



max planck institut
informatik

Language-Model based Ranking in Entity-Relationship Graphs

Shady Elbassuoni, Maya Ramanath, Gerhard Weikum

Max-Planck Institute for Informatics

Saarbrücken, Germany

Motivation

- Politicians who are also scientists ?
- Academy Awarded movie directed by an Australian director?
- Fiction books written by a Nobel prize winner?
- Author of two papers in WWW 2008?



politician scientist

Search

[Advanced Search](#)
[Preferences](#)

Web

Results 1 - 10 of about 1,150,000 for [politician scientist](#). (0.05 seconds)

[Charles, 3rd earl Stanhope, England, radical **politician/scientist** ...](#)

August 1, 1753 in History. Born: Charles, 3rd earl Stanhope, England, radical politician/scientist, Related Topics: Charles · earl · England · politician ...
www.brainyhistory.com/events/1753/august_1_1753_41696.html - 7k -

[Cached](#) - [Similar pages](#)

[John Lubbock, 1st Baron Avebury \(1834-1913\), Banker, **politician** ...](#)

John Lubbock, 1st Baron Avebury (1834-1913), Banker, politician, scientist and writer Sitter in 7 portraits Page 1 of 1 ...

www.npg.org.uk/live/search/person.asp?LinkID=mp00183 - 13k -

[Cached](#) - [Similar pages](#)

[JSTOR: Charles Tait, Planter, **Politician**, and **Scientist** of the Old ...](#)

CHARLES TAIT, PLANTER, POLITICIAN, SCIENTIST 207 his father's tobacco farm in Virginia, but when decreasing yields and rising costs of production made ...

[links.jstor.org/sici?sici=0022-4642\(194805\)14%3A2%3C206%3ACTPPAS%3E2.0.CO%3B2-F](http://links.jstor.org/sici?sici=0022-4642(194805)14%3A2%3C206%3ACTPPAS%3E2.0.CO%3B2-F) -

[Similar pages](#)

[Center for Health and the Global Environment](#)

Letters to politicians, scientists, community members, the media, etc can be a great way to apply material learned in the video presentations and articles. ...

chge.med.harvard.edu/education/secondary/hhgec/secondhhfw.html - 27k -

[Cached](#) - [Similar pages](#)

[Amazon.com: Science Under Siege: The **Politicians'** War on Nature ...](#)

Amazon.com: Science Under Siege: The Politicians' War on Nature and Truth: Books: Todd Wilkinson by Todd Wilkinson.

www.amazon.com/Science-Under-Siege-Politicians-Nature/dp/1555662110 - 126k -

[Cached](#) - [Similar pages](#)

[NPR : Remembering Ben Franklin on his 300th Birthday](#)

All this year the politician, scie

www.npr.org/ter

[More **politicians** write blogs to bypass mainstream media ...](#)

Yet politicians are beginning to see blogs are more than forums for snoops. ... Michael Cornfield, a political scientist at George Washington University. ...

www.csmonitor.com/2005/0324/p02s01-uspo.html - 59k - 29 May 2007 -

[Cached](#) - [Similar pages](#)

What is lacking ?

- Data is not knowledge
 - Extraction and Organization
- Keywords cannot express advanced user intentions
 - Expression of concepts, relationships, etc.

