



# Entity Linking

ENITVΛ ΓIPKIND

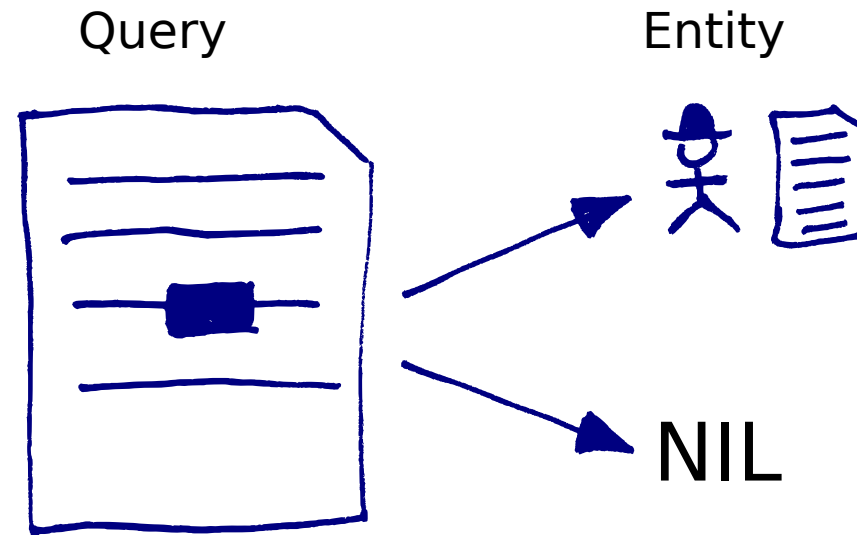


**Laura Dietz** [dietz@cs.umass.edu](mailto:dietz@cs.umass.edu)

University of Massachusetts

Use cursor keys to flip through slides.

# Problem: Entity Linking



Given query mention in a source document, identify which Wikipedia entity it represents

# Problem: Example

## Example Query:

Northern Ireland has a population of about one and a half million people. At the time of partition in 1921 Protestants / unionists had a two-thirds majority in the region. The first Prime Minister of **Northern Ireland**, Sir James Craig, described the state as having 'a Protestant Parliament for a Protestant people.' The state effectively discriminated against Catholics in housing, jobs, and political representation.

<http://cain.ulst.ac.uk/othelem/incorepaper09.htm>

## Search for:

**Northern Ireland**



**WIKIPEDIA**  
The Free Encyclopedia

[Main page](#)  
[Contents](#)  
[Featured content](#)  
[Current events](#)  
[Random article](#)  
[Donate to Wikipedia](#)  
[Wikimedia Shop](#)

Interaction

- [Help](#)
- [About Wikipedia](#)
- [Community portal](#)
- [Recent changes](#)
- [Contact Wikipedia](#)

[Toolbox](#)

[Print/export](#)

Languages

- [Afrikaans](#)
- [Ænglisc](#)
- [العربية](#)
- [Aragonés](#)
- [Asturianu](#)
- [Azərbaycanca](#)
- [বাংলা](#)
- [Bân-lâm-gú](#)
- [Беларуская](#)
- [Беларуская \(тарашкевіца\)](#)
- [Български](#)
- [Boarisch](#)

Article [Talk](#)

[Read](#) [View source](#) [View history](#)

# Northern Ireland

From Wikipedia, the free encyclopedia

*For the European Parliament constituency, see [Northern Ireland \(European Parliament constituency\)](#).*

**Northern Ireland** (Irish: *Tuaisceart Éireann*

pronounced [ˈt̪ˠuə̃ːfˠc̪ˠəɾˠ t̪ˠˠ ˈeːr̪ˠəl̪ˠn̪ˠ] ( listen), Ulster Scots: *Norlin Airlann* or *Norlin Airlan*) is a part of the [United Kingdom](#) in the north-east of the island of Ireland. It is variously described as a [country](#), [province](#) or [region](#) of the UK, amongst other terms.<sup>[3][4][5]</sup> Northern Ireland shares a border with the [Republic of Ireland](#) to the south and west. As of 2011, its population was 1,810,863,<sup>[2]</sup> constituting about 30% of the island's total population and about 3% of the population of the [United Kingdom](#). Since the signing of the [Good Friday Agreement](#) in 1998, Northern Ireland is largely self-governing. According to the agreement, Northern Ireland co-operates with the rest of Ireland – from which it was [partitioned in 1921](#) – on some policy areas, while other areas are reserved for the [Government of the United Kingdom](#), though the Republic of Ireland "may put forward views and proposals".<sup>[6]</sup>

Northern Ireland was for many years the site of a violent and bitter inter-communal conflict – [the Troubles](#) – which was caused by divisions between [nationalists](#), who see themselves as Irish and are predominantly Roman Catholic, and [unionists](#), who see themselves as British and are predominantly Protestant. (Additionally, people from both sides of the community may describe themselves as Northern Irish.)<sup>[7]</sup> Unionists want Northern Ireland to remain as a part of the [United Kingdom](#),<sup>[8]</sup> while nationalists want [reunification](#) with the rest of Ireland, independent of British rule.<sup>[9][10][11][12]</sup> Since 1998, most of the paramilitary groups involved in the Troubles have ceased their armed campaigns.

[Northern Ireland](#) has traditionally been the most industrialised region of the island. After declining as a result of [political and](#)

## Northern Ireland

*Tuaisceart Éireann*  
*Norlin Airlann*



Location of **Northern Ireland** (dark green)  
 – in the European continent (light green & dark grey)  
 – in the [United Kingdom](#) (light green)

<b>Capital and largest city</b>	<b>Belfast</b> <span><span><span><span><span>54°35.456′N</span> <span>5°50.000′W</span></span></span><span><span>﻿</span> / <span>﻿</span></span><span><span>54.591°N 5.833°W</span><span><span>﻿</span> / <span>54.591; -5.833</span></span></span></span></span>
<b>Official languages</b>	English Irish Ulster Scots <sup>1</sup>
<b>Ethnic groups</b>	99.15% <a href="#">White</a> (91% Northern Ireland born) 8.15% other white 0.41% <a href="#">Asian</a> 0.10% <a href="#">Irish Traveller</a> 0.34% others. <sup>[1]</sup>
<b>Demonym</b>	Northern Irish
<b>Government</b>	Consociational de facto government with constitutional monarchy
<span> </span> - <a href="#">Monarch</a>	Elizabeth II
<span> </span> - <a href="#">First Minister</a>	Peter Robinson, M.P.

# Problem: Example

## Example Query:

Northern Ireland has a population of about one and a half million people. At the time of partition in 1921 Protestants / unionists had a two-thirds majority in the region. The first Prime Minister of Northern Ireland, Sir **James Craig**, described the state as having 'a Protestant Parliament for a Protestant people.' The state effectively discriminated against Catholics in housing, jobs, and political representation.

<http://cain.ulst.ac.uk/othelem/incorepaper09.htm>

## Search for:

**James Craig**

About 59,200 results (0.49 seconds)

[James Craig \(actor\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/James\\_Craig\\_\(actor\)](https://en.wikipedia.org/wiki/James_Craig_(actor)) ▾

**James Craig** (February 4, 1912 – June 28, 1985) was an American actor. After graduating from the Rice Institute, Craig began appearing in films in 1937, most ...

[James Craig - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/James\\_Craig](https://en.wikipedia.org/wiki/James_Craig) ▾

**James Craig** or **Jim Craig** may refer to: Contents. 1 Public officials; 2 Sports personalities; 3 Actors; 4 Other people; 5 Characters; 6 Other. Public officials[edit].

[James Craig, 1st Viscount Craigavon - Wikipedia, the free ...](#)

[en.wikipedia.org/wiki/James\\_Craig,\\_1st\\_Viscount\\_Craigavon](https://en.wikipedia.org/wiki/James_Craig,_1st_Viscount_Craigavon) ▾

**James Craig**, 1st Viscount Craigavon, PC, PC (NI) (8 January 1871 – 24 November 1940), was a prominent Irish unionist politician, leader of the Ulster Unionist ...

[James Craig \(Missouri\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/James\\_Craig\\_\(Missouri\)](https://en.wikipedia.org/wiki/James_Craig_(Missouri)) ▾

**James Craig** (February 28, 1818 – October 22, 1888) was an American lawyer and politician from Saint Joseph, Missouri. He represented Missouri in the U.S. ...

[James Craig \(architect\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/James\\_Craig\\_\(architect\)](https://en.wikipedia.org/wiki/James_Craig_(architect)) ▾

**James Craig** (31 October 1739 – 23 June 1795) was a Scottish architect. His brief career was concentrated almost entirely in Edinburgh, and he is remembered ...

[Jim Craig \(ice hockey\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Jim\\_Craig\\_\(ice\\_hockey\)](https://en.wikipedia.org/wiki/Jim_Craig_(ice_hockey)) ▾

**James Downey Craig** (born May 31, 1957) is a former American ice hockey goaltender who is most notable for being the goaltender for the 1980 U.S. Olympic ...

# James Craig



Actor

James Craig was an American actor. After graduating from the Rice Institute, Craig began appearing in films in 1937, most often in B-movies and serials. In 1939, he appeared in the Three Stooges film *Oily to Bed, Oily to Rise*.

Wikipedia

**Born:** February 4, 1912, [Nashville, TN](#)

**Died:** June 28, 1985, [Santa Ana, CA](#)

**Education:** [Rice University](#)

**Spouse:** [Sumie Craig](#) (m. 1969–1980), [Jil Jarmyn](#) (m. 1959–1962)

## Movies



The Devil



Kitty Foyle:



Drums in

James Craig site:en.wikipedia.org



Sign in

Web Images Maps Shopping More Search tools



About 51,000 results (0.30 seconds)

near miss! :(

[James Craig - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/James\\_Craig](http://en.wikipedia.org/wiki/James_Craig)

**James Craig** or **Jim Craig** may refer to: Contents. 1 Public officials; 2 Sports personalities; 3 Actors; 4 Other people; 5 Characters; 6 Other. [edit] Public officials ...  
Public officials - Sports personalities - Actors - Other people

[James Craig \(actor\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/James\\_Craig\\_\(actor\)](http://en.wikipedia.org/wiki/James_Craig_(actor))

**James Craig** (February 4, 1912 – June 28, 1985) was an American actor. After graduating from the Rice Institute, Craig began appearing in films in 1937, most ...

[James Craig, 1st Viscount Craigavon - Wikipedia, the free ...](#)

[en.wikipedia.org/wiki/James\\_Craig,\\_1st\\_Viscount\\_Craigavon](http://en.wikipedia.org/wiki/James_Craig,_1st_Viscount_Craigavon)

**James Craig**, 1st Viscount Craigavon, PC, PC (NI) (8 January 1871 – 24 November 1940), was a prominent Irish unionist politician, leader of the Ulster Unionist ...

[Jim Craig \(ice hockey\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Jim\\_Craig\\_\(ice\\_hockey\)](http://en.wikipedia.org/wiki/Jim_Craig_(ice_hockey))

**James Downey Craig** (born May 31, 1957) is a former American ice hockey goaltender who is most notable for being the goaltender for the 1980 U.S. Olympic ...

[James Craig \(barque\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/James\\_Craig\\_\(barque\)](http://en.wikipedia.org/wiki/James_Craig_(barque))

Capacity: 1,100 tons. Complement: 16. The **James Craig** is a three-masted, iron-hulled barque restored and sailed by the Sydney Maritime Museum. Contents ...

[James Craig \(architect\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/James\\_Craig\\_\(architect\)](http://en.wikipedia.org/wiki/James_Craig_(architect))

**James Craig** (31 October 1739 – 23 June 1795) was a Scottish architect. His brief career was concentrated almost entirely in Edinburgh, and he is remembered ...

[James Craig \(Missouri\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/James\\_Craig\\_\(Missouri\)](http://en.wikipedia.org/wiki/James_Craig_(Missouri))

See results about



**James Craig**  
Actor  
James Craig was an American actor. After ...



**James Craig, 1st Viscount Craigavon**  
James Craig, 1st Viscount Craigavon, PC, PC, was a ...



**James Craig**  
Barque  
The James Craig is a three-masted, ...



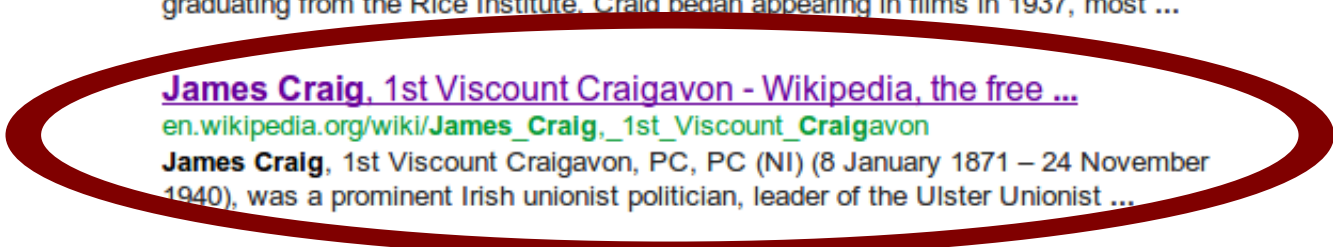
**James Craig**  
Architect  
James Craig was a Scottish architect. His ...



