

# **#Drugs4Covid**

**EUvsVirus**

**April 2020**

# Inspiration

The provision of medicines to COVID-19 patients has become a major challenge for practitioners and hospital pharmacists due to the increased demand in recent days.

In the absence of sufficient medication, disused drugs are employed or the doses of those available are modified.

However, identifying the **drug that can be used as a replacement** for others is not a simple task. In some cases the same active substances are present in medicines under different brand names.

# Data ATC Catalog

<https://library.linkeddata.es/solr/#/atc/core-overview>

Request-Handler (qt)

/select

— common —

q

\*.\*

fq

sort

start, rows


fl

df

Raw Query Parameters

key1=val1&key2=val2

wt



☐ indent off

☐ debugQuery

☐ dismax

☐ edismax

☐ hl

☐ facet

☐ spatial

☐ spellcheck

Execute Query

 [https://library.linkeddata.es/solr/atc/select?q=%3A\\*](https://library.linkeddata.es/solr/atc/select?q=%3A*)

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 2,
    "params": {
      "q": "*.*",
      "_": "1587690763419"
    }
  },
  "response": {
    "numFound": 6446,
    "start": 0,
    "docs": [
      {
        "id": "G03G",
        "label_t": "gonadotropins and other ovulation stimulants",
        "cui_s": "C3653589",
        "parent_s": "G03",
        "level_i": 3,
        "_version_": 1664580781048070144
      },
      {
        "id": "R07AB",
        "label_t": "respiratory stimulants",
        "cui_s": "C0282685",
        "parent_s": "R07A",
        "level_i": 4,
        "_version_": 1664580781048070145
      },
      {
        "id": "A10BX",
        "label_t": "other blood glucose lowering drugs, excl. insulins in atc",
        "synonyms": ["other blood glucose lowering drugs, excl. insulins"],
        "cui_s": "C3653889",
        "parent_s": "A10B",
        "level_i": 4,
        "_version_": 1664580781048070146
      },
      {
        "id": "N02AG03",
        "label_t": "pethidine and antispasmodics",
        "cui_s": "C3653049",
        "parent_s": "N02AG",
        "level_i": 5,
        "_version_": 1664580781048070147
      }
    ]
  }
}
```

# Data

## Disease Catalog

<https://library.linkeddata.es/solr/#/diseases/core-overview>

Request-Handler (qt)

/select

— common —

q

\*.\*

fq

−

+

sort

start, rows

0

10

fl

df

Do not indent results.

Raw Query Parameters

key1=val1&key2=val2

wt

-----

⬆

☐ indent off

☐ debugQuery

☐ dismax

☐ edismax

☐ hl

☐ facet

☐ spatial

☐ spellcheck

Execute Query

https://library.linkeddata.es/solr/diseases/select?q=\*\*3A\*

