ARTICLES?

Marco Ciccalè Baztán
Jorge Martín Izquierdo
Mª Gloria Cumia Espinosa de los Monteros
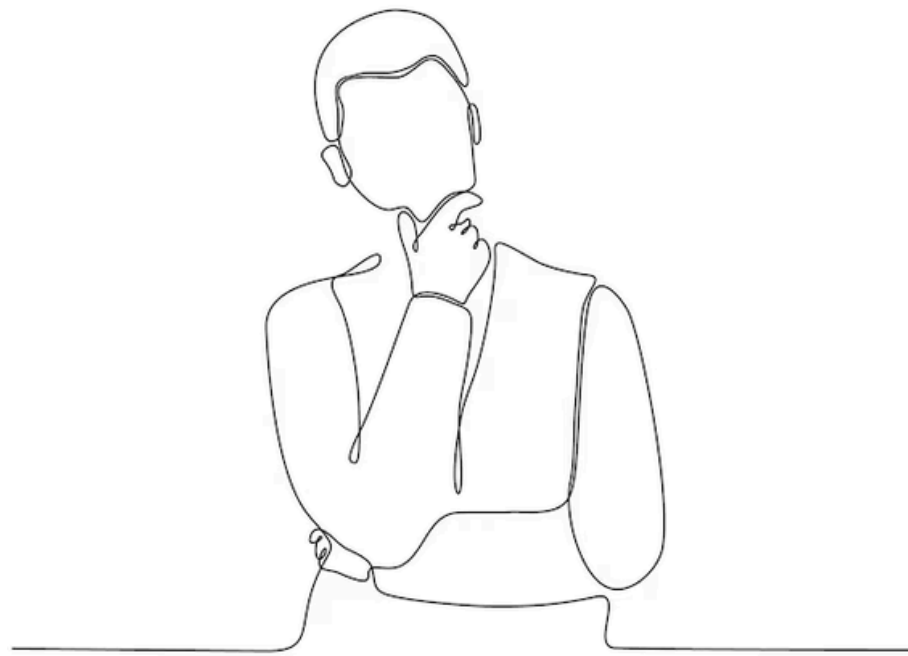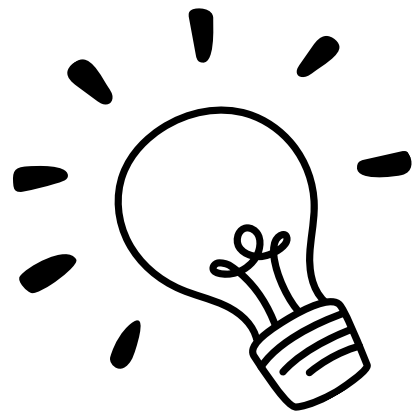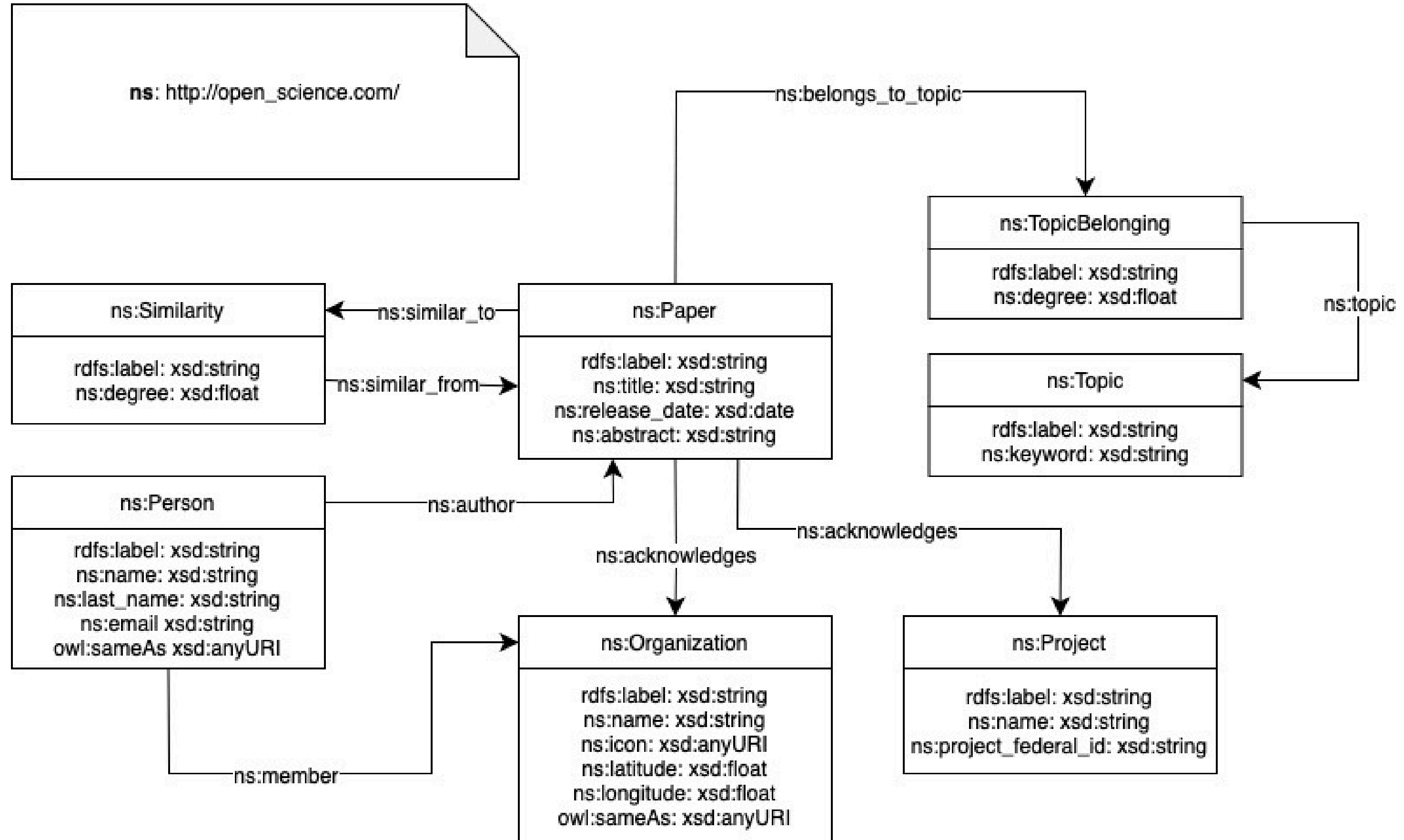
ARTICLES? ➡️

