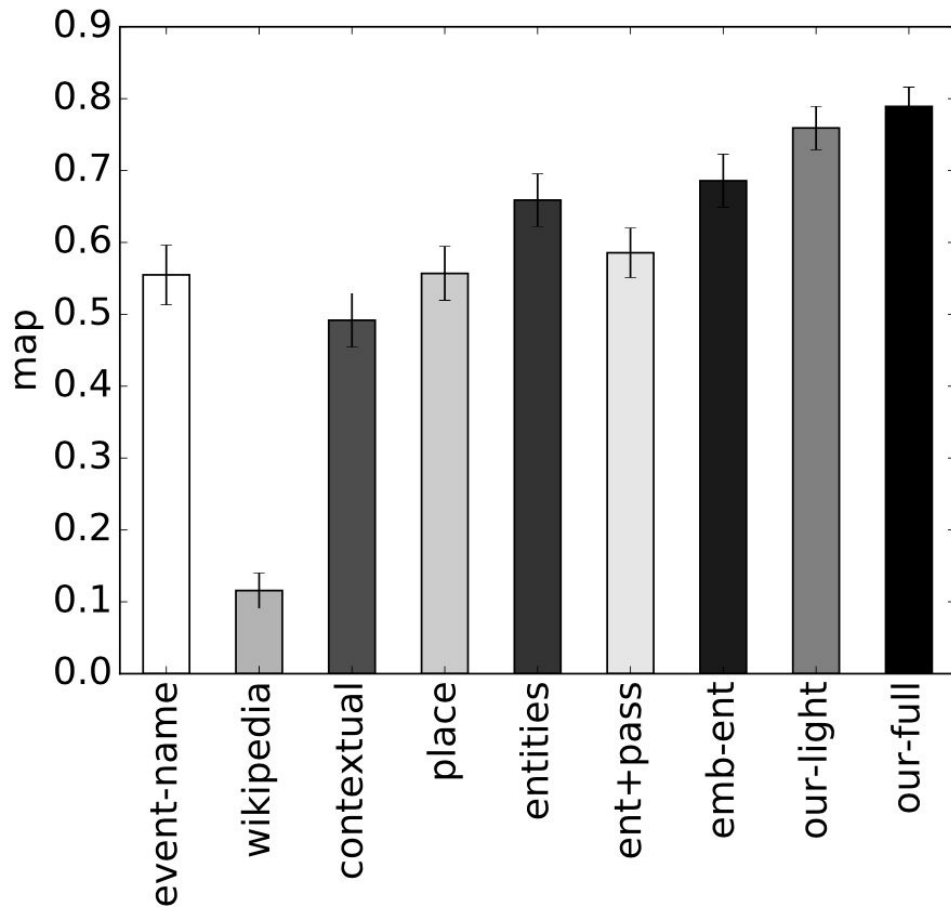
# Baselines

Event-name

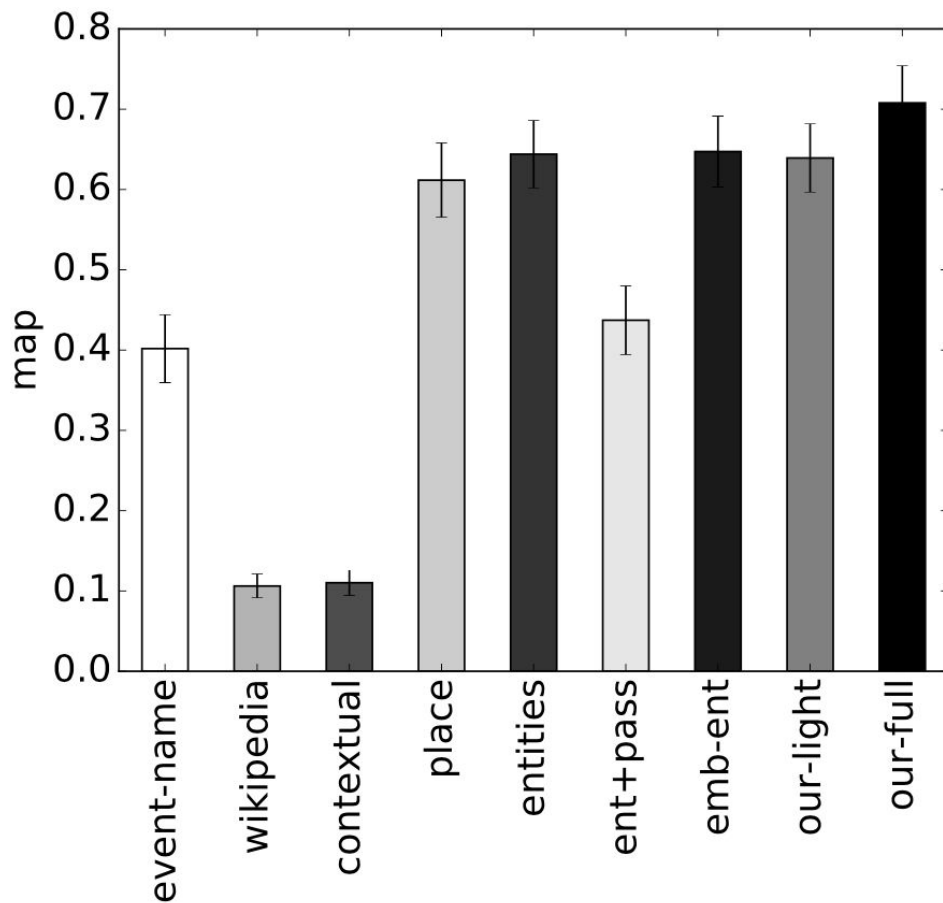
Wikipedia BoW

Contextual passages BoW

# Evaluation on NY Times



# Evaluation on US Congressional Record



# MAP for Types of Events

**Table 4: MAP for types of events on the NYT Corpus.**

Method	Elections	Crises	Wars
event-name	$0.64 \pm 0.06$	$0.39 \pm 0.06$	$0.61 \pm 0.06$
entities	$0.63 \pm 0.05$	$0.59 \pm 0.06$	$0.76 \pm 0.04$
our-light	$0.72 \pm 0.05$	$0.73 \pm 0.06$	$0.83 \pm 0.04$
our-full	<b><math>0.76 \pm 0.04</math></b>	<b><math>0.74 \pm 0.06</math></b>	<b><math>0.86 \pm 0.04</math></b>

**Table 5: MAP for types of events on the USC Corpus.**

Method	Elections	Crises	Wars
event-name	$0.32 \pm 0.07$	$0.38 \pm 0.06$	$0.52 \pm 0.06$
entities	$0.65 \pm 0.07$	$0.63 \pm 0.06$	$0.65 \pm 0.06$
our-light	$0.52 \pm 0.06$	<b><math>0.70 \pm 0.05</math></b>	$0.73 \pm 0.06$
our-full	<b><math>0.73 \pm 0.05</math></b>	$0.63 \pm 0.09$	<b><math>0.77 \pm 0.08</math></b>

# MAP for Types of Events

Table 4: MAP for types of events on the NYT Corpus.

Method	Elections	Crises	Wars
event-name	$0.64 \pm 0.06$	$0.39 \pm 0.06$	$0.61 \pm 0.06$
entities	$0.63 \pm 0.05$	$0.59 \pm 0.06$	$0.76 \pm 0.04$
