

INFORMATION RETRIEVAL PROJECT

By:

Denis FESTA

Salima MAMMA

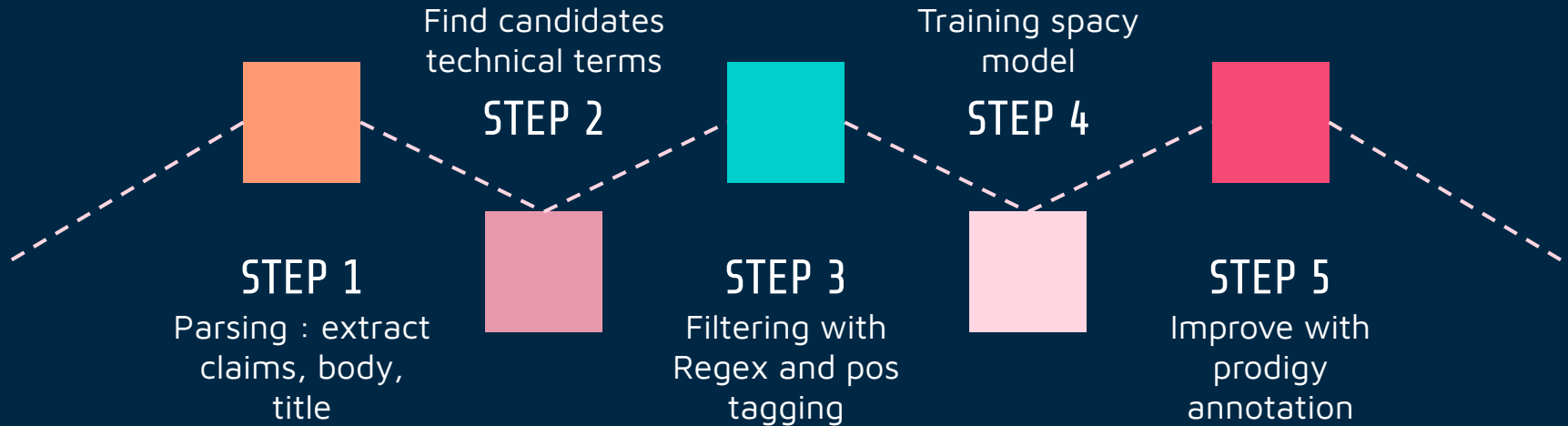
CONTENT OF THE PRESENTATION :

- Our Technical terms extraction approach
- Our relation extractions approach
- Researches and further improvements

DOMAIN DESCRIPTION :

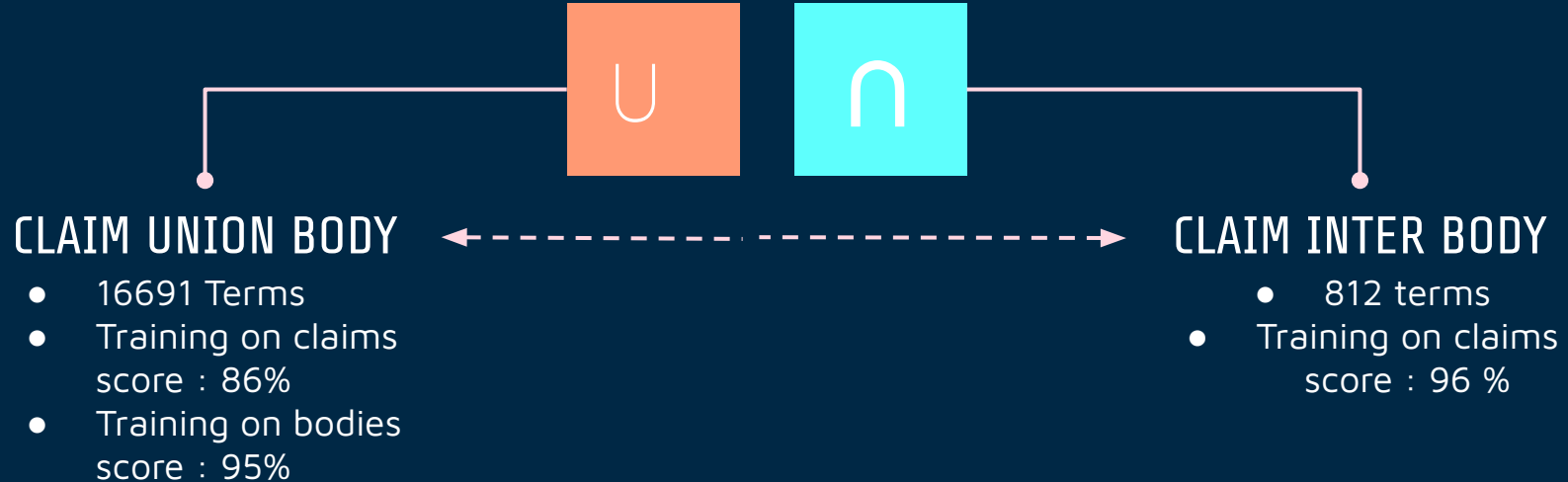
- Domain : H04N Pictorial communication
- Important topics covered : Electronic acquisition, scanning, reproduction of images, conversion of optical information into electrical signals, transmission, handling pictorial communication signals .. ect