**James Henry Craig**  
General Sir James Henry Craig KB was a British military officer ...



**James Craig**  
Loyalist  
James Pratt "Jim" Craig was a Northern Irish ...



# Overview

M1: Popularity Method

M2: Machine Learned Similarity

M3: Context with IR

M4: Joint Assignment Model

M5: Joint Retrieval Model


Experimental Results

Online Demos



# Challenges


## am·bi·gu·i·ty

/,ambi'gyoo-itē/ 

*noun*

1. uncertainty or inexactness of meaning in language.  
"we can detect no ambiguity in this section of the Act"  
*synonyms:* vagueness, **obscurity**, abstruseness, doubtfulness, **uncertainty**; [More](#)

## syn·o·nym

/'sinə,nim/ 

*noun*

1. a word or phrase that means exactly or nearly the same as another word or phrase in the same language, for example *shut* is a synonym of *close*.  
*synonyms:* **alternate**, **substitute**, **alternative**, **equivalent**, **euphemism**; [More](#)

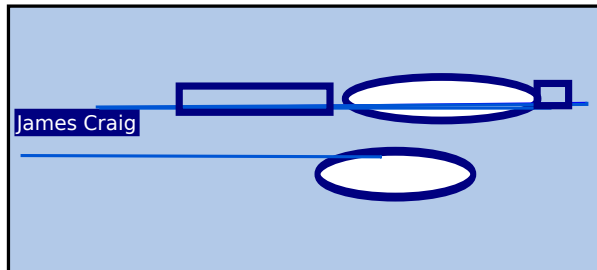
# Problem: Example

## Example Query:

Northern Ireland has a population of about one and a half million people. At the time of partition in 1921 Protestants / unionists had a two-thirds majority in the region. The first Prime Minister of Northern Ireland, Sir **James Craig**, described the state as having 'a Protestant Parliament for a Protestant people.' The state effectively discriminated against Catholics in housing, jobs, and political representation.

<http://cain.ulst.ac.uk/othelem/incorepaper09.htm>

# Document Analysis



Name Variants:

Within-doc Coreference

Neighbor Mentions:

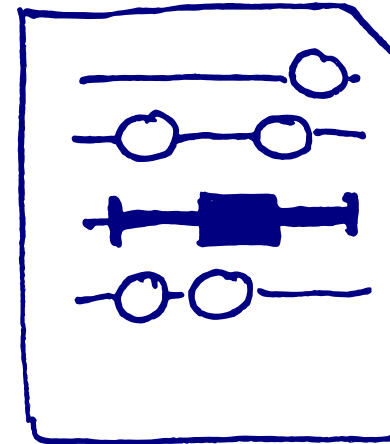
NER Tagger

(Alternative Mention Detection)

Sentence:

Term models

Symbol Notation:



- Q: Query String
- V: Name Variants
- M: Neighbor Mentions
- ⊢ S: Sentence

# Method 1: Popularity of Links

Step 1:

Build a dictionary of names for each entity.

Step 2:

Inspect all KB entities that have the query mention as a name variant.

Step 3:

Choose the entity with the most inlinks through this name.

# Names and Links on Wikipedia

## Northern Ireland

**Northern Ireland** (Irish: *Tuaisceart Éireann*<sup>[4]</sup> [t̪ˠuəʃc̪əɾˠt̪ˠ ˈeːɾʲəɲˠ] ( listen); Ulster Scots: *Norlin Airlann* or *Norlin Airlan*) is a part of the United Kingdom in the north-east of the island of Ireland. It is variously described as a country, province or region of the UK, amongst other terms.<sup>[5][6][7]</sup> Northern Ireland shares a border to the south and west with the Republic of Ireland. As of 2011, its

## James Craig, 1st Viscount Craigavon

From Wikipedia, the free encyclopedia

*For other uses, see [James Craig \(disambiguation\)](#).*

**James Craig, 1st Viscount Craigavon, PC, PC (NI)** (8 January 1871 – 24 November 1940), was a prominent **Irish unionist** politician, leader of the Ulster Unionist Party and the first **Prime Minister of Northern Ireland**. He was created a **baronet** in 1918 and raised to the Peerage in 1927.

**The Right Honourable**

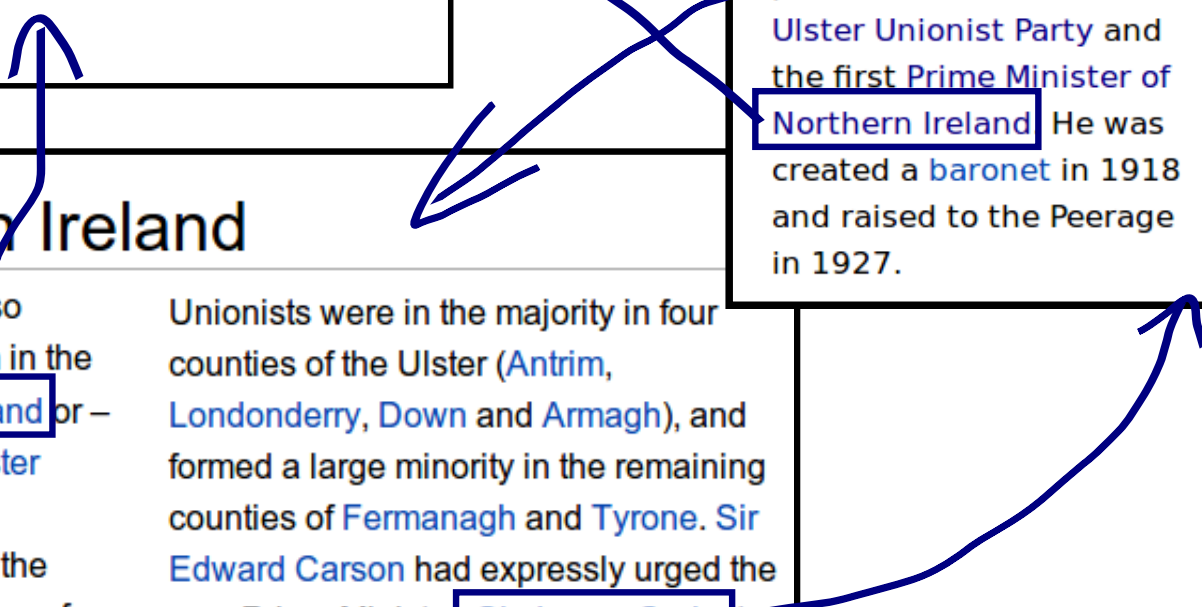
**The Viscount Craigavon**  
**PC**



## Unionism in Ireland

In Ireland, unionism, also called **Ulster unionism** in the context of **Northern Ireland** or – chiefly historically – **Ulster** specifically, refers to an ideology which favours the continuation of some form of **political union** between the islands of Ireland and Great

Unionists were in the majority in four counties of the Ulster (**Antrim**, **Londonderry**, **Down** and **Armagh**), and formed a large minority in the remaining counties of **Fermanagh** and **Tyrone**. **Sir Edward Carson** had expressly urged the new Prime Minister, **Sir James Craig**, to ensure absolute equality in the treatment of Catholics, so to guarantee the stability



# Mining Name Variants and Neighbors

## Northern Ireland



Northern  
Ireland

## James Craig, 1st Viscount Craigavon



Sir James Craig

1st Viscount  
Craigavon

Ulster Unionists

Northern Ireland

Prime Minister of  
Northern Ireland

## Unionism in Ireland



Irish Unionist

Unionism in  
Ireland

Ulster

Northern Ireland

James Craig, 1st  
Viscount Craigavon

## Pros & Cons: Popularity of Links

Works for very popular entities  
such as "Northern Ireland"

Fails for entities with confusable names  
"James Craig", "Springfield", "Jaguar"

# Method 1: Popularity of Links

Step 1:

Build a dictionary of names for each entity.

Step 2:

Inspect all KB entities that have the query mention as a name variant.

Step 3:

Choose the entity with the most inlinks through this name.



## Method 2: Machine Learn Similarity

Step 1:

Collect different **similarity features** of query mention and entities

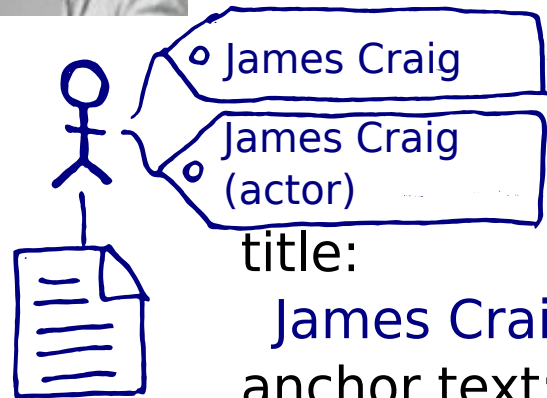
Step 2:

**Machine learn** the feature weights on training data (e.g. learning to rank)

Step 3:

Apply similarity to query and each entity, select the most similar entity.

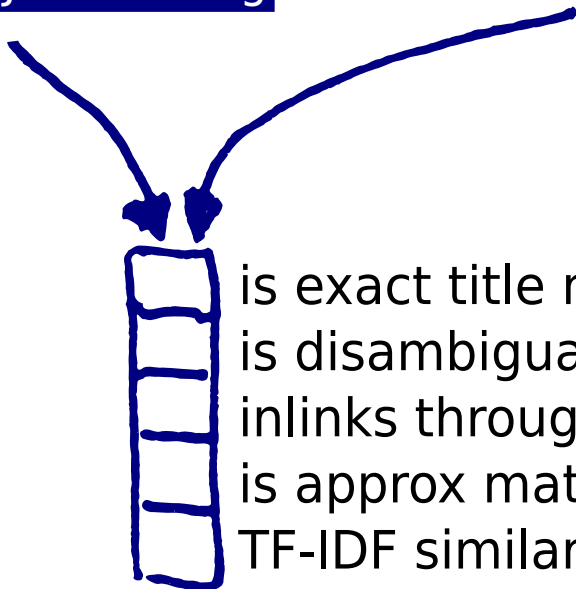
# Method 2: Similarity Features



title:  
 James Craig (actor)  
 anchor text:  
 James Craig  
 James Craig in  
 disambiguation:  
 James Craig  
 freebase name:  
 James Craig (actor)

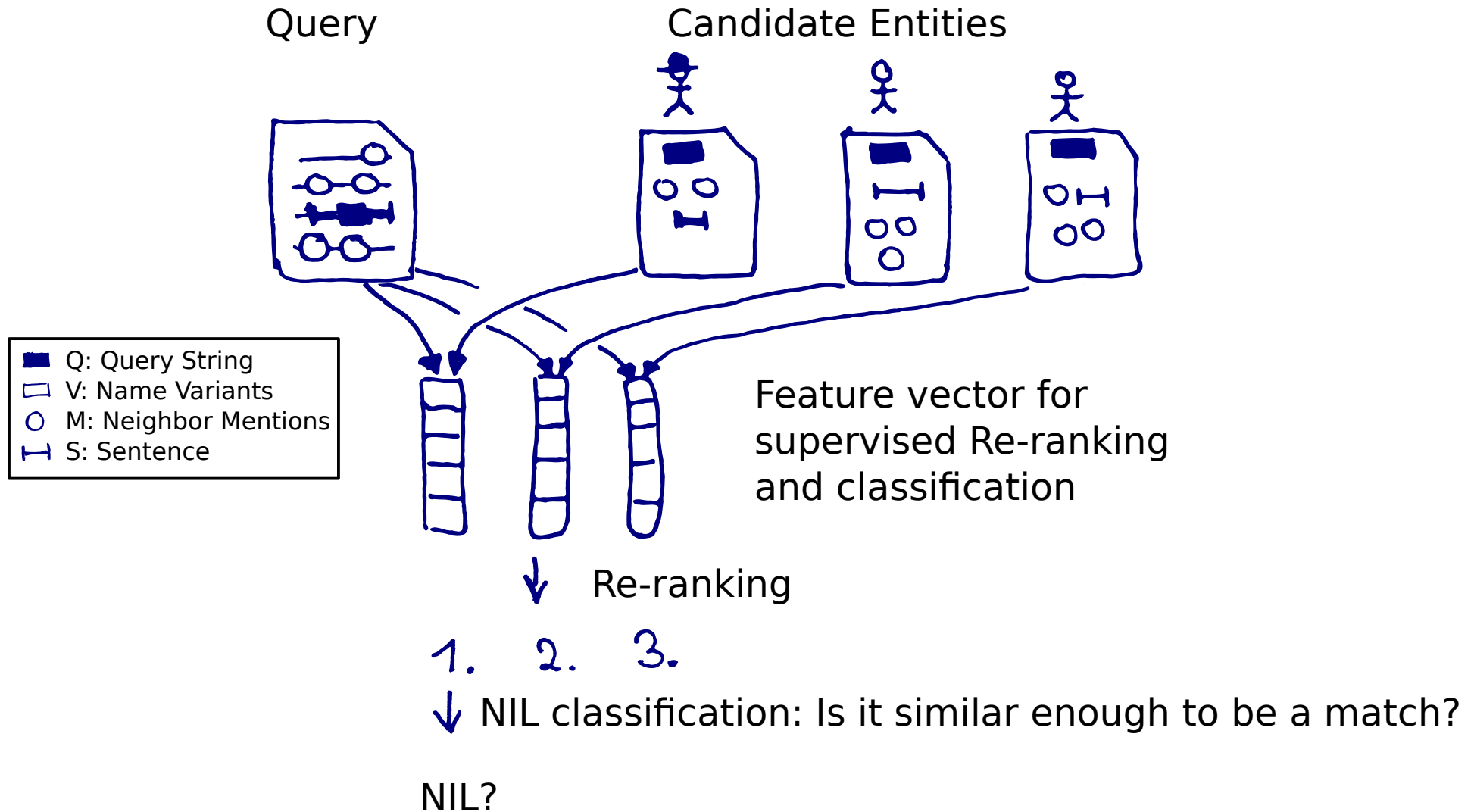
title:  
 James Craig, 1st  
 Viscount Craigavon  
 anchor text:  
 Sir James Craig's  
 Craig Administration  
 disambiguation:  
 James Craig  
 freebase name:  
 Lord Craigavon