ARTICLE'S GRAPH

https://github.com/JorgeMIng/Article_Graph

Marco Ciccalè Baztán
Jorge Martín Izquierdo
Mª Gloria Cumia Espinosa de los Monteros

UNIVERSIDAD POLITÉCNICA DE MADRID
POLITÉCNICA

ONTOLOGY

ns: http://open_science.com/

ns:TopicBelonging

rdfs:label: xsd:string
ns:degree: xsd:float

ns:Paper

rdfs:label: xsd:string
ns:title: xsd:string
ns:release_date: xsd:date
ns:abstract: xsd:string

ns:belongs_to_topic

ns:topic

ns:Topic

rdfs:label: xsd:string
ns:keyword: xsd:string

ns:Similarity

rdfs:label: xsd:string
ns:degree: xsd:float

ns:similar_to

ns:similar_from

ns:Person

rdfs:label: xsd:string
ns:name: xsd:string
ns:last_name: xsd:string
ns:email xsd:string
owl:sameAs xsd:anyURI

ns:author

ns:acknowledges

ns:acknowledges

ns:Organization

rdfs:label: xsd:string
ns:name: xsd:string
ns:icon: xsd:anyURI
ns:latitude: xsd:float
ns:longitude: xsd:float
owl:sameAs: xsd:anyURI