TECHNICAL TERM EXTRACTION PROCESS :



TECHNICAL TERM EXTRACTION VOCABULARY :

DIFFERENT VOCABULARIES



PRODIGY ANNOTATIONS :

- Training : 26 sentences
- Gold Patent (Test) : 107
- Performances on the gold patent:

before Prodigy:

- 'p': 0.6682832201745877
- 'r': 1.0
- 'f': 0.8011627906976745

with Prodigy:

- 'p': 0.879948914431673
- 'r': 1.0
- 'f': 0.936141304347826

EXAMPLE OF ANNOTATED TERMS IN TEXT :

Before Prodigy

Golden Annotations:

Laser displacement meter 30 measures a displacement of the surface of the object with a laser beam casted to the object, and outputs a result of measurement as displacement data . To be more specific, laser displacement meter 30 casts the laser beam to a part of the measuring region to be photographed with infrared camera 20, thereby measuring the displacement of the part. For instance, as shown in FIG. 2, laser displacement meter 30 casts a laser beam to measuring point P on structure 102 of bridge 101, where structure 102 is located in the measuring region (= the image region shown in FIG. 2), thereby measuring the displacement of measuring point P. The displacement data includes information to be converted into stress, and is an example of data on the stress. The displacement data is not necessarily an amount of displacement (e.g. 1 mm), but it can be a general idea including the information indicating an amount of displacement (e.g. an output signal supplied from laser displacement meter 30).

Model Predictions:

Laser displacement meter 30 measures a displacement of the surface of the object with a laser beam casted to the object, and outputs a result of measurement as displacement data . To be more specific, laser displacement meter 30 casts the laser beam to a part of the measuring region to be photographed with infrared camera 20, thereby measuring the displacement of the part. For instance, as shown in FIG. 2, laser displacement meter 30 casts a laser beam to measuring point P on structure 102 of bridge 101, where structure 102 is located in the measuring region (= the image region shown in FIG. 2), thereby measuring the displacement of measuring point P. The displacement data includes information to be converted into stress, and is an example of data on the stress. The displacement data is not necessarily an amount of displacement (e.g. 1 mm), but it can be a general idea including the information indicating an amount of displacement (e.g. an output signal supplied from laser displacement meter 30).

Golden Annotations:

Laser displacement meter 30 measures a displacement of the surface of the object with a laser beam casted to the object, and outputs a result of measurement as displacement data . To be more specific, laser displacement meter 30 casts the laser beam to a part of the measuring region to be photographed with infrared camera 20, thereby measuring the displacement of the part. For instance, as shown in FIG. 2, laser displacement meter 30 casts a laser beam to measuring point P on structure 102 of bridge 101, where structure 102 is located in the measuring region (= the image region shown in FIG. 2), thereby measuring the displacement of measuring point P. The displacement data includes information to be converted into stress, and is an example of data on the stress. The displacement data is not necessarily an amount of displacement (e.g. 1 mm), but it can be a general idea including the information indicating an amount of displacement (e.g. an output signal supplied from laser displacement meter 30).