James Craig



- is exact title match?
- is disambiguation match?
- inlinks through this name
- is approx match?
- TF-IDF similarity score

# Learn Similarity and NIL



Features: Name variants, Document Terms, Links, Popularity ...

## Pros & Cons: Machine Learn Similarity

Pro: Combination of different indicators of similarity; option to predict "NILs".

Pro: Can incorporate name variants found in the text (coreference tools)

Con: Requires selection of a pool of candidate entities, which can be large ("John Smith").

Will still fail on "James Craig", because the wrong James has more anchor text matches.

## Method 3: Context Disambiguation

Step 1:

Identify surrounding text, entities, etc.

Step 2:

Issue search query containing all of it.

# Different Kinds of Context

Example Query:

Northern Ireland has a population of about one and a half million people. At the time of partition in 1921 Protestants / unionists had a two-thirds majority in the region. The first Prime Minister of Northern Ireland, Sir James Craig, described the state as having 'a Protestant Parliament for a Protestant people.' The state effectively discriminated against Catholics in housing, jobs, and political representation.

<http://cain.ulst.ac.uk/othelem/incorepaper09.htm>

Search for:

**James Craig** + Name Variants + Neighbors + Sentence



James Craig Northern Ireland Catholics site:en.wikipedia.org



Sign in

Web Images Maps Shopping More Search tools



About 30,000 results (0.32 seconds)

[James Craig, 1st Viscount Craigavon - Wikipedia, the free ...](#)

[en.wikipedia.org/wiki/James\\_Craig,\\_1st\\_Viscount\\_Craigavon](#)

Member must remember that in the South they boasted of a **Catholic** State. ... Sir **James Craig**, Unionist Party, then Prime Minister of **Northern Ireland**, 24 April ...



[Partition of Ireland - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Partition\\_of\\_Ireland](#)

The Prime Minister of **Northern Ireland**, Sir **James Craig**, speaking in the ... persecution of Protestants by a Roman **Catholic**-dominated Irish government.

[James Craig - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/James\\_Craig](#)

James Henry Craig (1748–1812), British soldier and colonial administrator; James ... **James Craig** (County Antrim, 20th century) (1931–1974), **Northern Ireland** ...

[File:James Craig Viscount Craigavon.jpg - Wikipedia, the free ...](#)

[en.wikipedia.org/wiki/File:James\\_Craig\\_Viscount\\_Craigavon.jpg](#)

Description, Sir **James Craig**, 1st Viscount of Craigavon, first Prime Minister of **Northern Ireland**. Date. Source. This file is lacking source information. Please edit ...

[James Craig \(loyalist\) - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/James\\_Craig\\_\(loyalist\)](#)

James Pratt "**Jim**" **Craig** (c.1941 – 15 October 1988) was a **Northern Irish** .... Tyrrie suggested that Craig was a suspect on account that his wife was **Catholic**.

[Orange Order - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/Orange\\_Order](#)

**Catholics**, and those whose close relatives are **Catholic**, are banned from ..... **James Craig**, the first Prime Minister of **Northern Ireland**, maintained always that ...

[The Troubles - Wikipedia, the free encyclopedia](#)

[en.wikipedia.org/wiki/The\\_Troubles](#)

The conflict was the result of discrimination against the Nationalist/**Catholic** .... Sir

## Method 3: Pros and Cons

Works for "James Craig"!

Problematic when neighbors are ambiguous:

"Lisa witnessed a shooting at Springfield high school".

(Unclear which "Lisa" and which "Springfield")



## Method 3: Pros and Cons

Also problematic when neighbors don't provide enough disambiguation power

Example, all other James Craigs of Ireland which are less popular.

# James Craig

From Wikipedia, the free encyclopedia

**James Craig** or **Jim Craig** may refer to:

## Public officials [edit]

- [James Henry Craig](#) (1748–1812), British soldier
- [James Craig \(County Antrim, 19th century\)](#) (1759–1833), British politician
- [James Craig \(Ontario\)](#) (1823–1874), Canadian politician
- [James Craig \(Missouri\)](#) (1818–1888), American politician and soldier
- [James Craig \(Newcastle\)](#) (1834–1902), British politician
- [James Craig \(University of Dublin\)](#) (died 1933), British and Irish politician, physician, and educator
- [James Craig, 1st Viscount Craigavon](#) (1871–1940), British politician and Prime Minister of Northern Ireland
- [James Craig, 2nd Viscount Craigavon](#) (1906–1974), British politician
- [Sir James Craig \(diplomat\)](#) (born 1924), British ambassador, lecturer, writer
- [James Craig \(County Antrim, 20th century\)](#) (1931–1974), Northern Ireland politician

### Contents [hide]

- 1 Public officials
- 2 Sports personalities
- 3 Actors
- 4 Other people
- 5 Characters
- 6 Other

## Actors [edit]

- [James Craig \(actor\)](#) (1912–1985), American actor
- [James Philip Craig \(actor\)](#) (born 1950), British actor

## Other people [edit]

- [James Craig \(architect\)](#) (1739–1795), Scottish architect
- [Sir James Gibson-Craig](#) (1765–1850), 1st [Baronet Gibson-Craig](#)
- [Sir James Henry Gibson-Craig](#) (1841–1908), 3rd [Baronet Gibson-Craig](#)
- [James Craig \(VC\)](#) (1824–1861), British soldier
- [James Ireland Craig](#) (1868–1952), British cartographer; creator of [Craig retroazimuthal projection](#)
- [James Craig \(loyalist\)](#) (c. 1941–1988), Ulster Defence Association member
- [James Humbert Craig](#) (1877–1944), Irish painter

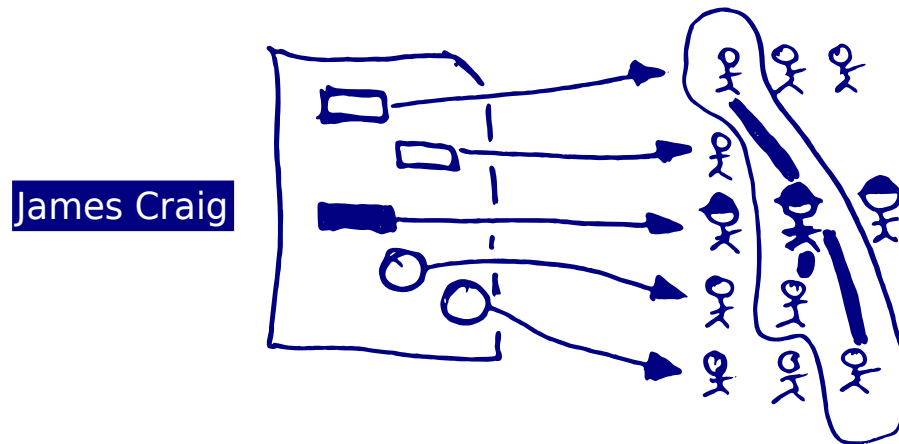
## Characters [edit]

- [Jim Craig](#) (*The Man from Snowy River*) in films
- [Jim Craig](#) (*One Life to Live*) on television
- [James Craig](#) (*General Hospital*) on television

# Method 4: Joint Assignment Models

Step 1: Identify all entity mentions in text

Step 2: For each mention retrieve candidates



Step 3: Select the entity  $e$  that maximizes:

$$p(\text{James Craig}, e) \prod (p(o, e) p(e, e))$$

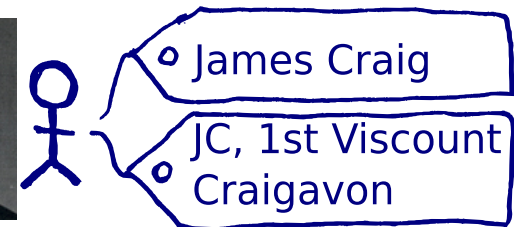
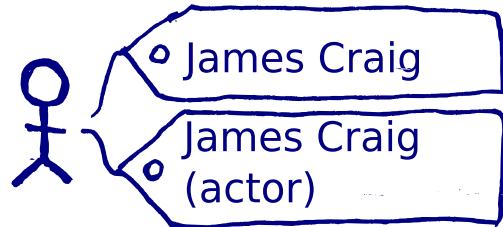
across all neighbor entities  $e$

# Method 4 Example: Candidates

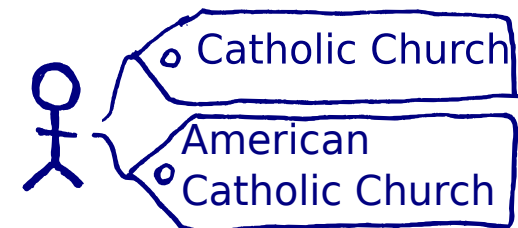
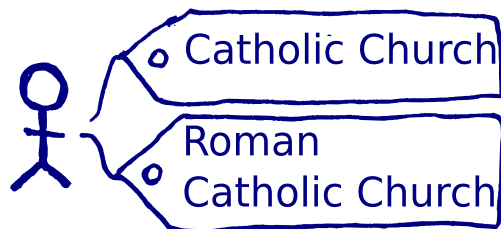
Northern Ireland