ns:Project

rdfs:label: xsd:string
ns:name: xsd:string
ns:project_federal_id: xsd:string

ns:member

NER

NER
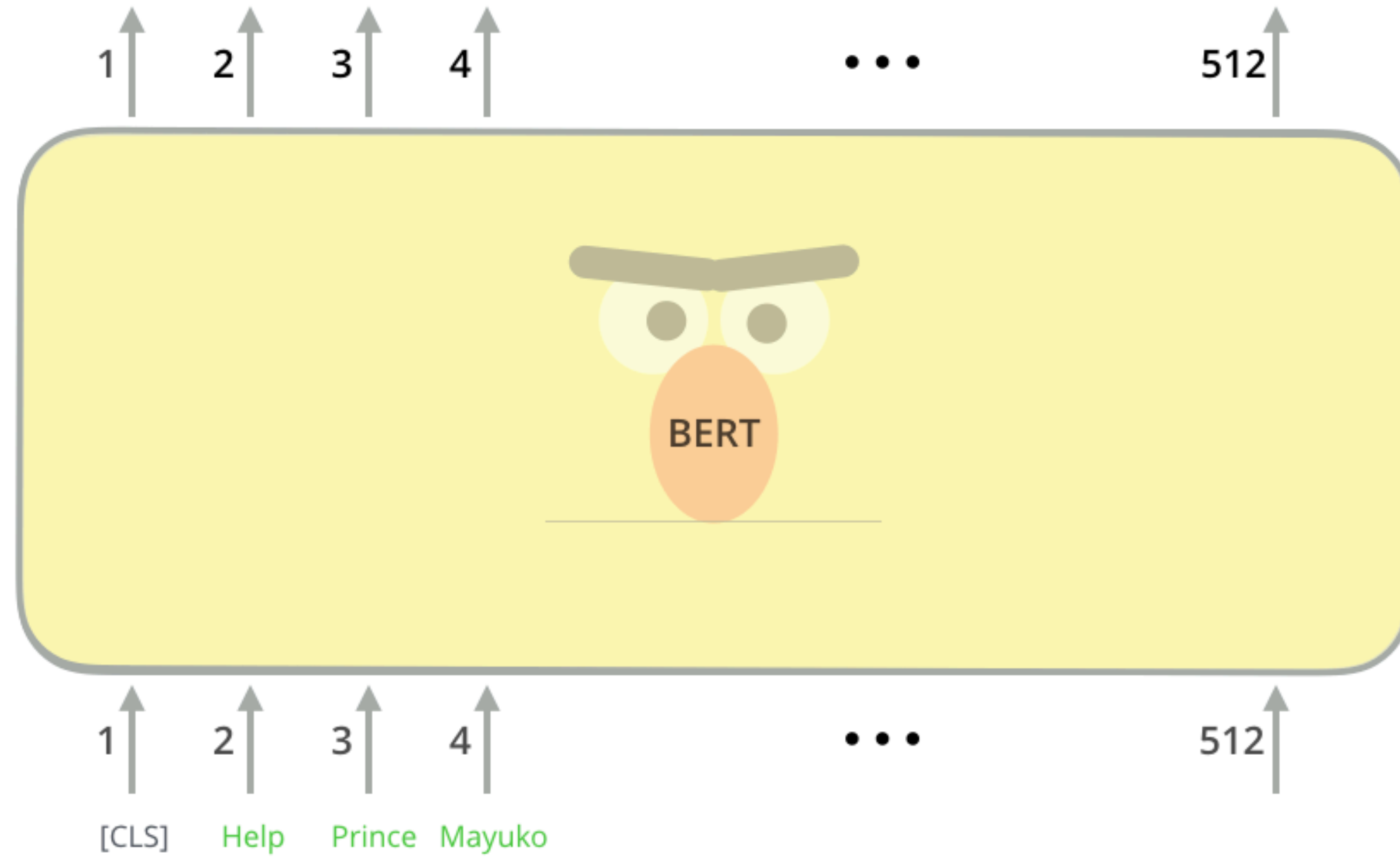
PROJECTS

```python
regex_patterns = {
    "NIH": r'(?:#)?\b[1-9][A-Z\d]{3}[A-Z]{2}\d{6}(?:-[AS]?\d+)?\b',
    "DOD": r'(?:#)?\b[A-Z\d]{6}-\d{2}-[123]-\d{4}\b',
    "NASA": r'(?:#)?\b(?:80|NN)[A-Z]+\d{2}[A-Z\d]+\b',
    "Education": r'(?:#)?\b[A-Z]+\d+[A-Z]\d{2}[A-Z\d]+\b',
    "Universal":r'[A-Z]{3,}[0-9]+-[0-9]+'
}
```
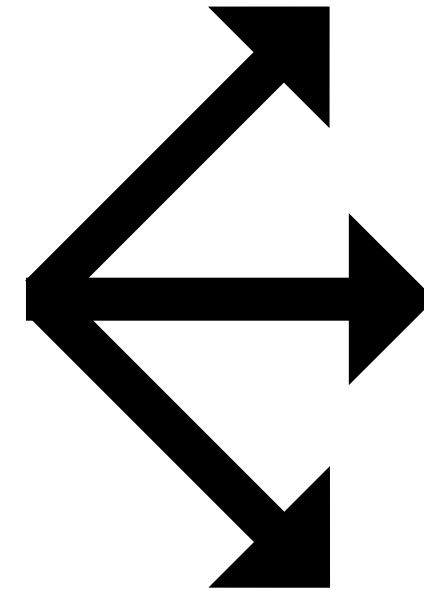
```python
match = re.search(rf'(\w+)\s+(?:grant|grants|award|awards|code|codes)\s+'+
                  '{re.escape(identifier)}', text, re.IGNORECASE)
```