Model Predictions:

Laser displacement meter 30 measures a displacement of the surface of the object with a laser beam casted to the object, and outputs a result of measurement as displacement data . To be more specific, laser displacement meter 30 casts the laser beam to a part of the measuring region to be photographed with infrared camera 20, thereby measuring the displacement of the part. For instance, as shown in FIG. 2, laser displacement meter 30 casts a laser beam to measuring point P on structure 102 of bridge 101, where structure 102 is located in the measuring region (= the image region shown in FIG. 2), thereby measuring the displacement of measuring point P. The displacement data includes information to be converted into stress, and is an example of data on the stress. The displacement data is not necessarily an amount of displacement (e.g. 1 mm), but it can be a general idea including the information indicating an amount of displacement (e.g. an output signal supplied from laser displacement meter 30).

With Prodigy

EXAMPLE OF ANNOTATED TERMS IN TEXT :

Golden Annotations:

The **controller TT** finds first **waveform data TT** respectively on the one part of the **measuring region TT** and a part other than the one part of the **measuring region TT** based on a change with time of the **thermal data TT**, and then finds second **waveform data TT** based on a change with time of the data related to the stress. The **controller TT** also finds **disturbance data TT** through a deduction of the second **waveform data TT** from the first **waveform data TT** on the one part of the **measuring region TT**, and finds **stress data TT** indicating stress occurring in the part other than the one part of the **measuring region TT** through a deduction the **disturbance data TT** from the first **waveform data TT** on the part other than the one part of the **measuring region TT**.

Model Predictions:

The **controller TT** finds first **waveform TT** **data TT** **respectively TT** on the one part of the measuring **region TT** and a part other than the one part of the measuring **region TT** based on a change with **time TT** of the thermal **data TT**, and then finds second **waveform TT** **data TT** based on a change with **time TT** of the **data TT** related to the stress. The **controller TT** also finds disturbance **data TT** through a deduction of the second **waveform TT** **data TT** from the first **waveform TT** **data TT** on the one part of the measuring **region TT**, and finds stress **data TT** indicating stress occurring in the part other than the one part of the measuring **region TT** through a deduction the disturbance **data TT** from the first **waveform TT** **data TT** on the part other than the one part of the measuring **region TT**.

Golden Annotations:

The **controller TT** finds first **waveform data TT** respectively on the one part of the **measuring region TT** and a part other than the one part of the **measuring region TT** based on a change with time of the **thermal data TT**, and then finds second **waveform data TT** based on a change with time of the data related to the stress. The **controller TT** also finds **disturbance data TT** through a deduction of the second **waveform data TT** from the first **waveform data TT** on the one part of the **measuring region TT**, and finds **stress data TT** indicating stress occurring in the part other than the one part of the **measuring region TT** through a deduction the **disturbance data TT** from the first **waveform data TT** on the part other than the one part of the **measuring region TT**.

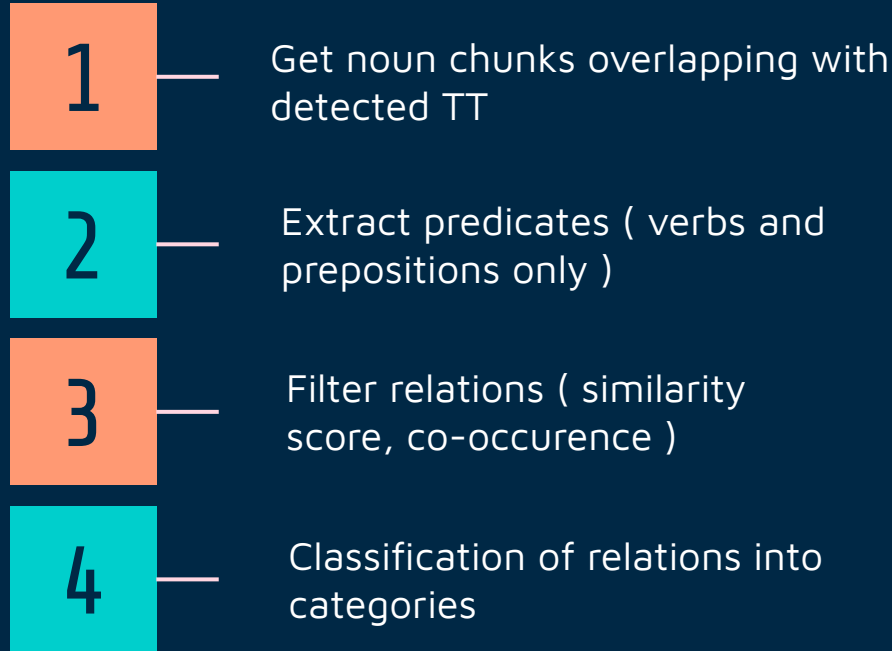
Model Predictions:

The **controller TT** finds first **waveform data TT** respectively on the one part of the **measuring region TT** and a part other than the one part of the **measuring region TT** based on a change with time of the **thermal data TT**, and then finds second **waveform data TT** based on a change with time of the data related to the stress. The **controller TT** also finds **disturbance data TT** through a deduction of the second **waveform data TT** from the first **waveform data TT** on the one part of the **measuring region TT**, and finds **stress data TT** indicating stress occurring in the part other than the one part of the **measuring region TT** through a deduction the **disturbance data TT** from the first **waveform data TT** on the part other than the one part of the **measuring region TT**.

SOME IMPROVEMENTS TO DO :

- Training on more patents with prodigy
- Handling special cases like hyphens
- Improve detection of compound terms (3 - 4 ngrams)

RELATION EXTRACTION PIPELINE :



RELATIONS EXTRACTED BEFORE FILTERING :

infrared camera 20 photographs an object at a predetermined frame rate TT, thereby producing produce thermal image data TT (video data TT) indicating a temperature of of the object TT. infrared camera TT 20 is disposed under under bridge TT 101, and photographs a predetermined region (measuring region TT) of the underside of bridge 101, thereby producing produce data TT of of a thermal image TT (thermal image data TT) indicating a temperature of the predetermined region. each pixel of the thermal image TT has have its own temperature information TT. fig. 2 schematically shows an example of the thermal image TT produced by produce infrared camera TT 20. the data of the thermal image TT is an example of of thermal data TT.

laser displacement meter 30 measures TT a displacement of the surface of the object with with a laser beam TT casted to the object, and outputs a result of of measurement TT as as displacement data TT. to be more specific, laser displacement meter TT 30 casts cast the laser beam TT to a part of of the measuring region TT to be photographed with with infrared camera TT 20, thereby measuring the displacement of the part. for instance, as shown in fig. 2, laser displacement meter TT 30 casts cast a laser beam TT to measuring to point p TT on structure 102 of bridge 101, where structure 102 is located in in the measuring region TT (= the image region shown in fig. 2), thereby measuring the displacement of measuring point TT p. the displacement data TT includes include information to be converted into stress, and is an example of of data TT on the stress. the displacement data TT is not necessarily an amount of displacement (e.g. 1 mm), but it can be a general idea including the information indicating an amount of displacement (e.g. an output signal TT supplied from supply laser displacement meter TT 30).

stress measurement device TT 10 measures a stress dis measure tribution occurring in a predetermined region TT of of bridge TT 101 based on on the thermal image data TT supplied from supply infrared camera TT 20 and the displacement data TT

RELATIONS FILTERING

EMBEDDING

Apply Bert Embedding for “source” and “target” (all-mnpet-base-v2)

Compute cosine similarity

Filter relations using a threshold

CO-OCCURRENCE

Compute the co-occurrence matrix of the technical terms

Filter the relations using a threshold

Combine it with the embedding





RELATIONS CLASSIFICATIONS :

- Done manually
- Used NEO4J to visualize and query using cypher in order to get more understanding on the semantic of relationships
- Remove irrelevant relations ("other" category)

Relationship types

*(159)