James Craig



Catholics

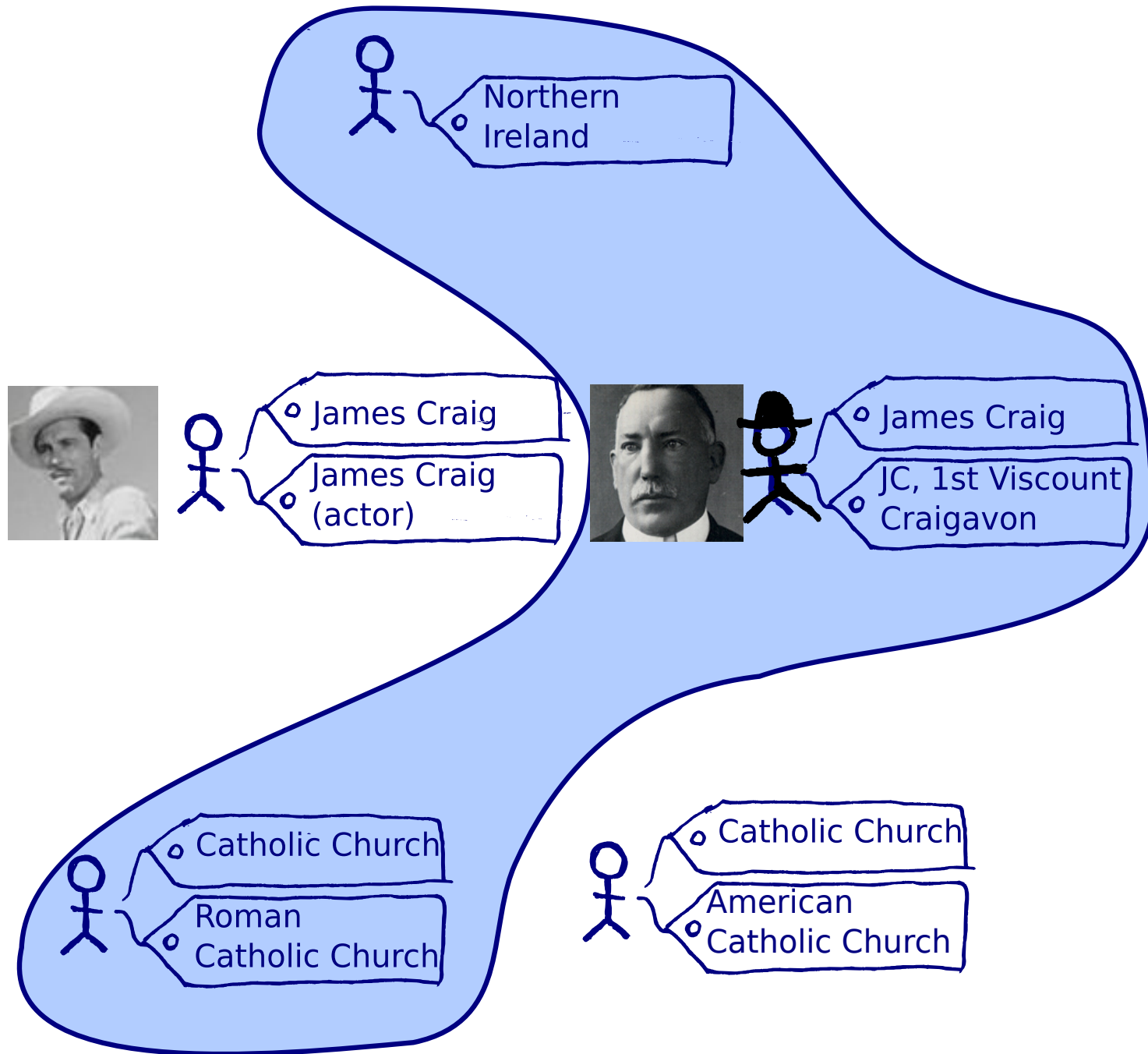


# Method 4 Example: Correct Selection

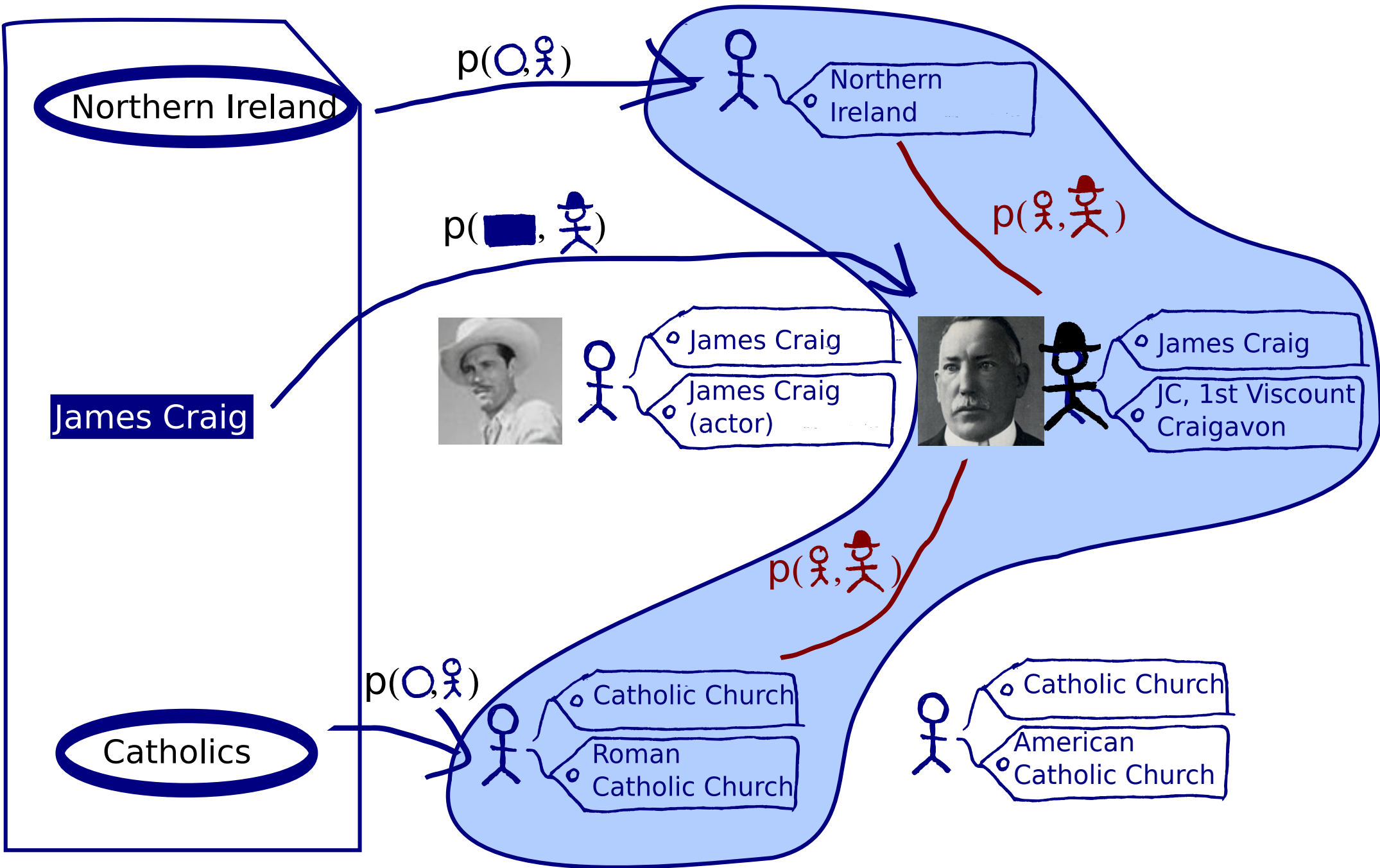
Northern Ireland

James Craig

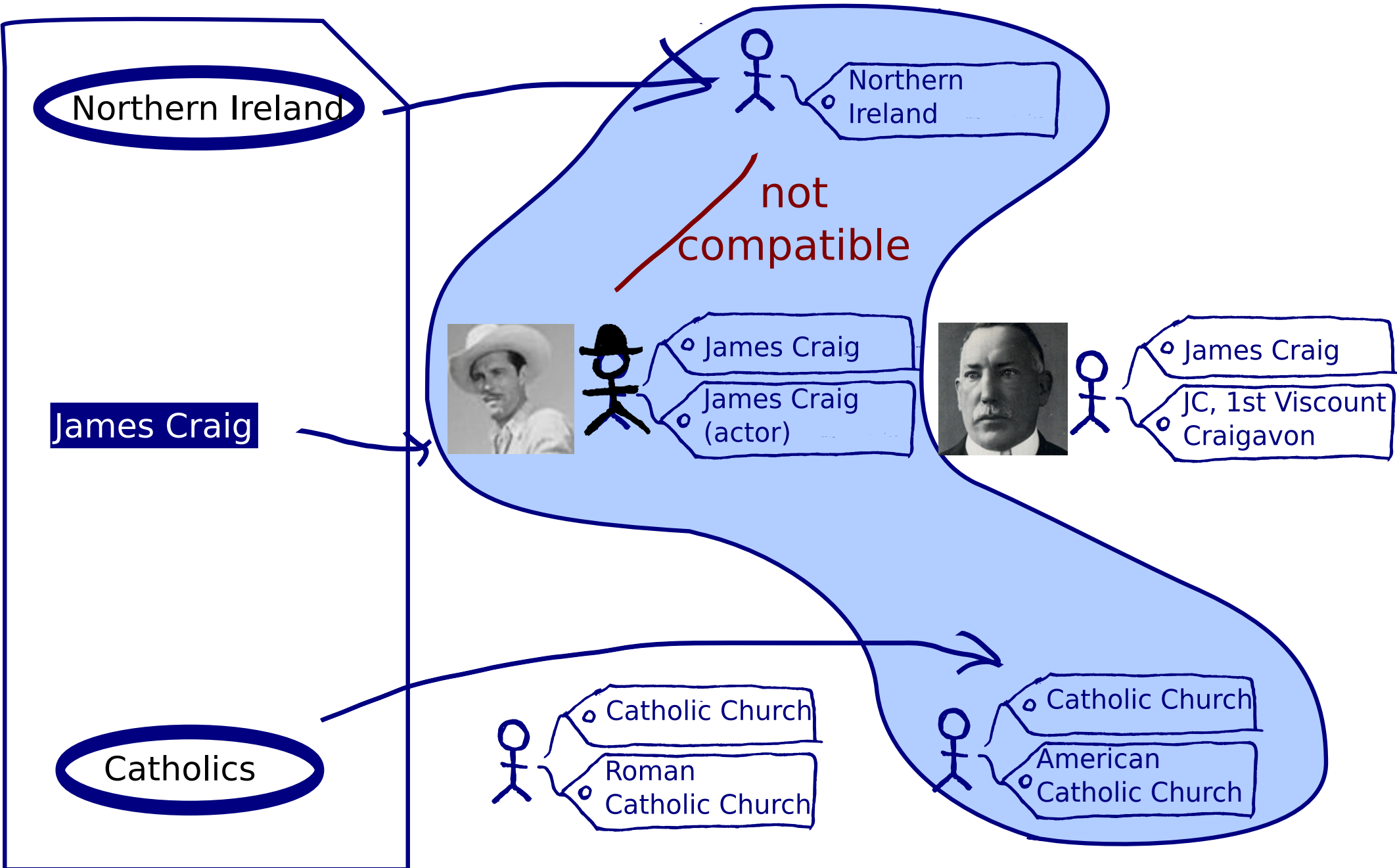
Catholics



# Method 4 Example: Scoring



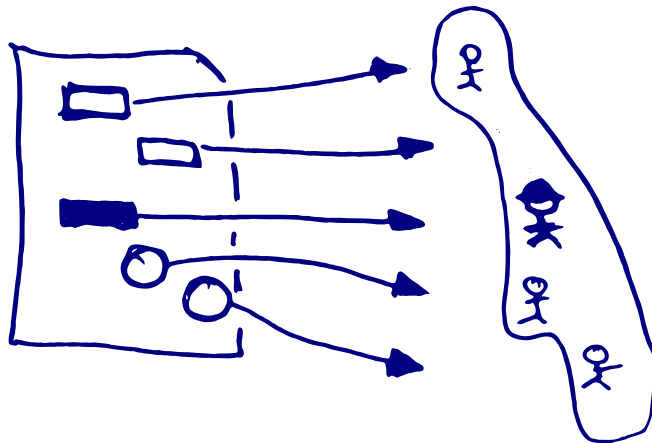
# Method 4 Example: Wrong Selection



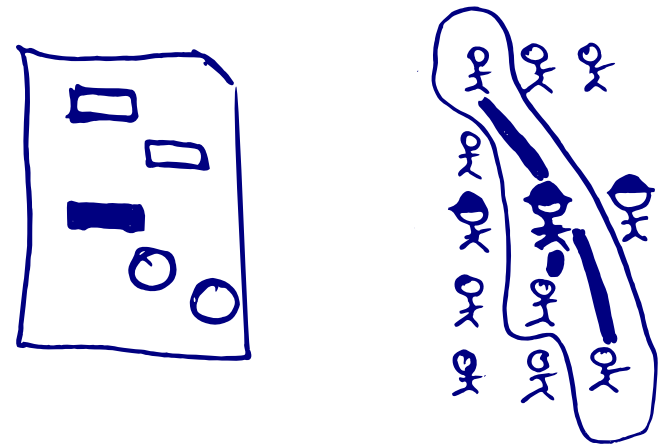
# Method 4: Learn Similarities

As in Method 2, learn feature-based similarity

mention-entity  
similarity  $p(\blacksquare, \text{人})$   
 $p(\bigcirc, \text{人})$



entity-entity  
similarity  $p(\text{人}, \text{人})$



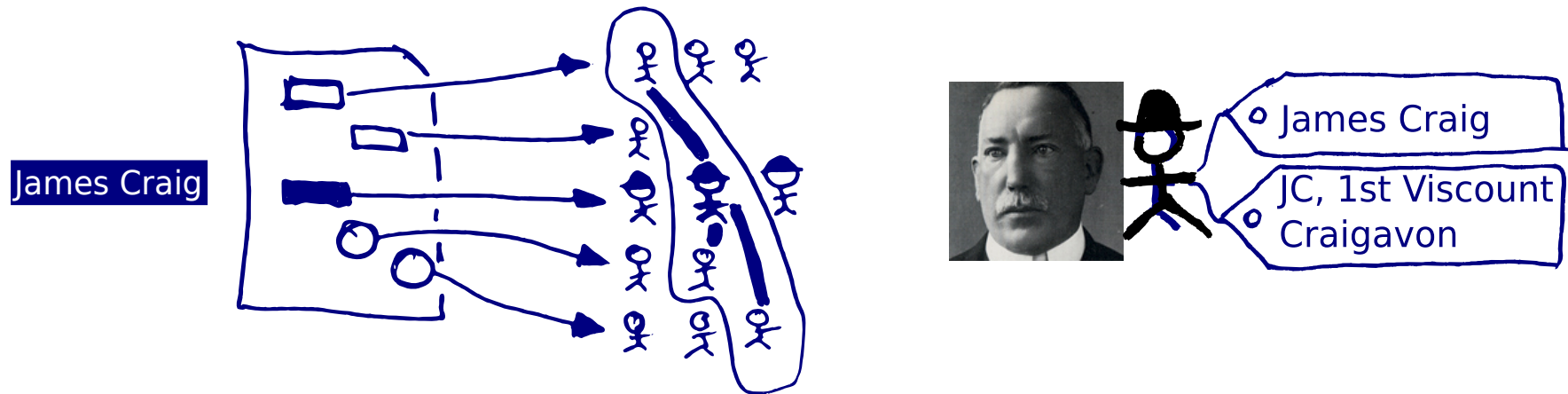
entity-entity similarity features:  
mutual links, same categories, RDF relations