our-light	$0.72 \pm 0.05$	$0.73 \pm 0.06$	$0.83 \pm 0.04$
our-full	<b><math>0.76 \pm 0.04</math></b>	<b><math>0.74 \pm 0.06</math></b>	<b><math>0.86 \pm 0.04</math></b>

Table 5: MAP for types of events on the USC Corpus.

Method	Elections	Crises	Wars
event-name	$0.32 \pm 0.07$	$0.38 \pm 0.06$	$0.52 \pm 0.06$
entities	$0.65 \pm 0.07$	$0.63 \pm 0.06$	$0.65 \pm 0.06$
our-light	$0.52 \pm 0.06$	<b><math>0.70 \pm 0.05</math></b>	$0.73 \pm 0.06$
our-full	<b><math>0.73 \pm 0.05</math></b>	$0.63 \pm 0.09$	<b><math>0.77 \pm 0.08</math></b>



## % of Docs Missed Using the Event-Name

Type of Event	Before	After
Elections	16% ± 6	22% ± 7
<b>Crises</b>	63% ± 9	31% ± 6
Wars	14% ± 4	8% ± 2
All	30% ± 5	20% ± 4

# Qualitative Study on TREC KBA



# The story so far 2...

Federico Nanni, Simone Paolo Ponzetto, Laura Dietz:  
**Building Entity-Centric Event Collections.** *JCDL 2017*: 199-208

# The story so far 2...

Federico Nanni, Simone Paolo Ponzetto, Laura Dietz:  
**Building Entity-Centric Event Collections.** *JCDL 2017*: 199-208

To build comprehensive event-collections it is **essential** to use a combination of relevant **entities** and **concepts**.