configuration

inclusion

indication

manipulation

measure

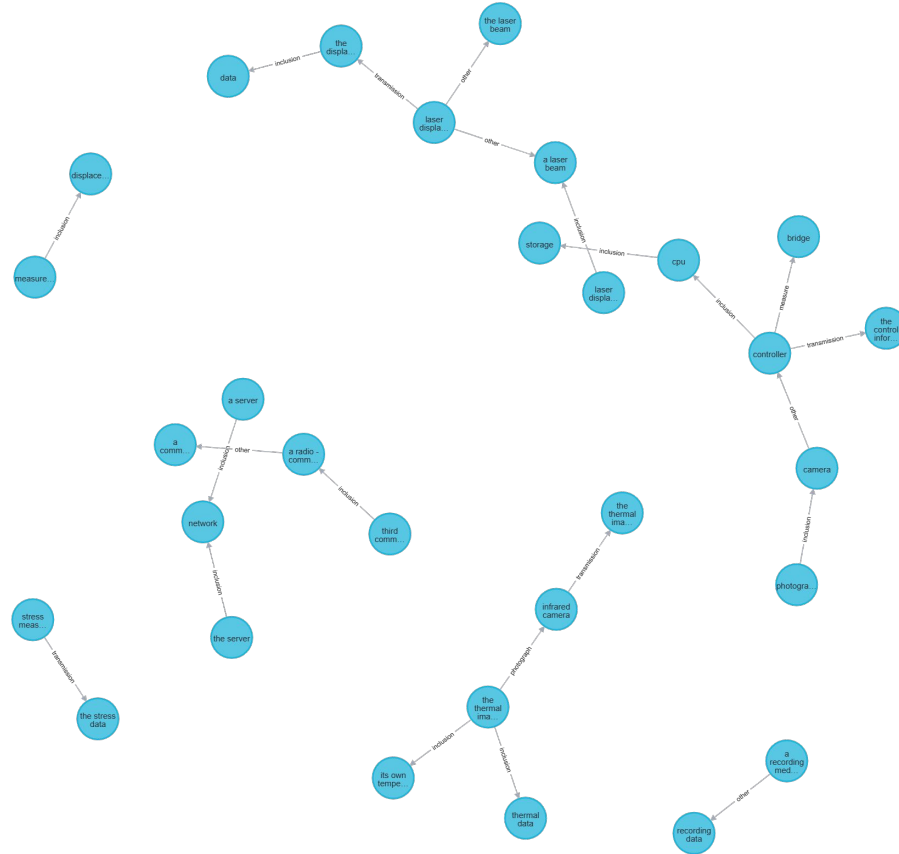
photograph

synonymy

transmission

using

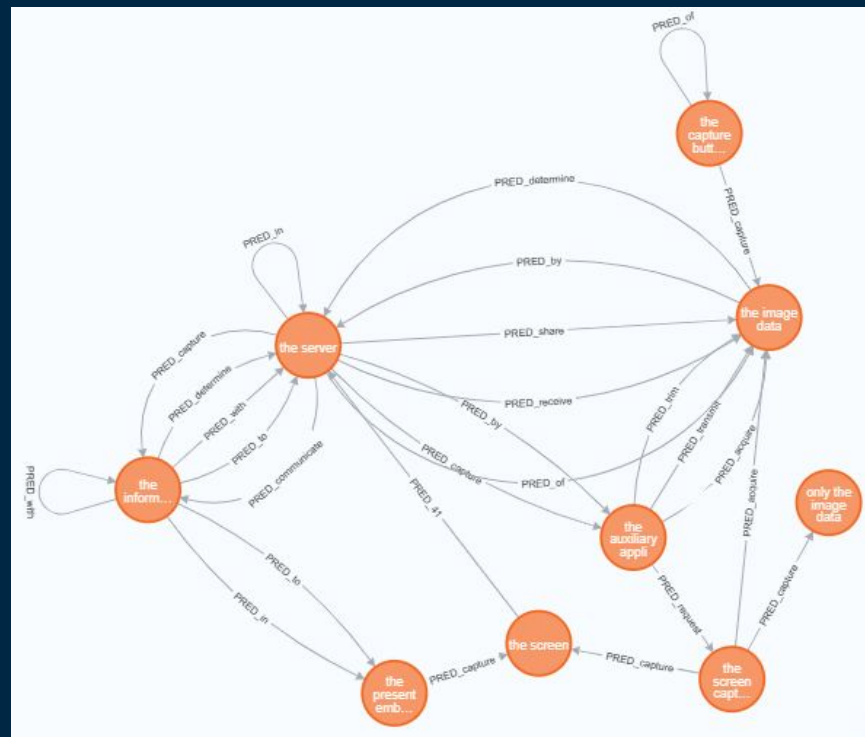
NEO4J GRAPH :