# Method 4: Joint Assignment Models

Step 1: Identify all entity mentions in text

Step 2: For each mention retrieve candidates



Step 3: Select the entity  that maximizes:

$$p(\text{James Craig}, \text{stick figure}) \prod (p(\text{O}, \text{stick figure}) p(\text{stick figure}, \text{stick figure}))$$

across all neighbor entities 

## Method 4: Pros and Cons

Pro: Can mutually resolve uncertainty

Con: Requires a pool of candidates  
(trade-off runtime versus recall)

Con: expensive inference problem

May still fail on less popular James Craigs  
or when context does not resolve  
ambiguities.

# Method 5: Joint Retrieval Model

Step 1: Identify all entity mentions in text

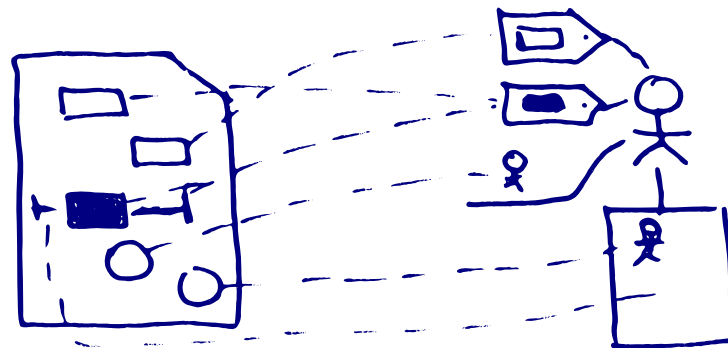
Step 2: For each query mention:

Issue a search query including

query, neighboring mentions, sentence

Weighting each "ingredient" differently

Intuition: structured matching of text to KB



# Names and Links on Wikipedia

## Northern Ireland

**Northern Ireland** (Irish: *Tuaisceart Éireann*<sup>[4]</sup> [t̪ˠuəʃc̪əɾˠt̪ˠ ˈeːɾʲəɲˠ] ( listen); Ulster Scots: *Norlin Airlann* or *Norlin Airlan*) is a part of the United Kingdom in the north-east of the island of Ireland. It is variously described as a country, province or region of the UK, amongst other terms.<sup>[5][6][7]</sup> Northern Ireland shares a border to the south and west with the Republic of Ireland. As of 2011, its

## James Craig, 1st Viscount Craigavon

From Wikipedia, the free encyclopedia

*For other uses, see James Craig (disambiguation).*

**James Craig, 1st Viscount Craigavon, PC, PC (NI)** (8 January 1871 – 24 November 1940), was a prominent Irish unionist politician, leader of the Ulster Unionist Party and the first Prime Minister of Northern Ireland. He was created a baronet in 1918 and raised to the Peerage in 1927.

The Right Honourable

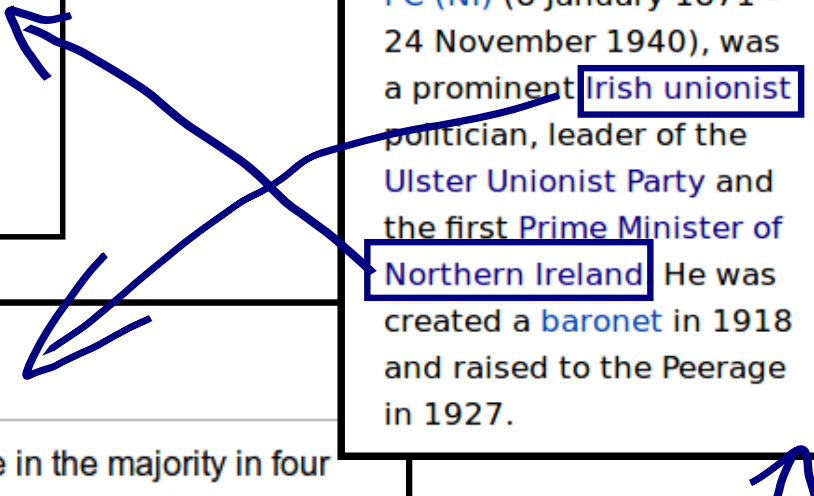
The Viscount Craigavon  
PC



## Unionism in Ireland

In Ireland, unionism, also called **Ulster unionism** in the context of Northern Ireland or – chiefly historically – **Ulster** specifically, refers to an ideology which favours the continuation of some form of political union between the islands of Ireland and Great

Unionists were in the majority in four counties of the Ulster (Antrim, Londonderry, Down and Armagh), and formed a large minority in the remaining counties of Fermanagh and Tyrone. Sir Edward Carson had expressly urged the new Prime Minister, Sir James Craig, to ensure absolute equality in the treatment of Catholics, so to guarantee the stability



# Mining Name Variants and Neighbors

## Northern Ireland



Northern  
Ireland

## James Craig, 1st Viscount Craigavon



Sir James Craig

1st Viscount  
Craigavon

Ulster Unionists

Northern Ireland

Prime Minister of  
Northern Ireland

## Unionism in Ireland



Irish Unionist

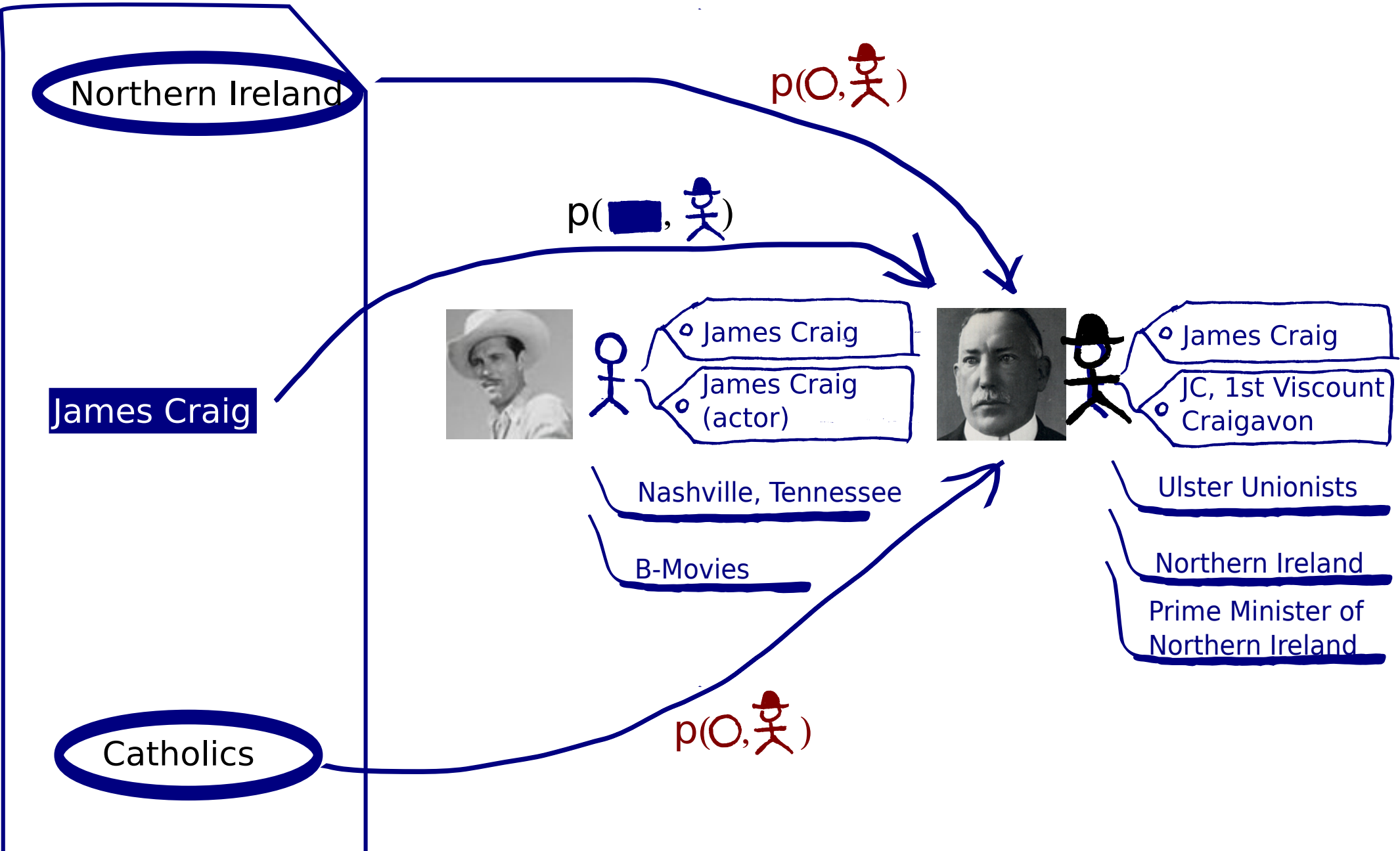
Unionism in  
Ireland

Ulster

Northern Ireland

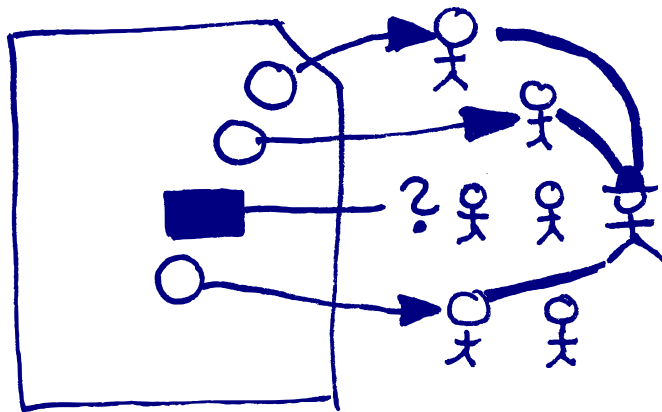
James Craig, 1st  
Viscount Craigavon

# Method 5 Example: Scoring



# Connection between 4 and 5

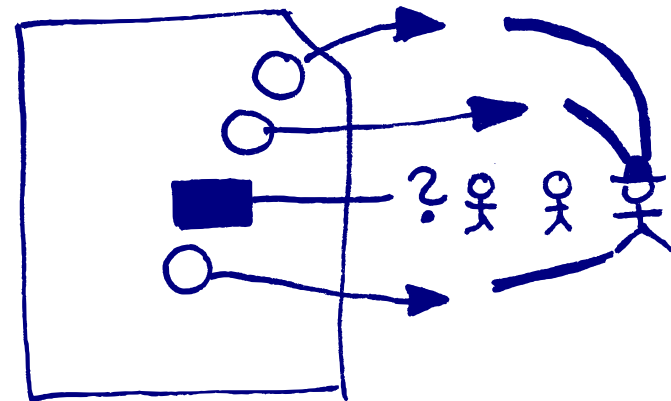
Method 4



$$p(\blacksquare, \text{stick}) \Pi ( p(\circ, \text{stick}) p(\text{stick}, \text{stick}) )$$

