

Trabalho 1 – parte 3: Resolução de Correferência

A resolução de correferência é a tarefa de detectar elementos linguísticos que se referem a uma mesma entidade de mundo, ou seja, detectar cadeias de correferência completas, as quais são compostas por entidades nomeadas, pronomes e outros. Por exemplo:

Esse **carro** é maravilhoso. Também, estamos falando de um **automóvel** de 100 mil reais.

Cadeia de Correferência: carro, automóvel

- Sinonímia é uma relação entre dois termos, em que estes, mesmo sendo distintos lexicalmente, possuem significados muito próximos (carro - automóvel).

João e Maria estão muito felizes com o seu **cão**. O **animal** é fiel e companheiro.

Cadeia de Correferência: cão, animal

- Hiperonímia é uma relação semântica que expressa um sentido amplo entre dois termos, partindo de uma classe mais ampla para uma subclasse mais específica (animal - cão). Já Hiponímia representa uma relação contrária. Ou seja, partindo de uma classe mais específica para uma classe mais ampla (cão - animal).

Maria está com febre. Acho que **ela** está doente.

Cadeia de Correferência: Maria, ela

- Anáfora pode ser definida como a retomada de uma expressão apresentada anteriormente em um texto. A interpretação referencial do sintagma nominal “ela” depende da sua relação anafórica com o sintagma nominal “Maria”.

Sua tarefa

1. Criar regras que identifique as cadeias de correferência no conjunto de dados fornecido (Dataset – Corref-PT) em português.

Referências

<https://www.rangakrish.com/index.php/2019/02/03/coreference-resolution-using-spacy/>

<https://stanfordnlp.github.io/CoreNLP/coref.html>