TOPIC MODELING

TOPIC 0: 0.9987

TOPIC 1: 0.0080

TOPIC 0

TOPIC 1

TOPIC 0: 0.0058

TOPIC 1: 0.9765

LATENT DIRICHLET ALLOCATION (LDA)
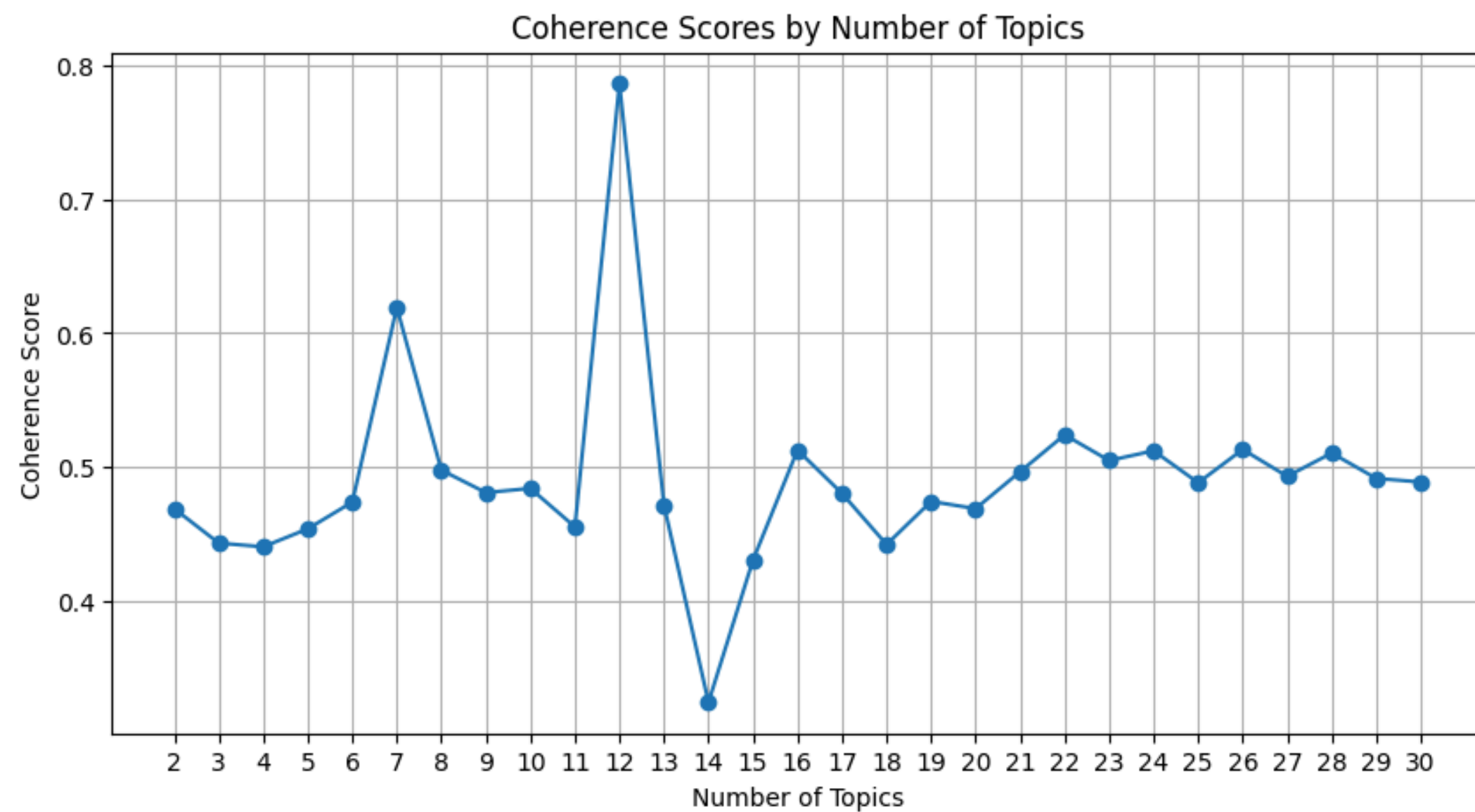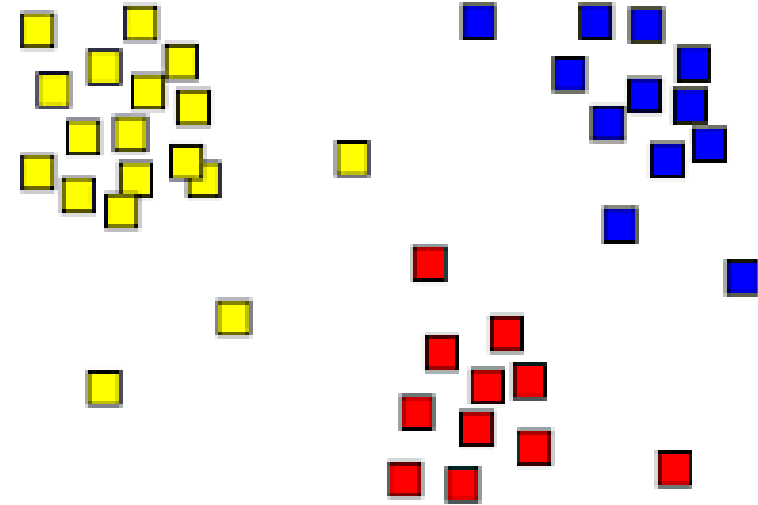
TOPIC 0: 0.9659