# The story so far 2...

Federico Nanni, Simone Paolo Ponzetto, Laura Dietz:  
**Building Entity-Centric Event Collections.** *JCDL 2017*: 199-208

To build comprehensive event-collections it is **essential** to use a combination of relevant **entities** and **concepts**.

This is particularly true when studying the premises and early-stages of events such as **political crises**.

Data, resources, etc. @ <http://federiconanni.com/event-collections/>

# The story so far 2...

Federico Nanni, Simone Paolo Ponzetto, Laura Dietz:  
**Building Entity-Centric Event Collections.** *JCDL 2017*: 199-208

To build comprehensive event-collections it is **essential** to use a combination of relevant **entities** and **concepts**.

This is particularly true when studying the premises and early-stages of events such as **political crises**.

Data, resources, etc. @ <http://federiconanni.com/event-collections/>

Next Step: Identifying **entity-aspects** and **topics** for improving event-related document retrieval and collection building.

# Entity Linking

# Entity Linking

I'm watching the debate between Clinton and Sanders.



# Entity Linking

I'm watching the debate between Clinton and Sanders

# Entity Linking



I'm watching the debate between Clinton and Sanders



# Entity Linking



I'm watching the debate between Clinton and Sanders



# Limits of Entity Linking

# Limits of Entity Linking

I'm watching the debate between Clinton and Sanders.

Hillary Rodham Clinton majored in political science.

Hillary Clinton is a former Senator for New York.

# Limits of Entity Linking

I'm watching the debate between **Clinton** and Sanders.

**Hillary Rodham Clinton** majored in political science.

**Hillary Clinton** is a former Senator for New York.



# Limits of Entity Linking



I'm watching the debate between Clinton and Sanders.

Hillary Rodham Clinton majored in political science.

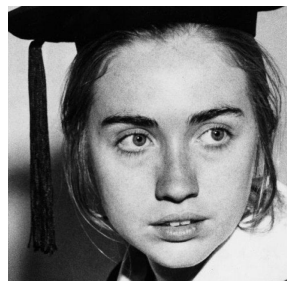
Hillary Clinton is a former Senator for New York.

# Limits of Entity Linking

I'm watching the debate between Clinton and Sanders.

Hillary Rodham Clinton majored in political science.

Hillary Clinton is a former Senator for New York.





# Limits of Entity Linking

I'm watching the debate between Clinton and Sanders.

Hillary Rodham Clinton majored in political science.

Hillary Clinton is a former Senator for New York.



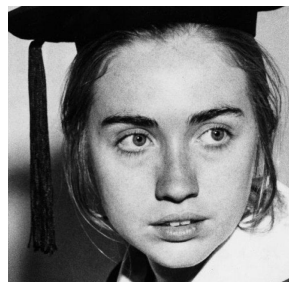
# Limits of Entity Linking



I'm watching the debate between Clinton and Sanders.

Hillary Rodham Clinton majored in political science.

Hillary Clinton is a former Senator for New York.



# Solution: Entity-Aspect Linking!

I'm watching the debate between Clinton and Sanders.

Hillary Rodham Clinton majored in political science.

Hillary Clinton is a former Senator for New York.

# Solution: Entity-Aspect Linking!

I'm watching the debate between Clinton and Sanders.

Hillary Rodham Clinton majored in political science.

Hillary Clinton is a former Senator for New York.

Hillary Clinton



67th United States Secretary of State

**In office**

January 21, 2009 – February 1, 2013

**President** Barack Obama

**Deputy** James Steinberg  
William Joseph Burns

**Preceded by** Condoleezza Rice

**Succeeded by** John Kerry

**United States Senator  
from New York**

**In office**

January 3, 2001 – January 21, 2009

**Preceded by** Daniel Patrick Moynihan

**Succeeded by** Kirsten Gillibrand

# Solution: Entity-Aspect Linking!

I'm watching the debate between **Clinton** and Sanders.