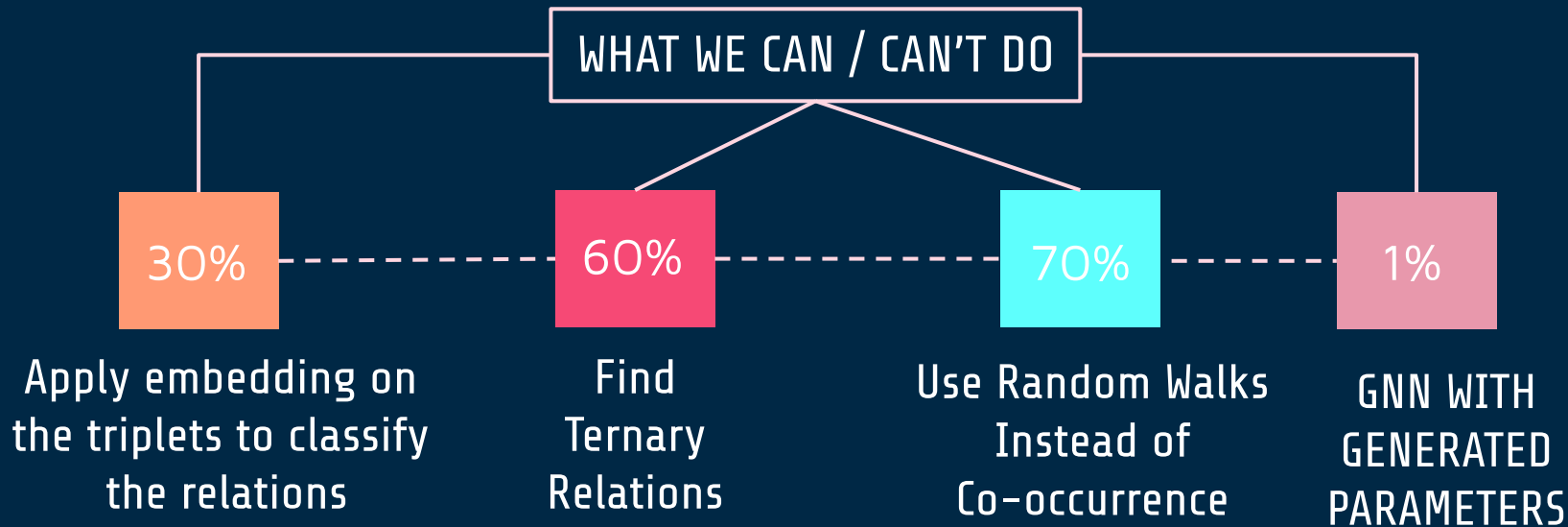
SOME ISSUES

- One predicate can have different meanings and describe different relationships.
- Lemmatization leads to the loss of direction in relations
- We may lose some important relations with the co-occurrence threshold.

We tried and didn't work :



SOME IMPROVEMENTS :



GP - GNN :

- Article Name : Graph Neural Networks with Generated Parameters for Relation Extraction

- How it works :

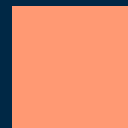
1- Encoding Module :

- > Combine positional embeddings with word embeddings (Bi-LSTM)
- > Extract representation of entity pairs (LSTM, GRU)
- > Generate edge parameters using MLP to produce a transition matrix

2- Propagation Module : Learn Node representations layer by layer by taking into account the transition matrix

3- Classification Module : Make predictions about the relationship between entities

SOME TRIES ON THE QUERYING SYSTEM



Subgraphs matching	JACCARD SIMILARITY	USE GRAPH EMBEDDING
<ul style="list-style-type: none">Not very good result, apply only to small and direct queries	<ul style="list-style-type: none">Biased by the difference of size between text and query	<ul style="list-style-type: none">TO DO

CONCLUSION

References

- Graph Neural Networks with Generated Parameters for Relation Extraction
<https://paperswithcode.com/paper/graph-neural-networks-with-generated>
- An Analysis of Relation Extraction within Sentences from Wet Lab Protocols
https://cse.cs.ucsb.edu/sites/default/files/publications/an_analysis_of_relation_extraction_within_sentences_from_wet_lab_protocols.pdf
- Word2Vec and Semantic Similarity using spacy | NLP spacy Series | Part 7 – Data Science Duniya (ashutoshtripathi.com)