Method 5

Integrate over



$$p(\blacksquare, \text{stick}) \Pi ( p(\circ, \text{stick}) )$$

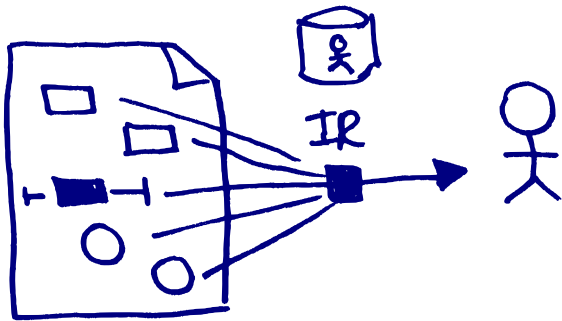
Requires iterative optimization

Can be solved inside a search engine

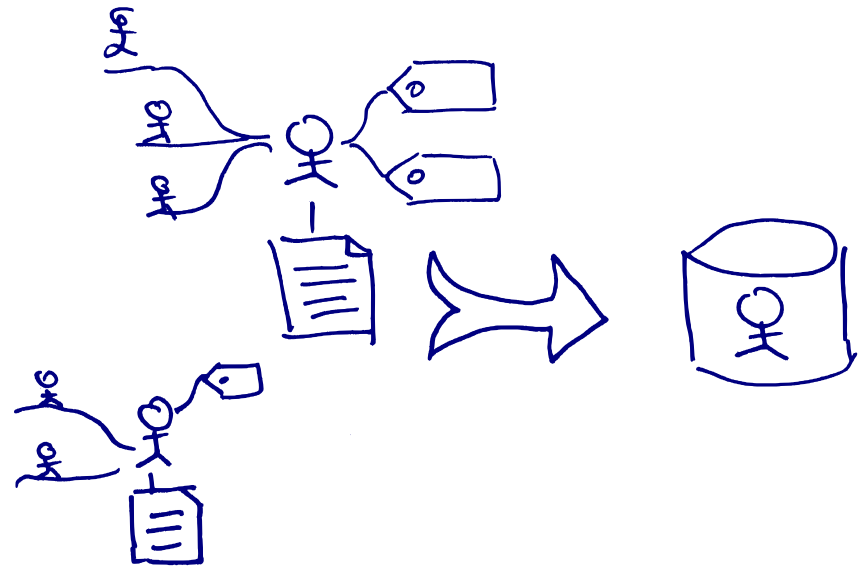
- Q: Query String
- V: Name Variants
- M: Neighbor Mentions
- S: Sentence

# Need a Search Index for the KB

Identify context of query mention



Preprocessing:  
build a special KB Index



neighbor-entity similarity features:  
neighbor occurs in entity's text  
neighbor is title of inlinks/outlinks



# Special Wikipedia Index

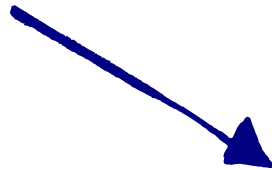
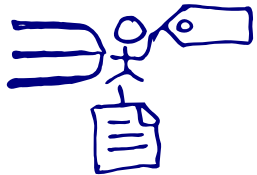


- James Craig
- JC, 1st Viscount Craigavon

Ulster Unionists

Northern Ireland

Prime Minister of Northern Ireland



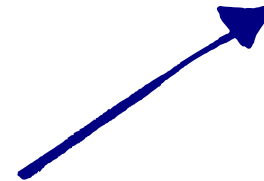
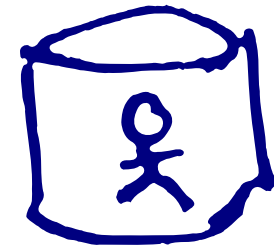
JC, 1st Viscount Craigavon

Ulster Unionists

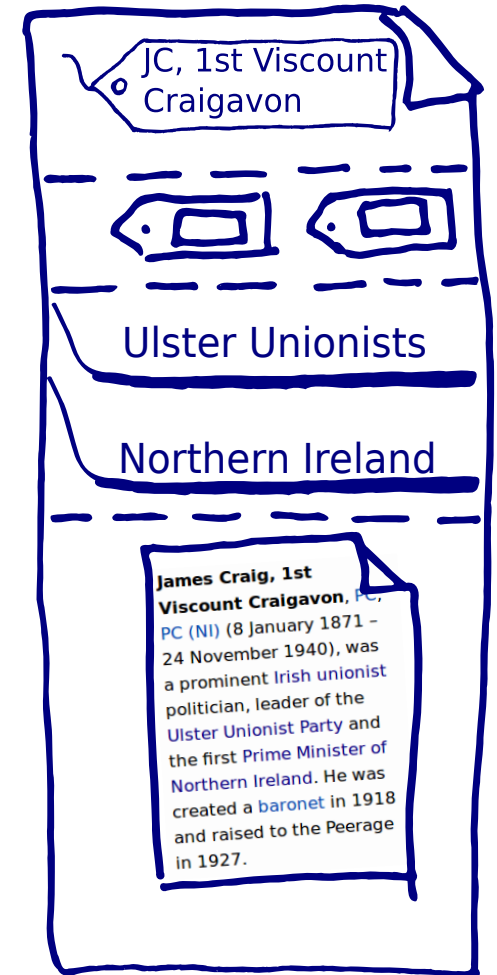
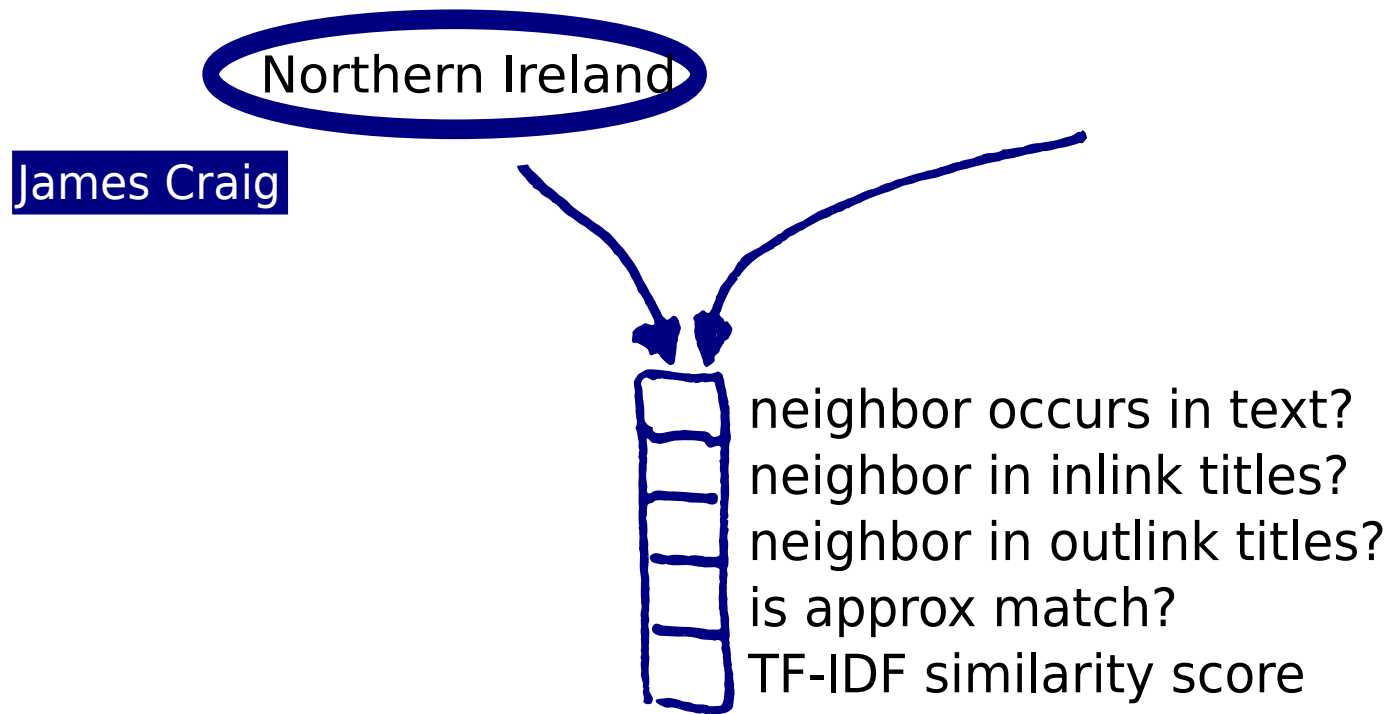
Northern Ireland

**James Craig, 1st Viscount Craigavon, PC, PC (NI)** (8 January 1871 – 24 November 1940), was a prominent Irish unionist politician, leader of the Ulster Unionist Party and the first Prime Minister of Northern Ireland. He was created a [baronet](#) in 1918 and raised to the Peerage in 1927.

Search Index with special Fields



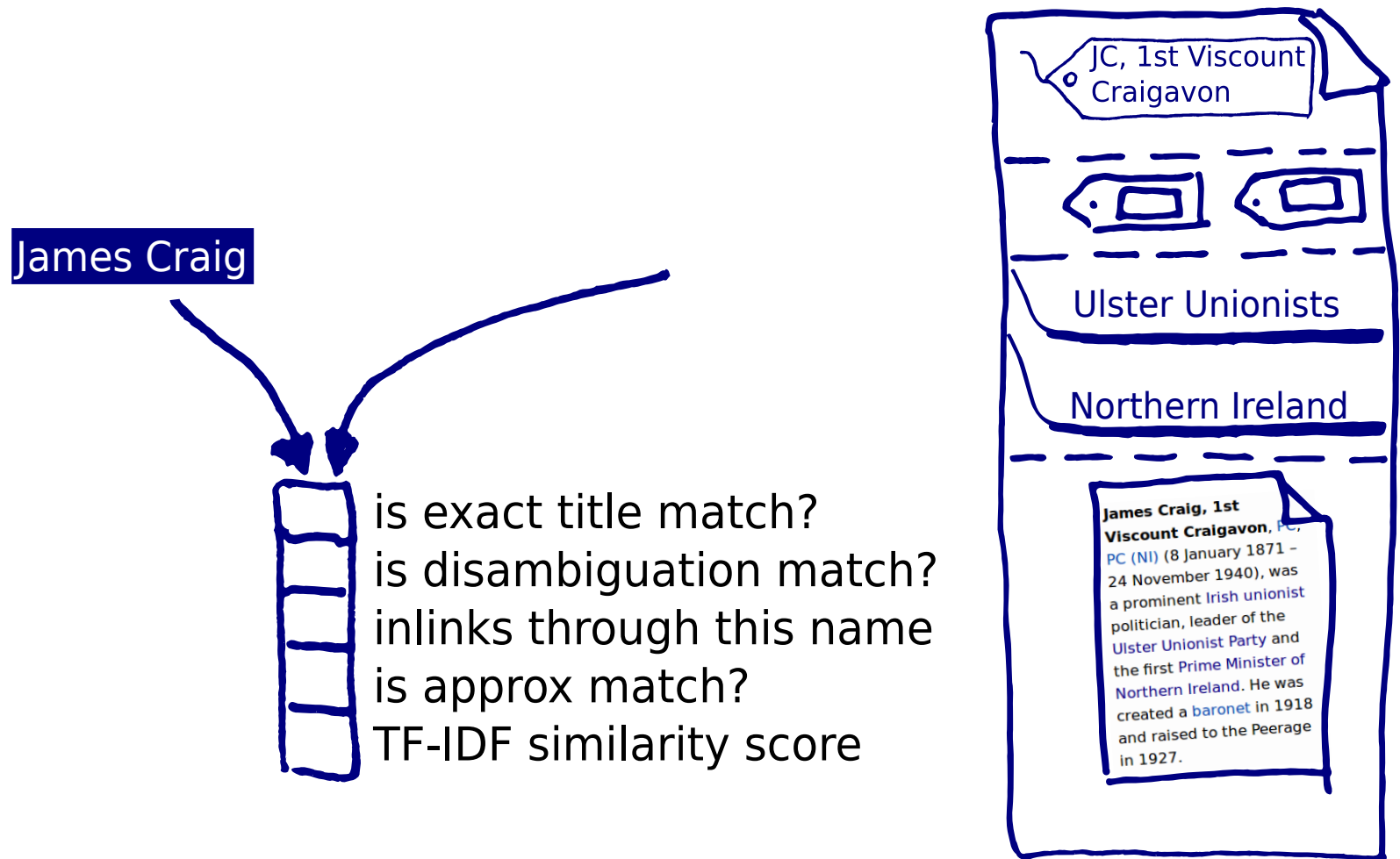
# Neighbor-Entity Features



Machine learn the feature weights on training data (e.g. learning to rank)

$$p(O, \text{人})$$

# Query mention-Entity Features

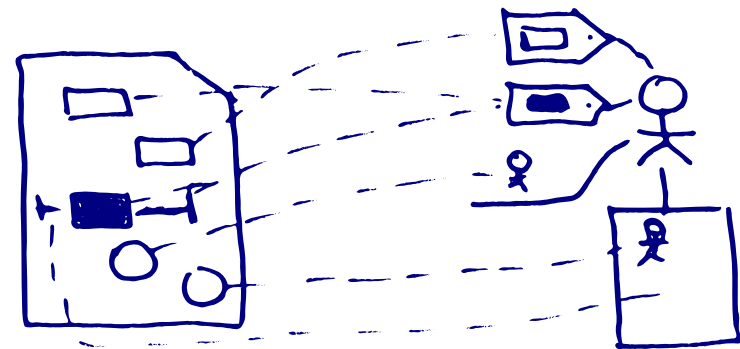
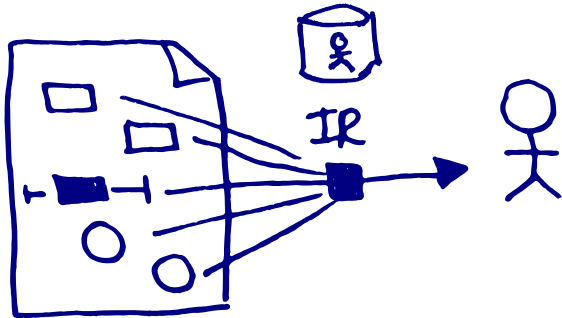


