

# *Entity Linking - Part I*

Pawan Goyal

CSE, IIT Kharagpur

Week 10, Lecture 1

# Entity Linking: Introduction

- Entity Linking is the task of identifying all mentions in text of a specific entity from a database or ontology

# Entity Linking: Introduction

- Entity Linking is the task of identifying all mentions in text of a specific entity from a database or ontology
- Researchers have used Wikipedia, YAGO, Freebase etc. as the databases

# Entity Linking: Introduction

- Entity Linking is the task of identifying all mentions in text of a specific entity from a database or ontology
- Researchers have used Wikipedia, YAGO, Freebase etc. as the databases
- Typically broken down into two main phases
  - ▶ Candidate Selection (entity annotation)
  - ▶ Reference disambiguation or entity resolution

# Entity Linking: Introduction

- Entity Linking is the task of identifying all mentions in text of a specific entity from a database or ontology
- Researchers have used Wikipedia, YAGO, Freebase etc. as the databases
- Typically broken down into two main phases
  - ▶ Candidate Selection (entity annotation)
  - ▶ Reference disambiguation or entity resolution
- Entity linking needs to handle
  - ▶ Name variations (entities are referred to in many different ways)
  - ▶ Entity ambiguity (the same string can refer to more than one entity)

# Entity Linking: Introduction

- We will take Wikipedia as the knowledge base to understand the task.
- With Wikipedia as the knowledge base, this task is commonly known as *Wikification*.

# What is Entity Linking?

Input Text

Italiano English

momentum) degeneracy is removed due to a geometric gradient onto a metasurface. The alliance of spin optics and metamaterials offers the dispersion engineering of a structured matter in a polarization helicity-dependent manner. We show that polarization-controlled optical modes of metamaterials arise where the spatial inversion symmetry is violated. The emerged spin-split dispersion of spontaneous emission originates from the spin-orbit interaction of light, generating a selection rule based on symmetry restrictions in a spin-optical metamaterial. The inversion asymmetric metasurface is obtained via anisotropic optical antenna patterns. This type of metamaterial provides a route for spin-controlled nanophotonic applications based on the design of the metasurface symmetry properties.

Many links

Few links

Reset

**TAGME!**

Tagged text

Topics

Spin [optics](#) provides a route to [control light](#), whereby the [photon helicity](#) (spin [angular momentum](#)) [degeneracy](#) is removed due to a [geometric gradient](#) onto a metasurface. The alliance of sp [matter](#) in a [p](#) [optical mode](#); [emerged spin](#) [interaction](#) of [optical meta](#) [optical antenna patterns](#). This [type](#) of metamaterial provides a route for spin-controlled

**Degenerate energy levels**

In physics, two or more different quantum states are said to be degenerate if they are all at the same energy level. Statistically this means that they are all equally probable of being filled, and in...

# What is Entity Linking?

## Iranian POW negotiator holds talks with Iraqi ministers

The head of [Iran's prisoner of war](#) commission met with two [Iraqi](#) Cabinet ministers Saturday in a bid to glean information about thousands of Iranian POWs allegedly in Iraq, the official Iraqi News Agency reported.

Iraqi Foreign Minister [Mohammed Saeed al-Sahhaf](#) told Abdullah al-Najafi that the two states needed to "speed up the closure of what remains from the POW and Missing-In-Action file," INA said.

The issue of POWs and missing persons remains a stumbling block to normalizing relations between the two neighbors.

Iraq has long maintained that it has released all Iranian prisoners captured in the 1980-88 [Iran-Iraq War](#). The countries accuse each other of hiding POWs and preventing visits by the [International Committee of the Red Cross](#) to prisoner camps.

The ICRC representative in [Baghdad](#), Manuel Bessler, told The [Associated Press](#) that his organization has had difficulty visiting POWs on both sides on a regular basis.

In April, Iran released 5,584 since [1990](#).

More than 1 million people w

### Baghdad

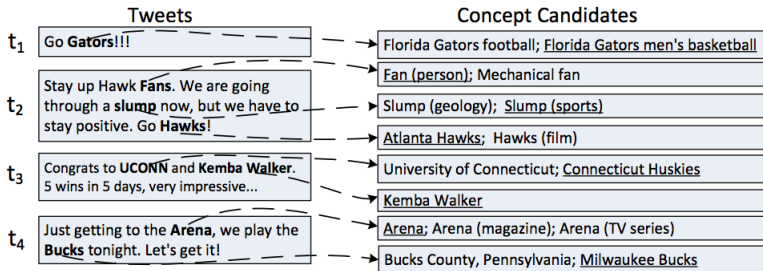
Baghdad is the capital of Iraq and of Baghdad Governorate. With a metropolitan area estimated at a population of 7,000,000, it is the largest city in Iraq. It is the second-largest city in the Arab world (after Cairo) and the second-largest city in southwest Asia (after Tehran).