**Hillary Rodham Clinton** majored in political science.

**Hillary Clinton** is a former Senator for New York.

## Contents [\[hide\]](#)

- 1 Early life and education
  - 1.1 Early life
  - 1.2 Wellesley College years
  - 1.3 Yale Law School and postgraduate studies
- 2 Marriage, family, law career and First Lady of Arkansas
  - 2.1 From the East Coast to Arkansas
  - 2.2 Early Arkansas years
  - 2.3 Later Arkansas years
  - 2.4 Bill Clinton presidential campaign of 1992
- 3 First Lady of the United States
  - 3.1 Health care and other policy initiatives
  - 3.2 Whitewater and other investigations
  - 3.3 Response to Lewinsky scandal
  - 3.4 Traditional duties
- 4 United States Senate
  - 4.1 2000 U.S. Senate election
  - 4.2 First term
  - 4.3 2006 re-election campaign
  - 4.4 Second term
- 5 2008 presidential campaign
- 6 U.S. Secretary of State
  - 6.1 Nomination and confirmation
  - 6.2 First half of tenure
  - 6.3 Second half of tenure
  - 6.4 Overall themes
  - 6.5 Benghazi attack and subsequent hearings
- 7 Email controversy
- 8 Clinton Foundation and speeches
- 9 2016 presidential campaign
- 10 Post-2016 election activities

# Solution: Entity-Aspect Linking!

I'm watching the debate between Clinton<sup>1</sup> and Sanders.

Hillary Rodham Clinton majored in political science.

Hillary Clinton is a former Senator for New York.

Contents [hide]	
1	Early life and education
1.1	Early life
1.2	Wellesley College years
1.3	Yale Law School and postgraduate studies
2	Marriage, family, law career and First Lady of Arkansas
2.1	From the East Coast to Arkansas
2.2	Early Arkansas years
2.3	Later Arkansas years
2.4	Bill Clinton presidential campaign of 1992
3	First Lady of the United States
3.1	Health care and other policy initiatives
3.2	Whitewater and other investigations
3.3	Response to Lewinsky scandal
3.4	Traditional duties
4	United States Senate
4.1	2000 U.S. Senate election
4.2	First term
4.3	2006 re-election campaign
4.4	Second term
5	2008 presidential campaign
6	U.S. Secretary of State
6.1	Nomination and confirmation
6.2	First half of tenure
6.3	Second half of tenure
6.4	Overall themes
6.5	Benghazi attack and subsequent hearings
7	Email controversy
8	Clinton Foundation and speeches
9	2016 presidential campaign <b>1</b>
10	Post-2016 election activities

# Solution: Entity-Aspect Linking!

I'm watching the debate between Clinton<sup>1</sup> and Sanders.

Hillary Rodham Clinton<sup>2</sup> majored in political science.

Hillary Clinton is a former Senator for New York.

Contents <a href="#">[hide]</a>	
1	Early life and education <span style="float: right;">2</span>
1.1	Early life
1.2	Wellesley College years
1.3	Yale Law School and postgraduate studies
2	Marriage, family, law career and First Lady of Arkansas
2.1	From the East Coast to Arkansas
2.2	Early Arkansas years
2.3	Later Arkansas years
2.4	Bill Clinton presidential campaign of 1992
3	First Lady of the United States
3.1	Health care and other policy initiatives
3.2	Whitewater and other investigations
3.3	Response to Lewinsky scandal
3.4	Traditional duties
4	United States Senate
4.1	2000 U.S. Senate election
4.2	First term
4.3	2006 re-election campaign
4.4	Second term
5	2008 presidential campaign
6	U.S. Secretary of State
6.1	Nomination and confirmation
6.2	First half of tenure
6.3	Second half of tenure
6.4	Overall themes
6.5	Benghazi attack and subsequent hearings
7	Email controversy
8	Clinton Foundation and speeches
9	2016 presidential campaign <span style="float: right;">1</span>
10	Post-2016 election activities

# Solution: Entity-Aspect Linking!

I'm watching the debate between Clinton<sup>1</sup> and Sanders.

Hillary Rodham Clinton<sup>2</sup> majored in political science.

Hillary Clinton<sup>3</sup> is a former Senator for New York.