Machine learn the feature weights  
on training data (e.g. learning to rank)  $p(\blacksquare, \text{stick figure})$

# Method 5: Joint Retrieval Model

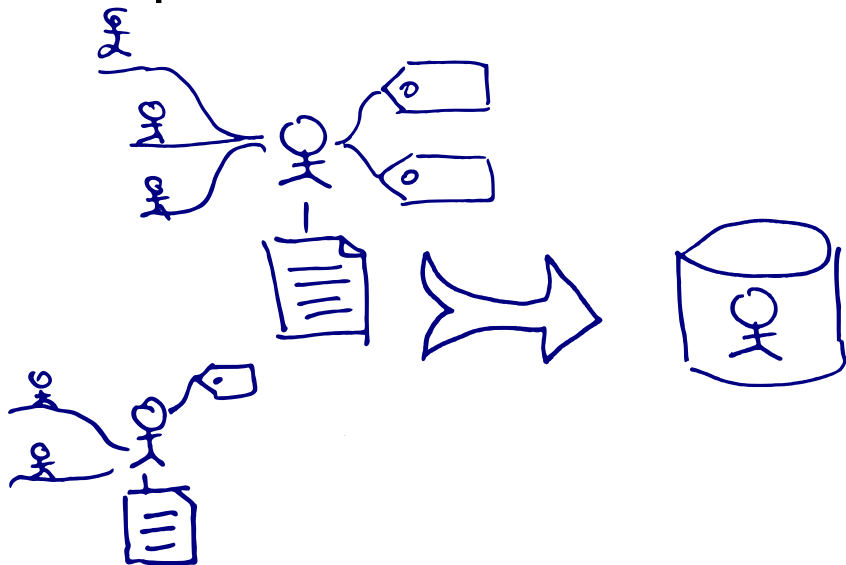
Issue the Entity Linking IR

Actually: structured matching



with special KB Index

Select the entity that maximizes:



$$p(\blacksquare, \text{stick figure}) \prod p(\circ, \text{stick figure})$$

## Method 5: Pros and Cons

Pro: Similar to joint assignment, but cheaper

Pro: Does not require pools (optimize in IR)

Pro: Can be combined with Machine Learning (Method 2) to improve precision.

Con: Fails when context is misleading

# Really Difficult Example

Example Query:

ABC shot "Lost" in Australia

True entity:

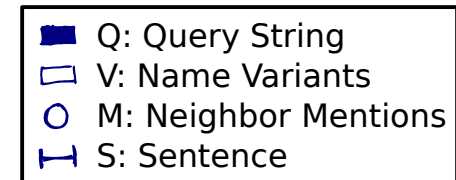
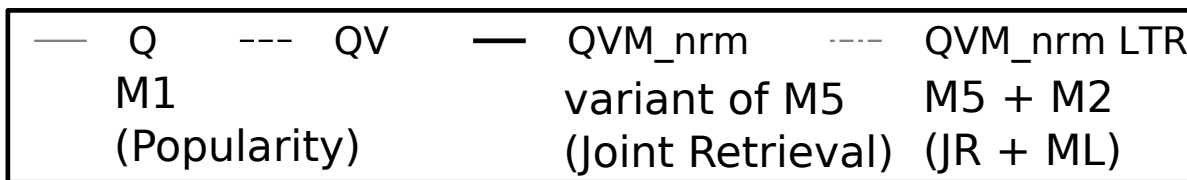
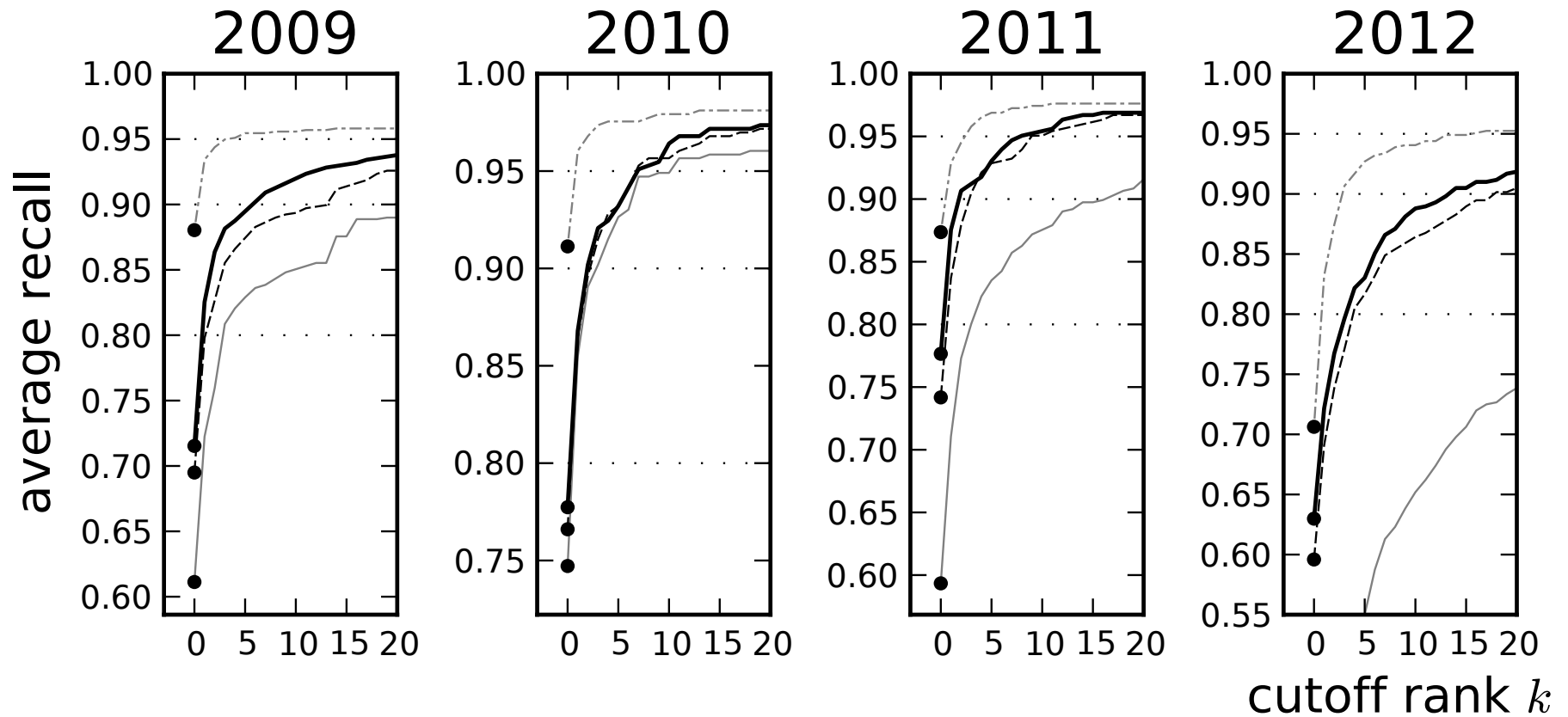
American Broadcasting Company

Context "Australia" and mention similarity  
will point instead to

Australian Broadcasting Corporation

Approach: Identify misleading neighbors (variant of M5)

# TAC KBP Entity Linking Task



# References

## **M1 Popularity / Keyphraseness:**

Mihalcea et al. In CIKM, 2007.

Wikify!: linking documents to encyclopedic knowledge.

## **M2 Machine Learn Mention-to-Entity Similarity**

Bunescu et al. In EACL, 2006.

"Using Encyclopedic Knowledge for Named entity Disambiguation."

M. Dredze, et al. In ACL, 2010

"Entity disambiguation for knowledge base population".

## **M4 Joint Assignment**

Silviu Cucerzan. In EMNLP-CoNLL, 2007.

"Large-scale named entity disambiguation based on wikipedia data."

Ratinov et al. ACL 2011.

"Local and global algorithms for disambiguation to wikipedia."

Entity-to-Entity Features: Ceccarelli et al. In CIKM, 2013.

"Learning relatedness measures for entity linking."

## **M5 Joint Retrieval Model**

Dalton et al. In OAIR, 2013.

"A neighborhood relevance model for entity linking."

**more:** <http://nlp.cs.rpi.edu/kbp/2014/elreading.html>

<http://www.mendeley.com/groups/3339761/entity-linking-and-retrieval/>



# Toolkits & Online Demos

List of toolkits:

<http://nlp.cs.rpi.edu/kbp/2014/tools.html>

Several Online Demos:

UIUC Wikifier

<http://cogcomp.cs.illinois.edu/demo/wikify/>

TagMe!

<http://tagme.di.unipi.it/>

AIDA

<https://gate.d5.mpi-inf.mpg.de/webaida/>

# UIUC Wikifier

Wikipedia Demo Results x

cogcomp.cs.illinois.edu/demo/wikify/results.php

## COGNITIVE COMPUTATION GROUP

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

News Research People Software Demos Publications Resources Schedule

### DEMOS

Problems? Email mssammor

#### Wikification Demo Results

The Wikification system has identified the following entities with Wikipedia articles. Click on an entity to visit the corresponding Wikipedia page. Hover over links to view the categories associated with each entity.

[Northern Ireland](#) has a [population](#) of about one and a half million people. At the time of partition in [1921](#) [Protestants / unionists](#) had a [two-thirds majority](#) in the region. The first Prime Minister of [Northern Ireland](#), [Sir James Craig](#), described the state as having 'a [Protestant Parliament](#) for a [Protestant](#) people.' The state effectively discriminated against [Catholics](#) in housing, [jobs](#), and [political representation](#).

#### Categories

- baronet
- viscount
- member
- constituency
- officer
- leaders

...wikipedia.org/.../James\_Craig,...

# TagMe!

The screenshot shows a web browser window with the URL `tagme.di.unipi.it/tagme.jsp`. The page header indicates it was developed by Paolo Ferragina and Ugo Scaiella at A<sup>3</sup> Lab, Dipartimento di Informatica, University of Pisa.

The main interface includes an "Input Text" area with language selection for Italian and English. A text box contains a paragraph about Northern Ireland. To the right of the text box is a vertical slider labeled "Many links" at the top and "Few links" at the bottom, with a "Reset" button and a "TAGME!" button below it.

A tooltip is displayed over the text "James Craig", showing his full name and a brief biography: "James Craig, 1st Viscount Craigavon, PC, PC (NI) (8 January 1871 – 24 November 1940), was a prominent Irish unionist politician, leader of the Ulster Unionist Party and the first Prime Minister of Nor...".

Below the text box, the "Tagged text" section shows the same paragraph with several words underlined in blue, indicating they have been tagged. A mouse cursor is hovering over the tag "Sir James Craig".

At the bottom left, a snippet of a search result is visible: `...wikipedia.org/.../James Craig, 1...`

# AIDA (prior+sim+coherence)

The screenshot shows the AIDA Web interface in a browser window. The URL is <https://gate.d5.mpi-inf.mpg.de/webaida/>. The interface is divided into several sections:

- Disambiguation Method:** Three buttons are shown: "prior" (blue), "prior+sim" (light blue), and "prior+sim+coherence" (green, selected).
- Parameters:** A section with a blue header containing sliders for:
  - Prior-Similarity-Coherence balancing ratio: "prior VS. sim. balance = 0.4" and "(prior+sim.) VS. coh. balance 0.6".
  - Ambiguity degree: 5.
  - Coherence robustness test threshold: 0.9.
- Entities Type Filters:** An empty text input field.
- Mention Extraction:** Two buttons: "Stanford NER" (green) and "Manual" (blue). Below them is a note: "You can manually tag the mentions by putting them between [] and []. HTML Tables are automatically disambiguated in the manual mode."
- Fast Mode:** A green button labeled "Enabled".
- Input Type:** "TEXT Overall runtime:0 sec(s)".
- Text Sample:** A paragraph of text with entity mentions in brackets: "Northern Ireland [Northern Ireland] has a population of about one and a half million people. At the time of partition in 1921 Protestants / unionists had a two-thirds majority in the region. The first Prime Minister of Northern Ireland [Northern Ireland], Sir James Craig [James Craig, 1st Viscount Craigavon], described the state as having 'a Protestant Parliament for a Protestant people.' The state effectively discriminated against Catholics [Irish Catholic] in housing, jobs, and political representation."
- Entity Cards:** Three cards are displayed below the text:
  - "Northern Ireland" with a map icon.
  - "Irish Catholic" with a cross icon.
  - "James Craig, 1st Viscount" with the YAGO logo and the tagline "select knowledge".
- Navigation:** At the bottom, there are buttons for "Examples" and "YAGOTypes", and a footer with "Run Information" and "Graph".