Represent, Search and Rank Knowledge on the Web

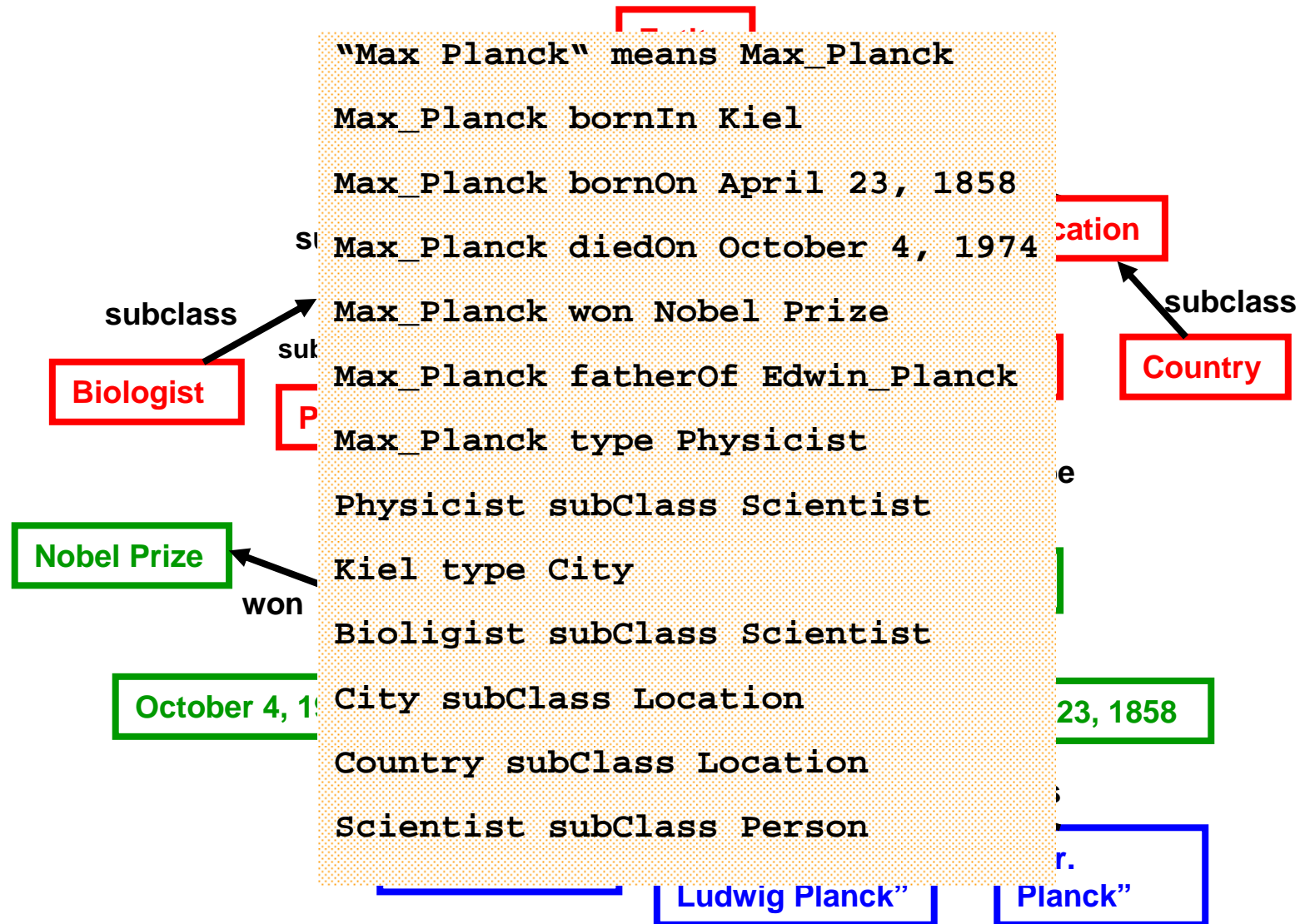
Knowledge Representation, Searching and Ranking

- Knowledge Representation
 - Entity-Relationship (ER) Graphs
 - Graphs of **RDF-triples** derived from web-sources
 - YAGO, FreeBase, DBPedia, DBLife, Libra, etc...
- Knowledge Searching
 - Expressive yet **simple** means of querying
 - Structured queries (as close to keywords as possible)
- Knowledge Ranking
 - IR-style ranking for structured queries over ER graphs

Outline

- Motivation
- Knowledge Representation
 - Entity-Relationship Graphs
- Knowledge Searching
 - Query Language
 - Result Matching
- Knowledge Ranking
 - Language-Model based ranking for structured queries
- Preliminary Evaluation

Entity-Relationship Graphs

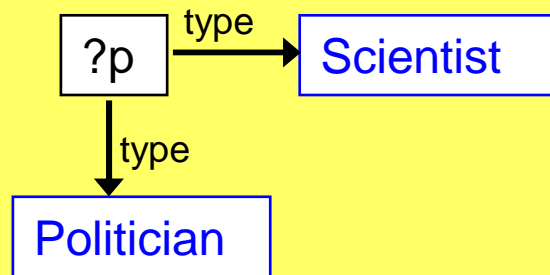


Outline

- Motivation
- Knowledge Representation
 - Entity-Relationship Graphs
- **Knowledge Searching**
 - Query Language
 - Result Matching
- Knowledge Ranking
 - Language-Model based ranking for structured queries
- Preliminary Evaluation

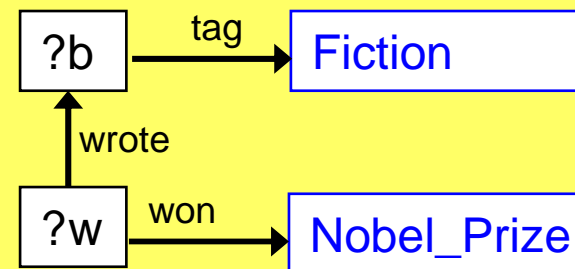
Query Language

- Politicians who are also scientists?