TOPIC 1: 0.0010

TOPIC MODELING

> Topic 0: ['investigación', 'resultados', 'poder', 'producción', 'doctoral', 'autor', 'presenta']
> Topic 1: ['investigación', 'resultados', 'poder', 'producción', 'doctoral', 'autor', 'presenta']
> Topic 2: ['investigación', 'resultados', 'poder', 'producción', 'doctoral', 'autor', 'presenta']
> Topic 3: ['escape', 'rates', 'nonmonotonic', 'increases', 'function', 'size', 'atmospheric']
> Topic 4: ['investigación', 'resultados', 'poder', 'producción', 'doctoral', 'autor', 'presenta']
> Topic 5: ['the', 'of', 'and', 'water', 'in', 'to', 'de']
> Topic 6: ['the', 'of', 'and', 'to', 'we', 'in', 'is']
> Topic 7: ['investigación', 'resultados', 'poder', 'producción', 'doctoral', 'autor', 'presenta']
> Topic 8: ['and', 'the', 'of', 'to', 'in', 'this', 'tourism']
> Topic 9: ['investigación', 'resultados', 'poder', 'producción', 'doctoral', 'autor', 'presenta']
> Topic 10: ['investigación', 'resultados', 'poder', 'producción', 'doctoral', 'autor', 'presenta']
> Topic 11: ['and', 'the', 'of', 'in', 'we', 'on', 'for']