# AIDA (prior only)

The screenshot shows the AIDA Web interface in a browser window. The address bar displays <https://gate.d5.mpi-inf.mpg.de/webaida/>. The interface is divided into several sections:

- Disambiguation Method:** Three buttons are shown: "prior" (highlighted in green), "prior+sim", and "prior+sim+coherence".
- Entities Type Filters:** An empty text input field.
- Mention Extraction:** Two buttons: "Stanford NER" (highlighted in green) and "Manual". Below them is a note: "You can manually tag the mentions by putting them between [[ and ]]. HTML Tables are automatically disambiguated in the manual mode."
- Fast Mode:** A button labeled "Enabled" (highlighted in green).

At the bottom left, there are two blue buttons: "Examples" and "YAGOTypes".

The main content area on the right shows the input text: "Input Type:TEXT Overall runtime:0 sec(s)". Below this, a text snippet is displayed with several entity mentions highlighted in blue boxes: "Northern Ireland", "James Craig (actor)", and "Catholics".

Below the text snippet, three knowledge cards are shown:

- Northern Ireland:** Includes a small map icon.
- Catholic Church:** Includes the YAGO logo and the text "select knowledge".
- James Craig (actor):** Includes a small portrait icon.

# Another Example: Lisa Fletcher

In 1990, Fletcher began her on-air career while still in college with [KEZI-TV](#) in [Eugene, Oregon](#).<sup>[3]</sup> After graduating, she accepted a job anchoring daily newscasts for [KTVZ-TV](#) in [Bend, Oregon](#). In 1993, she returned to Eugene and KEZI where she was the main anchor and investigative reporter. She covered natural disasters, including some of the Northwest's most devastating floods and forest fires. She also anchored coverage of one of the nation's worst high school shootings, the May 1998 tragedy at [Thurston High School](#) in [Springfield, Oregon](#), adjacent to Eugene. The coverage she anchored of the [Thurston High shootings](#) was honored with an [Edward R. Murrow Award](#) for broadcast excellence.<sup>[4]</sup>

**Lisa Fletcher**



**Alma mater** [University of Oregon](#) B.S., 1990

**Occupation** [Television journalist](#)

# UIUC Wikifier

Wikification Demo Results

The Wikification system has identified the following entities with Wikipedia articles.  
Click on an entity to visit the corresponding Wikipedia page.  
Hover over links to view the categories associated with each entity.

[Lisa Simpson](#) is a character in the [animated television series The Simpsons](#). She is the daughter of one of the **nation's** worst [high school](#) shootings, the [May 1998](#) [High School](#) in [Springfield, Oregon](#).

**Categories**

- character
- vegetarian
- introduction
- musician

[Back](#)

[en.wikipedia.org/wiki/Lisa\\_Simpson](http://en.wikipedia.org/wiki/Lisa_Simpson)

# TagMe!

The screenshot shows a web browser window with the URL `tagme.di.unipi.it/tagme.jsp`. The page content includes:

- A header section with a paragraph: "For instance, you can compute semantic relatedness between topics identified by TAGME, or enable the special parser for Twitter messages. Check the [RESTful API page](#) for further details."
- A developer credit: "Developed by [Paolo Ferragina](#) and [Ugo Scaiella](#) at [A<sup>3</sup> Lab](#) Dipartimento di Informatica, University of Pisa."
- An "Input Text" section with language selection buttons for "Italiano" and "English".
- A text input area containing: "Lisa also anchored coverage of one of the nation's worst high school shootings, the May 1998 tragedy at Thurston High School in Springfield, Oregon".
- A vertical slider on the right labeled "Many links" at the top and "Few links" at the bottom, with a "Reset" button and a "TAGME!" button.
- A tooltip for "Lisa Simpson" with the text: "Lisa Marie Simpson is a fictional main character in the animated television series The Simpsons. She is the middle child of the Simpson family. Voiced by Yeardley Smith, Lisa first appeared on televis...".
- The same text as the input area, but with blue underlines under "Lisa", "anchored coverage", "nation's", "worst high school shootings", "May 1998 tragedy", "Thurston High School", and "Springfield, Oregon".
- A red oval highlights a suggestion box at the bottom containing the text: `en.wikipedia.org/wiki/Lisa Simpson`.



# AIDA

The screenshot displays the AIDA Web interface in a browser window. The address bar shows the URL <https://gate.d5.mpi-inf.mpg.de/webaida/>. The interface is divided into several sections:

- Disambiguation Method:** Three buttons are shown: "prior", "prior+sim", and "prior+sim+coherence". The "prior+sim+coherence" button is highlighted in green.
- Parameters:** A section with a blue header containing sliders for:
  - Prior-Similarity-Coherence balancing ratio: "prior VS. sim. balance = 0.4" and "(prior+sim.) VS. coh. balance 0.6".
  - Ambiguity degree: 5.
  - Coherence robustness test threshold: 0.9.
- Entities Type Filters:** An empty input field.
- Mention Extraction:** Two buttons: "Stanford NER" (green) and "Manual" (blue). Below them is a note: "You can manually tag the mentions by putting them between [ and ]. HTML Tables are automatically disambiguated in the manual mode."
- Fast Mode:** A green button labeled "Enabled".
- Buttons:** "Examples" and "YAGOTypes" buttons are located at the bottom left.

The main content area on the right shows the input text and its disambiguated output:

**Input Type:TEXT Overall runtime:2 sec(s)**

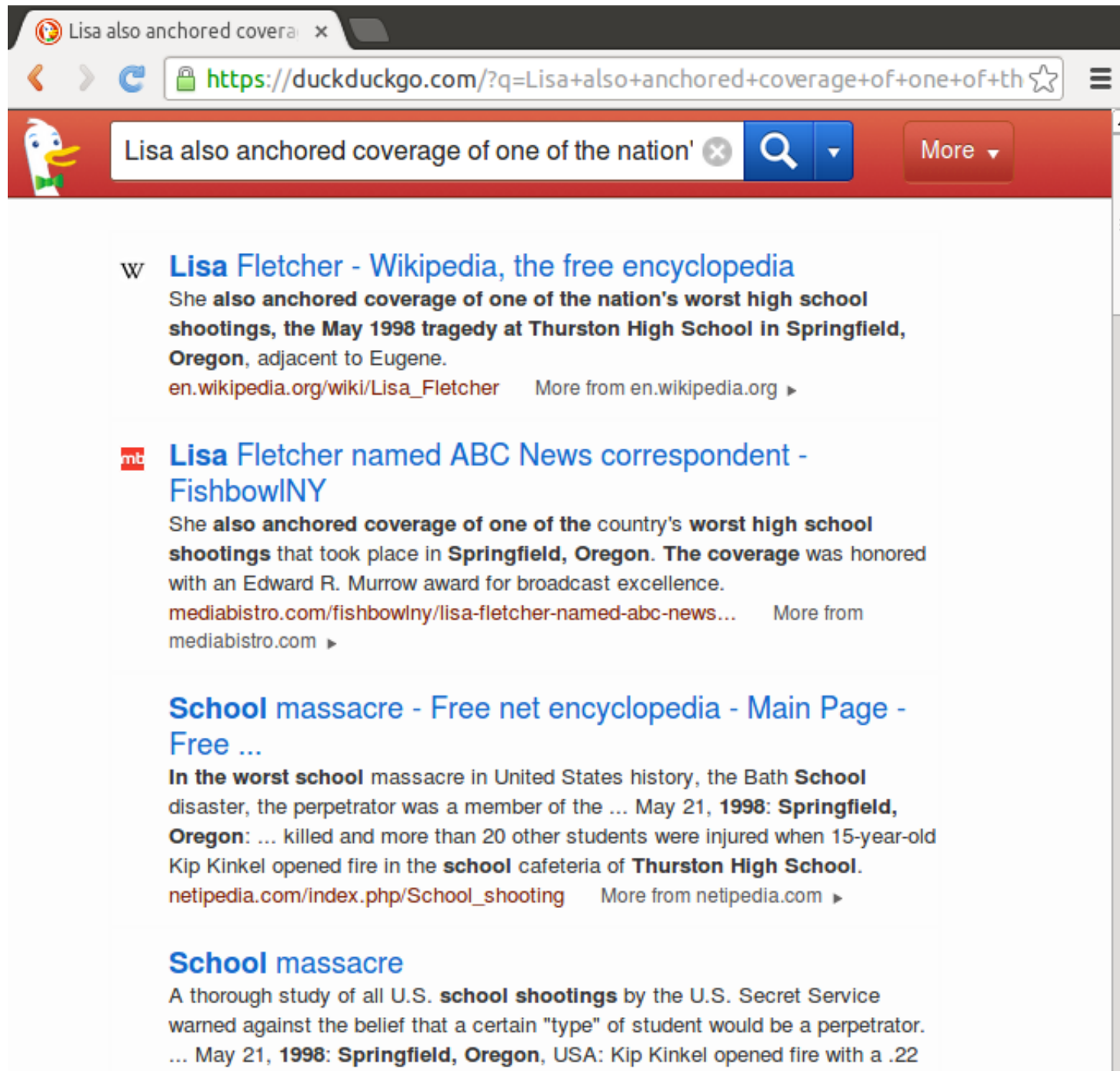
**Input Text:** Lisa [Lisa Simpson] also anchored coverage of one of the nation's worst high school shootings, the May 1998 tragedy at Thurston High School [Thurston High School] in Springfield [Springfield, Oregon], Oregon [Eugene, Oregon].

**Disambiguated Results:** A list of entities with associated images and YAGO logos:

- Eugene, Oregon
- Lisa Simpson (highlighted with a red oval)
- Springfield, Oregon
- Thurston High School

At the bottom right, there are buttons for "Run Information", "Graph", and "Removal Steps".

# Search Engine (DuckDuckGo)



The screenshot shows a web browser window with the DuckDuckGo search engine. The search bar contains the query "Lisa also anchored coverage of one of the nation's worst high school shootings". The search results are displayed below the search bar, featuring several entries with blue titles and red icons.

**W** [Lisa Fletcher - Wikipedia, the free encyclopedia](#)  
She **also anchored coverage of one of the nation's worst high school shootings, the May 1998 tragedy at Thurston High School in Springfield, Oregon**, adjacent to Eugene.  
[en.wikipedia.org/wiki/Lisa\\_Fletcher](https://en.wikipedia.org/wiki/Lisa_Fletcher) More from en.wikipedia.org ▶

**mb** [Lisa Fletcher named ABC News correspondent - FishbowlnY](#)  
She **also anchored coverage of one of the country's worst high school shootings** that took place in **Springfield, Oregon**. The coverage was honored with an Edward R. Murrow award for broadcast excellence.  
[mediabistro.com/fishbowlny/lisa-fletcher-named-abc-news...](https://mediabistro.com/fishbowlny/lisa-fletcher-named-abc-news...) More from mediabistro.com ▶

[School massacre - Free net encyclopedia - Main Page - Free ...](#)  
**In the worst school massacre** in United States history, the Bath School disaster, the perpetrator was a member of the ... **May 21, 1998: Springfield, Oregon:** ... killed and more than 20 other students were injured when 15-year-old Kip Kinkel opened fire in the **school** cafeteria of **Thurston High School**.  
[netipedia.com/index.php/School\\_shooting](https://netipedia.com/index.php/School_shooting) More from netipedia.com ▶

[School massacre](#)  
A thorough study of all U.S. **school shootings** by the U.S. Secret Service warned against the belief that a certain "type" of student would be a perpetrator. ... **May 21, 1998: Springfield, Oregon, USA:** Kip Kinkel opened fire with a .22

# Participate!

TAC KBP Entity Linking Task

<http://nlp.cs.rpi.edu/kbp/2014/>

SIGIR Entity Recognition and Disambiguation Challenge

<http://web-ngram.research.microsoft.com/erd2014/>

INEX 2014 Tweet Contextualization Track

<https://inex.mmci.uni-saarland.de/tracks/qa/>

## **Questions?**

email: [dietz@cs.umass.edu](mailto:dietz@cs.umass.edu)

web: <http://ciir.cs.umass.edu/~dietz/>