?p type Scientist; ?p type Politician

- Fiction books written by a Nobel prize winner?



?w wrote ?b; ?w won Nobel_Prize; ?b tag Fiction

Result Matching

- Based on graph matching
 - Results are **tuples** that match the query **triple-templates**

```
?p type Scientist; ?p type Politician
```

```
Benjamin_Franklin type Scientist; Benjamin_Franklin type Politician
```

```
Alan_Greenspan type Scientist; Alan_Greenspan type Politician
```

```
Jan_Smuts type Scientist; Jan_Smuts type Politician
```

```
Paul_Wolfowitz type Scientist; Paul_Wolfowitz type Politician
```

Ranking is required

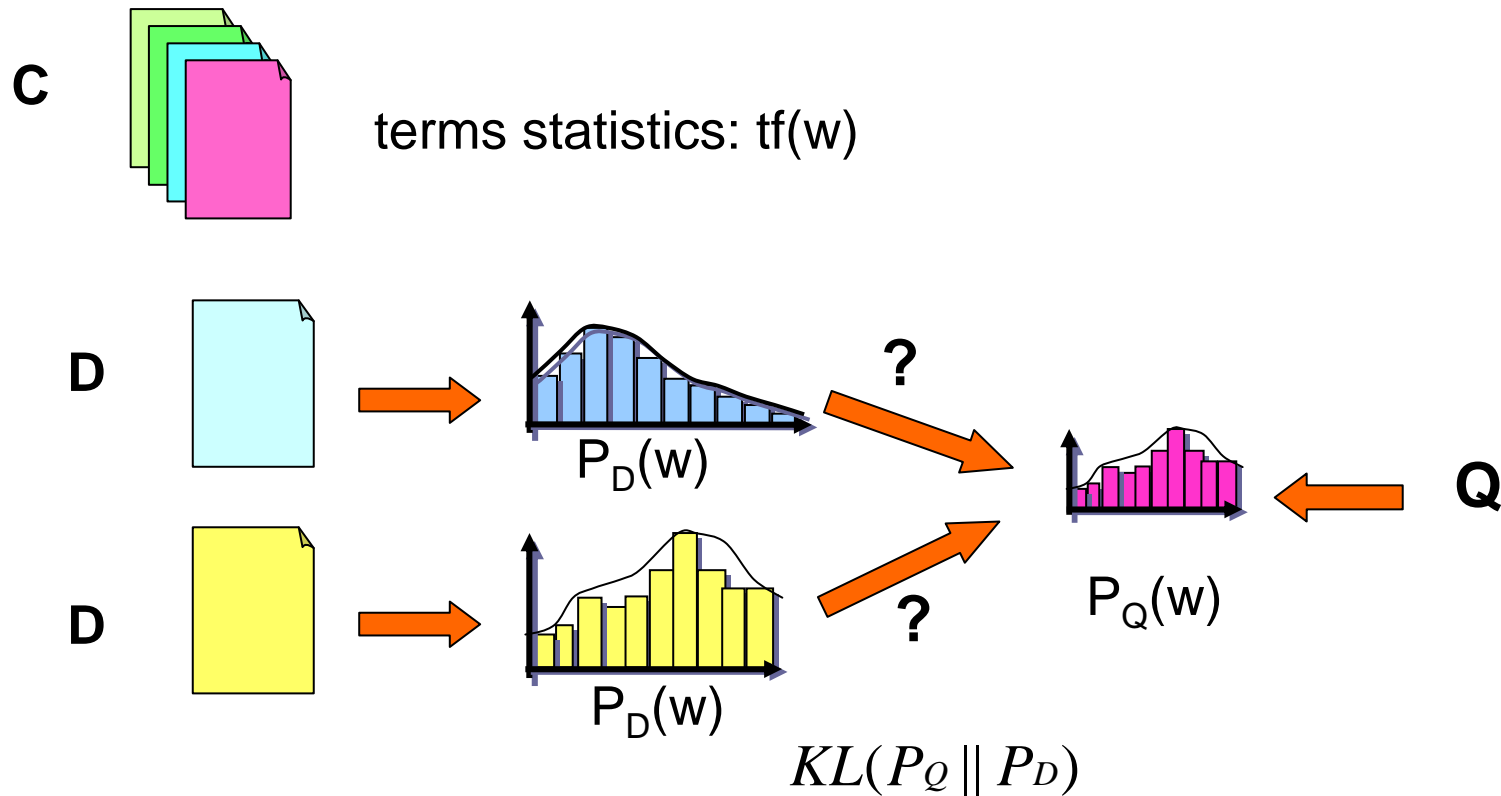
Outline

- Motivation
- Knowledge Representation
 - Entity-Relationship Graphs
- Knowledge Searching
 - Query Language
 - Result Matching
- **Knowledge Ranking**
 - Language-Model based ranking for structured queries
- Preliminary Evaluation

Language Model Ranking

- Results, Query: probability distributions over some **common** vocabulary
 - Construct Query Language Model P_Q
 - Construct Result Language Model P_D