```
{
  "responseHeader":{
    "status":0,
    "QTime":2,
    "params":{
      "q":"**3A*",
      "_":"1587690763419"}}},
  "response":{"numFound":13000,"start":0,"docs":[
    {
      "name_t":"Deafness, X-Linked 5",
      "id":"C564472",
      "parent":["D000160",
        "D006319",
        "D010523",
        "D040181"],
      "level_i":7,
      "synonyms":["Auditory Neuropathy, X-Linked, 1, with Peripheral Sensory Neuropathy",
        "AUNX1",
        "DEAFNESS, X-LINKED 5, WITH PERIPHERAL NEUROPATHY",
        "DFNX5",
        "with Peripheral Sensory Neuropathy 1 X-Linked Auditory Neuropathy",
        "WITH PERIPHERAL NEUROPATHY X-LINKED 5 DEAFNESS"],
      "mappings":["Ear-nose-throat disease",
        "Genetic disease (inborn)",
        "Nervous system disease",
        "Signs and symptoms"],
      "_version_":1664568456760000512},
    {
      "name_t":"DEAFNESS, X-LINKED 6",
      "id":"X300914",
      "parent":["D003638"],
      "level_i":3,
      "synonyms":["DFNX6"],
      "mappings":["Ear-nose-throat disease",
        "Nervous system disease",
        "Signs and symptoms"],
      "_version_":1664568456761049088},
    {
```



# Data

## CORD-19 Articles

<https://library.linkeddata.es/solr/#/covid/core-overview>

Request-Handler (qt)

/select

— common —

q

\*\*

fq

-

+

sort

start, rows

0

10

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

-----

↑

↓

☐ indent off

☐ debugQuery

☐ dismax

☐ edismax

☐ hl

☐ facet

☐ spatial

☐ spellcheck

Execute Query

https://library.linkeddata.es/solr/covid/select?q=%3A\*

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 5,
    "params": {
      "q": "**",
      "_": "1587690763419"
    }
  },
  "response": { "numFound": 60702, "start": 0, "docs": [
    {
      "name_s": "Hypothalamic integration of immune function and metabolism",
      "id": "0189f61cdf80e0271a3327abcb1ab3a20ac110dc",
      "abstract_t": "increased body weight, decreased cytokines production and increased expression of NPY and",
      "labels4_t": "A10BC C01AA J07BG J07AH C02KB J05AC A10BG B05CX S01EB N06AB A12AA G03BA V06DE V06DC G03AB",
      "labels5_t": "V08AD01 B05XB01 G03FB01 G03FA08 B03AB07 G03FA01 G03EA01 G03CA53 L03AB09 D01AE06 V08CB03 L0",
      "source_s": "custom_license",
      "annot_cliner_problems_t": "metabolism cancer_anorexia ",
      "annot_cliner_treatments_t": "serotonin hypothalamic_dopamine central_hypothalamic_levels serotonin_prodi",
      "topics2_t": "C01CA04",
      "topics1_t": "V06DC01",
      "topics0_t": "V04CK02",
      "scispacy_diseases_t": "infections traumas tumors tumor infection obesity hyperglycemia dyslipidemias me",
      "url": [ "https://cord-19.apps.allenai.org/paper/0189f61cdf80e0271a3327abcb1ab3a20ac110dc" ],
      "url_s": "https://cord-19.apps.allenai.org/paper/0189f61cdf80e0271a3327abcb1ab3a20ac110dc",
      "txt_t":
      "_version_": 1663122736360194048
    },
    {
      "name_s": "Journal Pre-proof Small Animal Reproduction: scientific facts versus dogmas or unverified bel",
      "id": "019c40eb0b4922ed72f621af542cfla512dca3d8",
      "abstract_t": "Small Animal Reproduction is a rather recent topic in the field of Animal Reproduction. A",
      "labels4_t": "L02AA J07BX G03EA J07BG M02AA J07AH J07BC G03CC G03CB G03DA B05AA M01AX J07BB J07AL J07AG",
      "labels5_t": "J07AH10 J07AL02 G02CB03 N05CH01 B05AA03 J07AH08 N04BC06 J07AH07 J07BC02 J07AL52 J07AG53 J0",
      "source_s": "custom_license",
      "annot_cliner_problems_t": "small_animal_reproduction_(sar) ",
      "topics2_t": "D07AB30",
      "topics1_t": "G03DA04",
      "topics0_t": "G02CB03",
      "scispacy_diseases_t": "Beauvais mammary_tumours urinary_incontinence mammary_tumours mammary_tumours mai",
      "url": [ "https://cord-19.apps.allenai.org/paper/019c40eb0b4922ed72f621af542cfla512dca3d8" ],

```



# Data

## CORD-19 Paragraphs

<https://library.linkeddata.es/solr/#/covid-paragraphs/core-overview>

Request-Handler (qt)