Contents <a href="#">[hide]</a>	
1	Early life and education <span style="float: right;">2</span>
1.1	Early life
1.2	Wellesley College years
1.3	Yale Law School and postgraduate studies
2	Marriage, family, law career and First Lady of Arkansas
2.1	From the East Coast to Arkansas
2.2	Early Arkansas years
2.3	Later Arkansas years
2.4	Bill Clinton presidential campaign of 1992
3	First Lady of the United States
3.1	Health care and other policy initiatives
3.2	Whitewater and other investigations
3.3	Response to Lewinsky scandal
3.4	Traditional duties
4	United States Senate <span style="float: right;">3</span>
4.1	2000 U.S. Senate election
4.2	First term
4.3	2006 re-election campaign
4.4	Second term
5	2008 presidential campaign
6	U.S. Secretary of State
6.1	Nomination and confirmation
6.2	First half of tenure
6.3	Second half of tenure
6.4	Overall themes
6.5	Benghazi attack and subsequent hearings
7	Email controversy
8	Clinton Foundation and speeches
9	2016 presidential campaign <span style="float: right;">1</span>
10	Post-2016 election activities



# Entity-Aspect Linking: A Few More Examples

Hillary Clinton  
2016 presidential  
campaign

Bernie Sanders  
2016 presidential  
campaign

1. I'm watching the debate between **Clinton** and **Sanders**



Brexit  
Impact on the United  
Kingdom

Issues in the United Kingdom  
European Union membership  
referendum, 2016  
Economy



2. **#Brexit** will destroy the **UK welfare state**



European migrant crisis  
Migrant routes, development  
and responses in individual  
countries

German Federal  
Election, 2017  
Government formation

3. The **refugee crisis** had an impact on the **German election**



# Entity-Aspect Linking

Federico Nanni, Simone Paolo Ponzetto, Laura Dietz:  
**Entity-Aspect Linking: Providing Fine-Grained Semantics of  
Entities in Context.** *JCDL 2018: 49-58*

Given: **an entity-mention in context** (e.g., tweet, sentence or paragraph)

Return: **a link to one from a set of predefined aspects** that captures the addressed topic

# Entity-Aspect Linking

Federico Nanni, Simone Paolo Ponzetto, Laura Dietz:  
**Entity-Aspect Linking: Providing Fine-Grained Semantics of Entities in Context.** *JCDL 2018*: 49-58

Given: **an entity-mention in context** (e.g., tweet, sentence or paragraph)

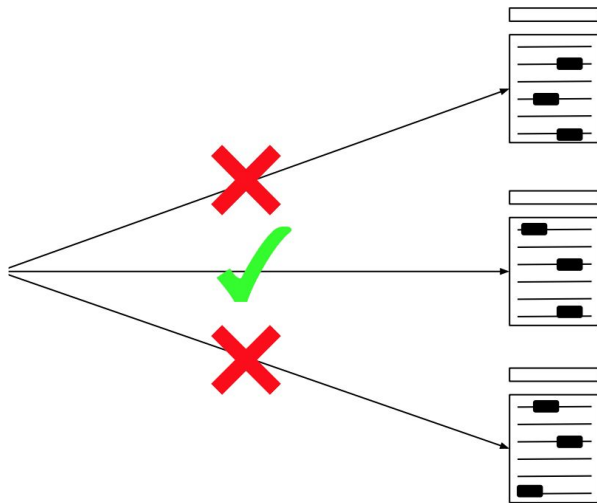
Return: **a link to one from a set of predefined aspects** that captures the addressed topic

- Each entity aspect is *accompanied by a textual description and heading*
- We focus on Wikipedia as reference resource, i.e., each *top-level section of a Wikipedia page* defines one of the entity's specific aspects