SIMILARITY

ALL-MINILM-L6-V2   9 s

ALL-MPNET-BASE-V2   43 s

PARAPHRASE-MULTILINGUAL-MINILM-L12-V2   15 s
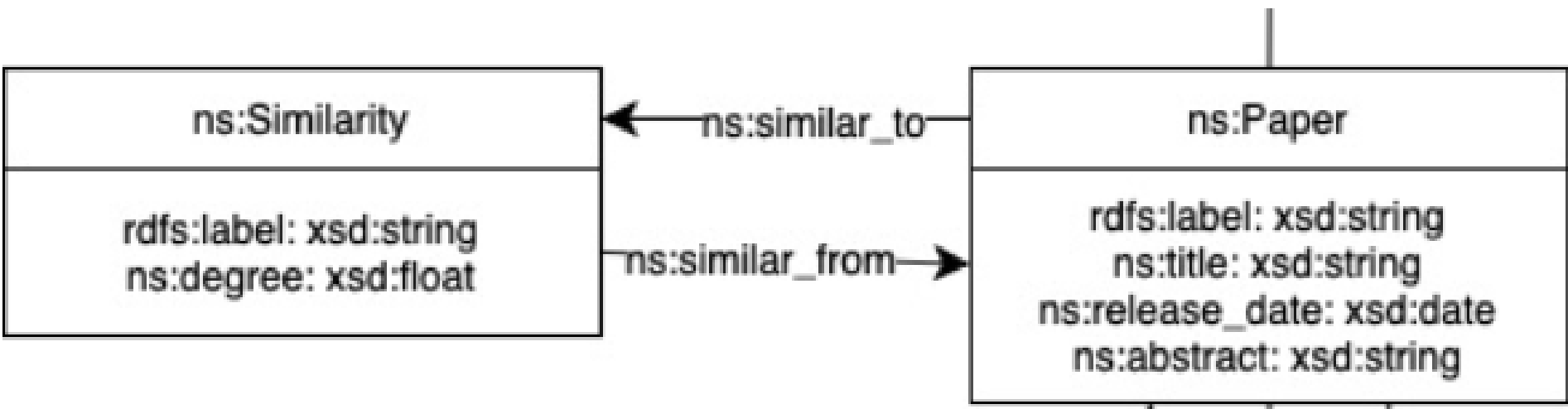
PARAPHRASE-MULTILINGUAL-MPNET-BASE-V2   1 min 11 s

LABSE   18 s
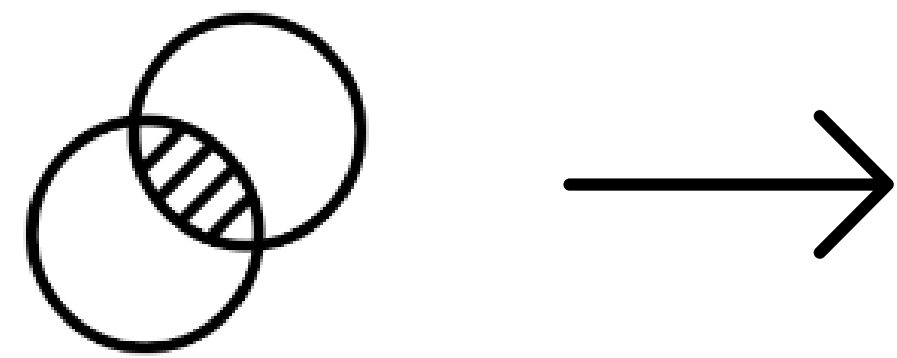
DISTILUSE   49 s

$$\text{similaridad coseno} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

SIMILARITY

ALL-MINILM-L6-V2


```
Similarity results:
{'text_id1': 0, 'text_id2': 1, 'similarity': 0.71557117}
{'text_id1': 0, 'text_id2': 2, 'similarity': 0.6356488}
{'text_id1': 0, 'text_id2': 3, 'similarity': 0.7295661}
{'text_id1': 0, 'text_id2': 4, 'similarity': 0.38767302}
{'text_id1': 0, 'text_id2': 5, 'similarity': 0.6962004}
{'text_id1': 0, 'text_id2': 6, 'similarity': 0.6496341}
{'text_id1': 0, 'text_id2': 7, 'similarity': 0.4257102}
{'text_id1': 0, 'text_id2': 8, 'similarity': 0.5233375}
{'text_id1': 0, 'text_id2': 9, 'similarity': 0.66041434}
```