[open in wikipedia](#)

ified as civil law detainees in the largest exchange



# What is Entity Linking?



# Entity Linking: Common Steps

*Determine “linkable” phrases*

mention detection - **MD**

# Entity Linking: Common Steps

*Determine “linkable” phrases*

mention detection - **MD**

*Rank/Select candidate entity links*

link generation - **LG**

# Entity Linking: Common Steps

*Determine “linkable” phrases*

mention detection - **MD**

*Rank/Select candidate entity links*

link generation - **LG**

*Use “context” to disambiguate/filter/improve*

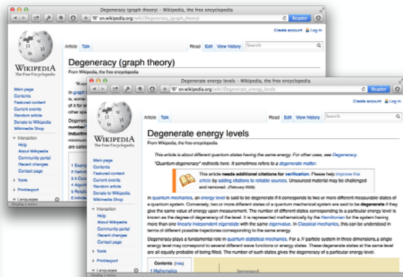
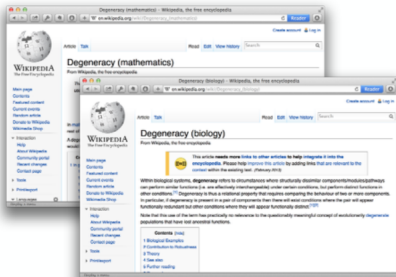
disambiguation - **DA**

# Mention Detection (MD)

Q ... degeneracy is removed ...

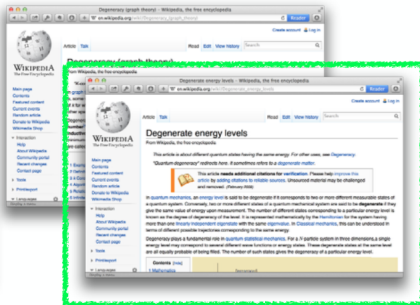
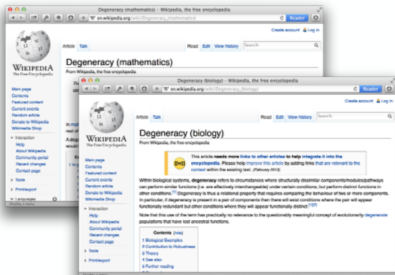
# Link Generation (LG)

Q ... degeneracy ...



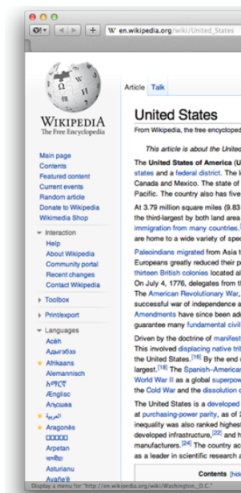
# Disambiguation (DA)

Q ... degeneracy ...



# Preliminaries: Wikipedia

- Basic element: article (proper)
- But also
  - redirect pages
  - disambiguation pages
  - category/template pages
  - admin pages
- Hyperlinks
  - use “unique identifiers” (URLs)
    - [[United States]] or [[United States|American]]
    - [[United States (TV series)]] or [[United States (TV series)|TV show]]





# Preliminaries: Disambiguation Pages

- Senses of an ambiguous phrase
- Short description
- (Possible) categorization
- Non-exhaustive



## WordNet

- 80k entity definitions
- 142k senses (entity - surface forms)

## WordNet

- 80k entity definitions
- 142k senses (entity - surface forms)

## Wikipedia

- 4M entity definitions
- 24M senses

*What can be a good measure for MD?*

*What can be a good measure for MD?*

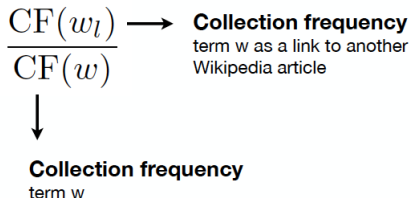
*keyphraseness( $w$ )*

Number of Wikipedia articles that use it as an anchor, divided by the number of articles that mention it at all.

*What can be a good measure for MD?*

*keyphraseness(w)*

Number of Wikipedia articles that use it as an anchor, divided by the number of articles that mention it at all.



*What can be a good measure for DA?*

*What can be a good measure for DA?*

*commonness(w, c)*

The fraction of times, a particular sense is used as a destination in Wikipedia.

$$\frac{|L_{w,c}|}{\sum_{c'} |L_{w,c'}|}$$

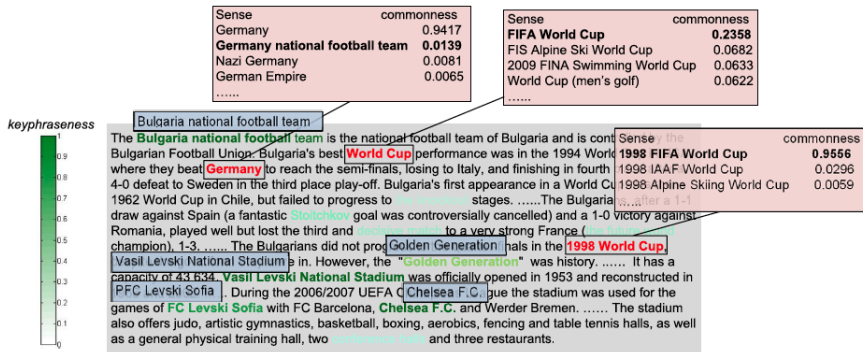


**Number of links**

with target  $c'$  and anchor text  $w$



# Commonness and keyphraseness



*Always the best decision?*

# *Always the best decision?*

- This can never help you build an accurate system, because you will always give some wrong links.
- Need to use the context.