# Entity-Aspect Linking: How?

I'm watching the  
debate between  
**Clinton** and Sanders



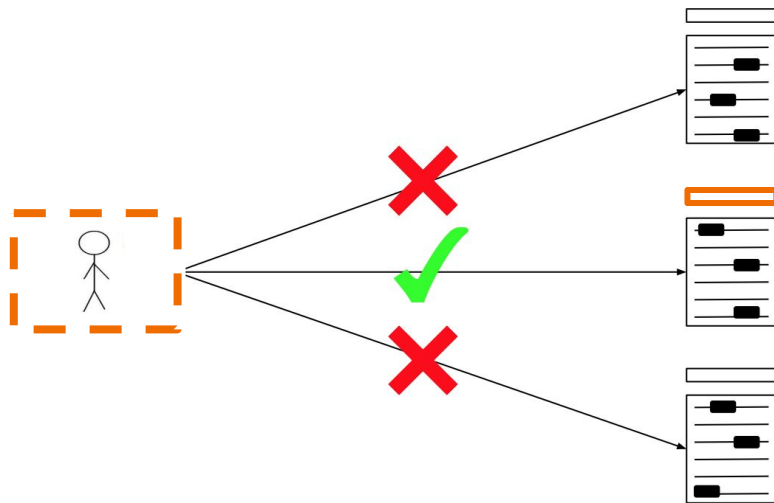
Early life and education

**2016 Elections**

US Senators

# Aspect Representations as Headers

I'm watching the  
debate between  
**Clinton** and Sanders



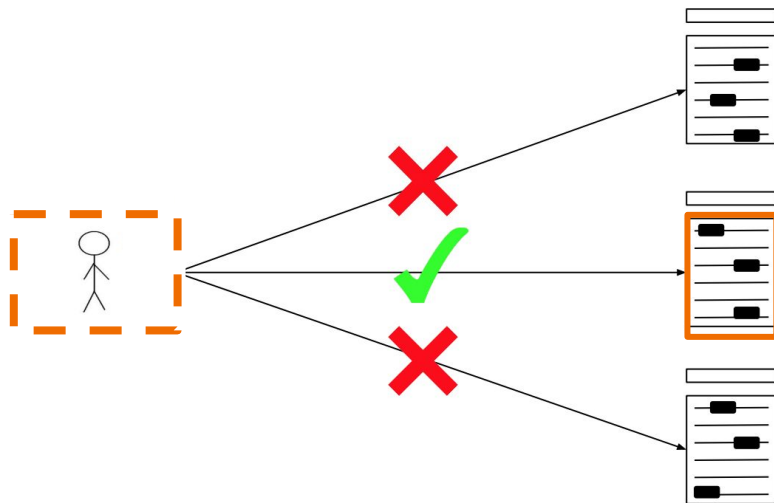
Early life and education

**2016 Elections**

US Senators

# Aspect Representations as Content

I'm watching the  
debate between  
**Clinton** and Sanders



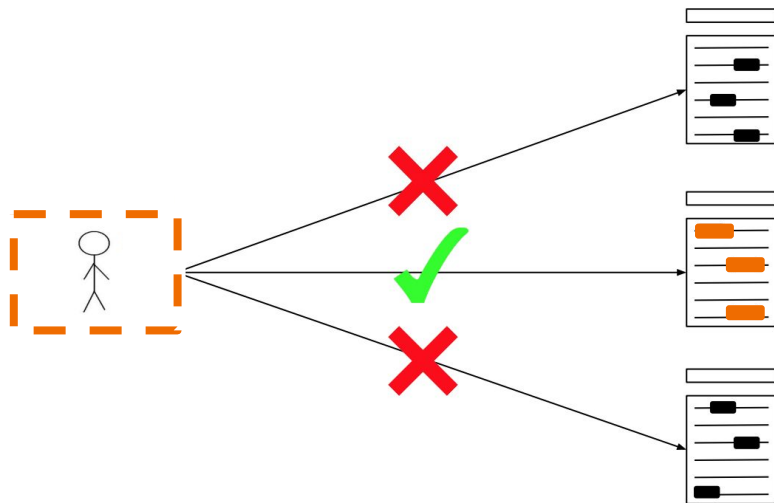
Early life and education

**2016 Elections**

US Senators

# Aspect Representations as Entities

I'm watching the  
debate between  
**Clinton** and Sanders



Early life and education

**2016 Elections**

US Senators

# Features

## Sparse vector representations

- **tf-idf (cs)**: vanilla VSM (TF-IDF, cosine)
- **BM25**: Okapi BM25 score

## Dense vector representations

- **w-emb (cs)**: cosine with pre-trained word embeddings (GloVe, 300d)
- **ent-emb (cs)**: cosine with pre-trained entity embeddings (RDF2Vec, 500d)



# Bringing it All Together...

<b>Representation</b>	<b>Features</b>
Header	tf-idf (cs), BM25, w-emb (cs)
Content	tf-idf (cs), BM25, w-emb (cs)
Entity	tf-idf (cs), BM25, ent-emb (cs)

- Similarities and relevance scores are input to a list-wise learning-to-rank (L2R) algorithm
- We optimize for P@1 using coordinate ascent with linear normalization