WIKIDATA RECONCILIATION

ARTICLE_GRAPH
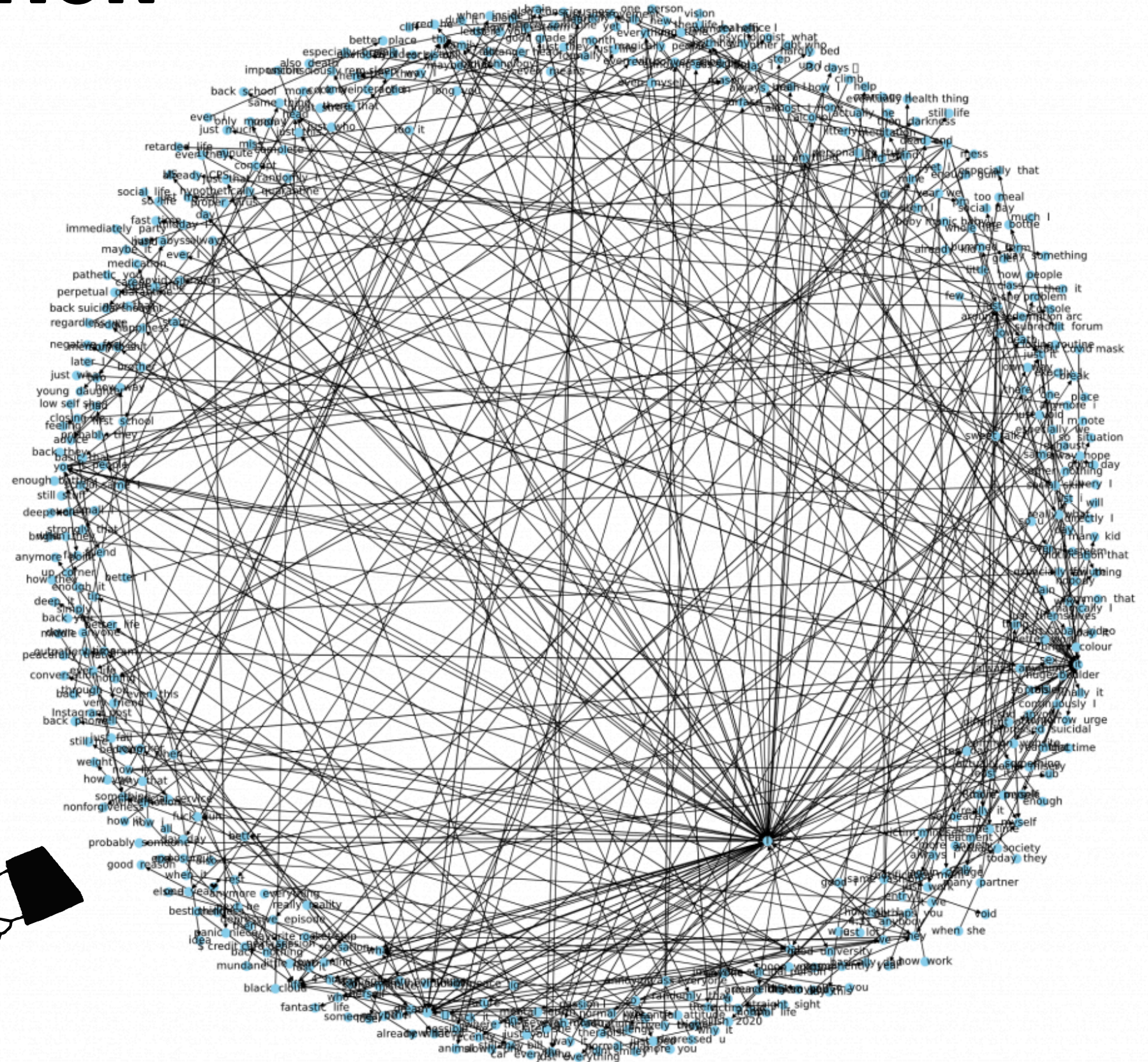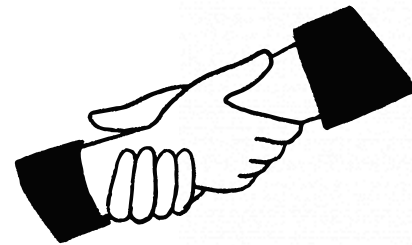
WIKIDATA

DEMOSTRACIÓN DEL FUNCIONAMIENTO

https://zenodo.org/records/11243656