Coreference Tasks

Natalie Parde

UIC CS 421

Coreference Tasks

- Now that we have some more linguistic background, we can formalize the task of coreference resolution:
  - **Given a text *T*, find all entities and the coreference links between them**
- This requires a few subtasks:
  - **Detect mentions**
    - Pronominal anaphoras
      - Filter out non-referential pronouns
    - Definite noun phrases
    - Indefinite noun phrases
    - Names
  - **Link those mentions into clusters**

What counts as a mention? What types of links are annotated?

- Depends on the task specifications and dataset
- Some coreference datasets do not include singletons as mentions
  - Makes the task easier
    - Singletons are often difficult to distinguish from non-referential noun phrases, and constitute a majority of mentions
- Some coreference datasets provide human-labeled mentions
  - Task is simply to cluster those mentions into groups

Sample Coreference Task

The University of Illinois at Chicago is an excellent place to study natural language processing. UIC has many faculty currently working in NLP, including but not limited to Natalie Parde, Barbara Di Eugenio, Cornelia Caragea, Bing Liu, and Philip Yu. The school is located in bustling downtown Chicago, and as a bonus it will be opening a snazzy new (non-brutalist) CS building in 2022.

Sample Coreference Task

The **University of Illinois at Chicago** is an excellent place to study **natural language processing**. **UIC** has many **faculty** currently working in **NLP**, including but not limited to **Natalie Parde**, **Barbara Di Eugenio**, **Cornelia Caragea**, **Bing Liu**, and **Philip Yu**. **The school** is located in bustling downtown **Chicago**, and as a bonus it will be opening a snazzy new (non-brutalist) **CS building** in 2022.
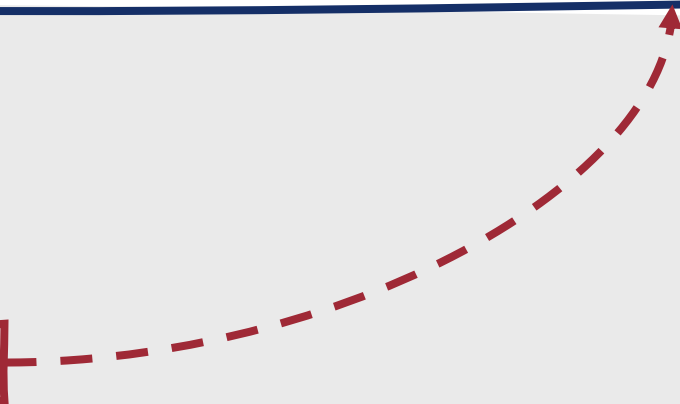
Detect mentions

Sample Coreference Task

The **University of Illinois at Chicago** is an excellent place to study **natural language processing**. **UIC** has many **faculty** currently working in **NLP**, including but not limited to **Natalie Parde**, **Barbara Di Eugenio**, **Cornelia Caragea**, **Bing Liu**, and **Philip Yu**. **The school** is located in bustling downtown **Chicago**, and as a bonus it will be opening a snazzy new (non-brutalist) **CS building** in 2022.

Detect mentions

Cluster mentions

Sample Coreference Task

The **University of Illinois at Chicago** is an excellent place to study **natural language processing**. **UIC** has many **faculty** currently working in **NLP**, including but not limited to **Natalie Parde**, **Barbara Di Eugenio**, **Cornelia Caragea**, **Bing Liu**, and **Philip Yu**. **The school** is located in bustling downtown **Chicago**, and as a bonus it will be opening a snazzy new (non-brutalist) **CS building** in 2022.

Detect mentions

Cluster mentions

Coreference Chains:
- {University of Illinois at Chicago, UIC, The school}
- {natural language processing, NLP}
- {faculty}
- {Natalie Parde}
- {Barbara Di Eugenio}
- {Cornelia Caragea}
- {Bing Liu}
- {Philip Yu}
- {Chicago}
- {CS building}