# Cool, But Are Entity-Aspects Useful?

# Entity Ranking for Queries

Given: **a query**

Return: **a ranking of relevant entities**

# Entity Ranking for Queries

Given: **a query**

Return: **a ranking of relevant entities**

“consequences brexit uk europe”

# Entity Ranking for Queries

Given: **a query**

Return: **a ranking of relevant entities**

“consequences brexit uk europe”



# Entity Ranking for Queries

---

<b>Model</b>	Robust04		ClueWeb12	
	<b>MAP</b>	<b>P@5</b>	<b>MAP</b>	<b>P@5</b>
REWQ [40]				
Fusion [14]				
<b>EAL</b>				

---

# Entity Ranking for Queries

<b>Model</b>	Robust04		ClueWeb12	
	<b>MAP</b>	<b>P@5</b>	<b>MAP</b>	<b>P@5</b>
REWQ [40]	-	0.79	-	<b>0.84</b>
Fusion [14]	0.57	0.85	0.39	0.82
<b>EAL</b>	<b>0.73</b>	<b>0.93</b>	<b>0.57</b>	0.74

# Entity Ranking for Named Events

Given: **a named event**

Return: **a ranking of relevant entities**



# Entity Ranking for Named Events

Given: **a named event**

Return: **a ranking of relevant entities**

*dbr:Brexit*

# Entity Ranking for Named Events

Given: a named event

Return: a ranking of relevant entities

*dbr:Brexit*



# Entity Ranking for Named Events

Model	MAP	P@5
RDF2Vec (cs)	0.65	0.80
<b>EAL (EvAsp)</b>		
<b>EAL (EntAsp)</b>		

**EvAsp:** rank on the basis of the similarity between the *entity name* and all sections on the *Wikipedia page of the event* (i.e., **event aspects**)

**EntAsp:** rank on the basis of the similarity between the *event name* and all sections on the Wikipedia page of the entity (i.e., **entity aspects**)

# Entity Ranking for Named Events

Model	MAP	P@5
RDF2Vec (cs)	0.65	0.80
<b>EAL (EvAsp)</b>	0.74	0.73
<b>EAL (EntAsp)</b>	<b>0.82</b>	<b>0.90</b>

**EvAsp:** rank on the basis of the similarity between the *entity name* and all sections on the *Wikipedia page of the event* (i.e., ***event aspects***)

**EntAsp:** rank on the basis of the similarity between the *event name* and all sections on the *Wikipedia page of the entity* (i.e., ***entity aspects***)

# Organizing Tweet Collections in Sub-Topics

# Organizing Tweet Collections in Sub-Topics



**Classify tweets** into any of the 8 aspects from the Wikipedia page [Issues in the United Kingdom European Union Membership Referendum, 2016](#)

# Organizing Tweet Collections in Sub-Topics

<b>Topic</b>	<b># Tweets</b>
Economy	
Immigration	
Sovereignty and influence	
Security, law enforcement and defense	
Risk to the Unity of the United Kingdom	
Transatlantic Trade and Investment Partnership	
Enlargement of the European Union	
Proposed consequences of a vote to leave	
<b>Total</b>	
<b>Excluded</b>	
General	
Out-of-topic	

# Organizing Tweet Collections in Sub-Topics

<b>Topic</b>	<b># Tweets</b>
Economy	
Immigration	
Sovereignty and influence	
Security, law enforcement and defense	
Risk to the Unity of the United Kingdom	
Transatlantic Trade and Investment Partnership	
Enlargement of the European Union	
Proposed consequences of a vote to leave	
<b>Total</b>	
<b>Excluded</b>	
General	270
Out-of-topic	108



# Organizing Tweet Collections in Sub-Topics

<b>Topic</b>	<b># Tweets</b>	
Economy		
Immigration		
Sovereignty and influence		
Security, law enforcement and defense		
Risk to the Unity of the United Kingdom		
Transatlantic Trade and Investment Partnership		
Enlargement of the European Union		
Proposed consequences of a vote to leave		
<b>Total</b>		
<b>Excluded</b>		
General	270	<i>The end is near #brexit</i>
Out-of-topic	108	<i>Congrats!!! ICELAND 2-1 <b>Brexit</b> in Euro 2016</i>