- Measure how close the distributions are
 - “Closeness” gives measure of “relevance”
 - Rank results based on **KL-Divergence** between P_Q and P_D

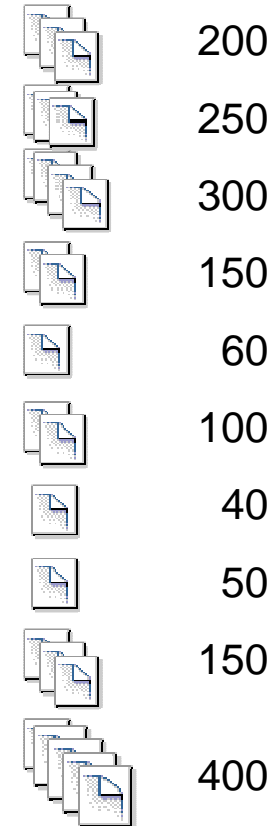
Language Models in Textual Retrieval



Language Models in ER Graphs

```

t1: Benjamin_Franklin type Scientist
t2: Benjamin_Franklin type Politician
t3: Paul_Wolfowitz type Scientist
t4: Paul_Wolfowitz type Politician
t5: Alan_Greenspan type Scientist
t6: Alan_Greenspan type Politician
t7: Jan_Smuts type Scientist
t8: Jan_Smuts type Politician
t9: Alan_Greenspan citizen American
t10: Benjamin_Franklin citizen American
  
```



```

T1: Benjamin_Franklin type Scientist;
    Benjamin_Franklin type Politician
  
```

```

T2: Alan_Greenspan type Scientist;
    Alan_Greenspan type Politician
  
```

Q

```

?p type Scientist;
?p type Politician
  
```

Terms in ER Graphs

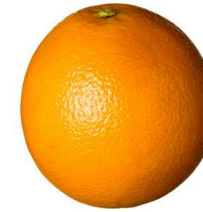


Triples



```
T1: Benjamin_Franklin type Scientist;  
Benjamin_Franklin type Politician
```

```
T2: Alan_Greenspan_type Scientist;  
Alan_Greenspan type Politician
```