# Organizing Tweet Collections in Sub-Topics

<b>Topic</b>	<b># Tweets</b>
Economy	
Immigration	
Sovereignty and influence	
Security, law enforcement and defense	
Risk to the Unity of the United Kingdom	
Transatlantic Trade and Investment Partnership	
Enlargement of the European Union	
Proposed consequences of a vote to leave	
<b>Total</b>	<b>372</b>
<b>Excluded</b>	
General	270
Out-of-topic	108

# Organizing Tweet Collections in Sub-Topics

<b>Topic</b>	<b># Tweets</b>
Economy	155
Immigration	52
Sovereignty and influence	50
Security, law enforcement and defense	3
Risk to the Unity of the United Kingdom	30
Transatlantic Trade and Investment Partnership	5
Enlargement of the European Union	12
Proposed consequences of a vote to leave	65
<b>Total</b>	<b>372</b>
<b>Excluded</b>	
General	270
Out-of-topic	108

# Organizing Tweet Collections in Sub-Topics

<b>Model</b>	<b>P@1</b>
random baseline	1
<b>Ranking Approaches</b>	
Content - BM25	
Content - w-emb (cs)	
<b>EAL</b>	
<b>Classification Approaches</b>	
Naive Bayes (tf-idf)	
SVM (tf-idf)	
Naive Bayes (w-emb)	
SVM (w-emb)	

# Organizing Tweet Collections in Sub-Topics

<b>Model</b>	<b>P@1</b>
random baseline	0.12
<b>Ranking Approaches</b>	
Content - BM25	
Content - w-emb (cs)	
<b>EAL</b>	
<b>Classification Approaches</b>	
Naive Bayes (tf-idf)	0.27
SVM (tf-idf)	0.27
Naive Bayes (w-emb)	0.38
SVM (w-emb)	0.37

# Organizing Tweet Collections in Sub-Topics

<b>Model</b>	<b>P@1</b>
random baseline	0.12
<b>Ranking Approaches</b>	
Content - BM25	0.37
Content - w-emb (cs)	0.36
<b>EAL</b>	<b>0.42</b>
<b>Classification Approaches</b>	
Naive Bayes (tf-idf)	0.27
SVM (tf-idf)	0.27
Naive Bayes (w-emb)	0.38
SVM (w-emb)	0.37

# Entity-Aspect Linking

Federico Nanni, Simone Paolo Ponzetto, Laura Dietz:

**Entity-Aspect Linking: Providing Fine-Grained Semantics of Entities in Context.** *JCDL 2018*: 49-58

1. A **New perspective** on entity linking.
2. Aimed at providing fine-grained **information** on entities in **context**.
3. Useful for **downstream tasks**.

Data, resources, etc. @ <http://federiconanni.com/entity-aspect-linking/>

**That's it Folks!**



# That's it Folks!

- Entities are essential for a wide range of high-end tasks that involve end users

# That's it Folks!

- Entities are essential for a wide range of high-end tasks that involve end users
- “Think outside the box” and come up with way to exploit entities for real-world applications!

# That's it Folks!

- Entities are essential for a wide range of high-end tasks that involve end users
- “Think outside the box” and come up with way to exploit entities for real-world applications!
- As usual for great research topics, we are left with more questions than answers

# Future Work

- Combine symbolic, explicit semantic (e.g., ontological) information with statistical, distributional semantic representations of concepts, entities, and their relations (cf. our JOIN-T DFG project)

# Future Work

- Combine symbolic, explicit semantic (e.g., ontological) information with statistical, distributional semantic representations of concepts, entities, and their relations (cf. our JOIN-T DFG project)
- Leverage not-so-recent-anymore advances on dense semantic spaces (e.g., neural embeddings)

# Future Work

- Combine symbolic, explicit semantic (e.g., ontological) information with statistical, distributional semantic representations of concepts, entities, and their relations (cf. our JOIN-T DFG project)
- Leverage not-so-recent-anymore advances on dense semantic spaces (e.g., neural embeddings)
- Keep on working on downstream applications (semantic indexing, search, summarization, etc.)

# That's it Folks!

- Entities are essential for a wide range of high-end tasks that involve end users
- “Think outside the box” and come up with way to exploit entities for real-world applications!
- As usual for great research topics, we are left with more questions than answers
- Data, resources, etc. @

[dws.informatik.uni-mannheim.de](http://dws.informatik.uni-mannheim.de)