Triple templates



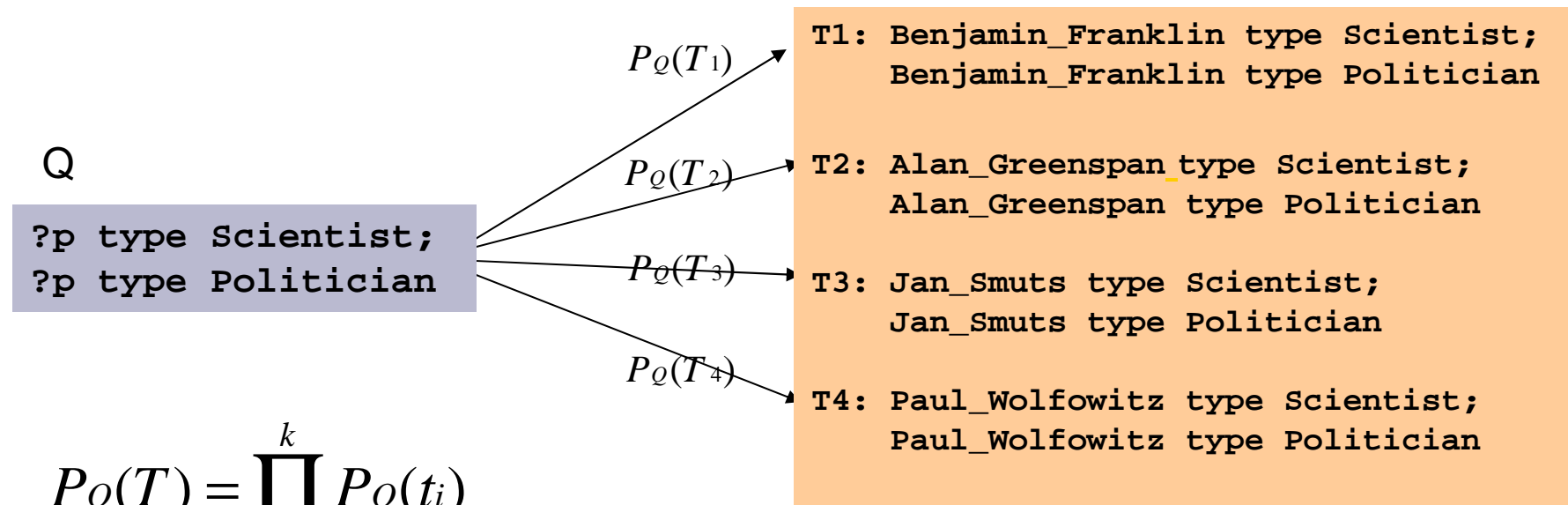
Q

```
?p type Scientist;  
?p type Politician
```



Query Language Model

Tuples of triples



$$P_Q(T) = \prod_{i=1}^k P_Q(t_i)$$

$P_Q(\text{Benjamin_Franklin type Scientist; Benjamin_Franklin type Politician})$

=

$P_Q(\text{Benjamin_Franklin type Scientist}) \cdot P_Q(\text{Benjamin_Franklin type Politician})$

Estimating Query LM

Q

?p type Scientist;
?p type Politician

$$P_Q(t_i) = \frac{c(t_i)}{\sum_j c(t_j)}$$

t_i	$C(t_i)$	$P_Q(t_i)$
Benjamin_Franklin type Scientist	200	200/600
Alan_Greenspan type Scientist	60	60/600
Jan_Smuts type Scientist	40	40/600
Paul_Wolfowitz type Scientist	300	300/600
	600	

Result Language Model

Results (Tuples of triples)

T1: Benjamin_Franklin type Scientist;
Benjamin_Franklin type Politician

T2: Alan_Greenspan type Scientist;
Alan_Greenspan type Politician

$$P_D(T) = \alpha P(T | D) + (1 - \alpha) P(T | C)$$

$$P(T | C) = \prod_{i=1}^k P(t_i | C)$$

Estimating Result LM

$$P(t_i | C) = \frac{c(t_i)}{\sum_j c(t_j | C)}$$

	C(t_i)	P (t_i C)
t1: Benjamin_Franklin type Scientist	200	200/1700
t2: Benjamin_Franklin type Politician	250	250/1700
t3: Paul_Wolfowitz type Scientist	300	300/1700
t4: Paul_Wolfowitz type Politician	150	150/1700
t5: Alan_Greenspan type Scientist	60	60/1700
t6: Alan_Greenspan type Politician	100	100/1700
t7: Jan_Smuts type Scientist	40	40/1700
t8: Jan_Smuts type Politician	50	50/1700
t9: Alan_Greenspan citizen American	150	150/1700
t10: Benjamin_Franklin citizen American	400	400/1700
	1700	

Example Results

IMDB Queries	Q	Top result
<i>Movies with genre comedy and two actors who acted in it.</i>	3	Sideways, Paul Giamatti, Sandra Oh
<i>Actor who acted in a family movie in 1995 and who also acts in comedies.</i>	5	Tom Hanks, Toy Story, You've Got Mail

LibraryThing Queries	Q	Top result
<i>A fiction book with tag Magic and its author</i>	3	Stephanie Meyer, New Moon
<i>Two books written by the same novelist</i>	3	Cormac McCarthy, The Road, No Country for Old Men

Conclusions

- Knowledge Representation, Searching and Ranking
- Language Model based Ranking in ER graphs
 - Estimate Query and Result LMs
 - Relevance determined by KL-divergence between the two estimated models
- Evaluation
 - Preliminary, but promising



max planck institut
informatik



THINK IBIZA, THINK AGAIN.

THINK KEYWORDS, THINK AGAIN.