/select

— common —

q

\*.\*

fq

sort

start, rows

0

10

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

-----

☐ indent off

☐ debugQuery

☐ dismax

☐ edismax

☐ hl

☐ facet

☐ spatial

☐ spellcheck

Execute Query

https://library.linkeddata.es/solr/covid-paragraphs/select?q=\*\*%3A\*

```
{
  "responseHeader":{
    "status":0,
    "QTime":0,
    "params":{
      "q":"**%",
      "_":"1587690763419"}},
  "response":{"numFound":2103892,"start":0,"docs":[
    {
      "section_s":"the most common antibody preparations administered orally",
      "text_t":"The main immunoglobulin class in mentioned antibody sources is IgG. A comparison of the IgG o:
      "id":"f2fe04da2f5438a7a9bcf27856e561f4",
      "article_id_s":"f8e7efae9fdae4f596b7cdb573ae5dee7eec4e70",
      "size_i":143,
      "_version_":1664508650482302977},
    {
      "section_s":"streptococcus pneumoniae",
      "text_t":"Streptococcus pneumoniae is the most common bacterial cause of communityacquired pneumonia. S
      "id":"b6c89dcdab9336f0c9a055ea654cd3a0",
      "article_id_s":"fafef140b44438528189daf3cd45a596bceb87ff",
      "size_i":451,
      "_version_":1664508650494885889},
    {
      "section_s":"enteric viruses",
      "text_t":"Although many pathogens can cause nosocomial gastroenteritis, rotavirus is responsible for 95'
      "id":"51e336fd7cc07e58c8755cfa3bcad321",
      "article_id_s":"922a678f10ad33ed9d778c2e3fb7f917bc52a6de",
      "size_i":676,
      "_version_":1664508650516905988},
    {
      "section_s":"family-centered care",
      "text_t":"As previously mentioned, a few infections do require brief separations of mother and infant. I
      "id":"56024037fb261109847b49f12e003d61",
      "article_id_s":"922a678f10ad33ed9d778c2e3fb7f917bc52a6de",
      "size_i":389,
      "_version_":1664508650563043330},
    {
      "section s":"introduction",
```



# Data

## CORD-19 Sentences

Request-Handler (qt)

/select

common

q

\*\*

fq

sort

start, rows

0

10

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

-----

☐ indent off

☐ debugQuery

☐ dismax

☐ edismax

☐ hl

☐ facet

☐ spatial

☐ spellcheck

Execute Query

https://library.linkeddata.es/solr/covid-sentences/select?q=%3A\*

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 0,
    "params": {
      "q": "**:",
      "_": "1587690763419"
    }
  },
  "response": {
    "numFound": 5274655,
    "start": 0,
    "docs": [
      {
        "id": "0adbf9d3e4a2ca61b0bbaf594355ca4358a69023-127",
        "text_t": "Influenza neuraminidase cleaves terminal sialic acid residues and destroys the receptors reco",
        "bionlp_atc3_t": "J07B J07C",
        "bionlp_atc4_t": "J07BX J07CA",
        "_version_": 1663065882764509189
      },
      {
        "id": "0adbf9d3e4a2ca61b0bbaf594355ca4358a69023-128",
        "text_t": "These drugs inhibit this action, which is essential for release of virus from infected cells,",
        "bionlp_atc3_t": "J07B J07C",
        "bionlp_atc4_t": "J07BX J07CA J07BH",
        "bionlp_atc5_t": "J07BB01 J07BH01 J07BG01 J07BC02 J07BH02 J07BA02 J07BF03 J07BA01",
        "_version_": 1663065882764509190
      },
      {
        "id": "0adbf9d3e4a2ca61b0bbaf594355ca4358a69023-141",
        "text_t": "The most common side effects of peramivir are nausea, diarrhea, and mild neutropenia.",
        "scispacy_diseases_t": "nausea diarrhea neutropenia",
        "scispacy_chemicals_t": "peramivir",
        "scispacy_diseases_complete_t": "nausea:46:52 diarrhea:54:62 neutropenia:73:84",
        "bionlp_atc5_t": "J05AH03",
        "_version_": 1663065882766606337
      },
      {
        "id": "0adbf9d3e4a2ca61b0bbaf594355ca4358a69023-171",
        "text_t": "Resistance is associated with increases in viral replication and aminotransferases.",
        "bionlp_atc3_t": "J07B J07C",
        "bionlp_atc4_t": "J07BX J07CA",
        "_version_": 1663065882772897796
      },
      {
        "id": "0adbf9d3e4a2ca61b0bbaf594355ca4358a69023-182",
        "text_t": "In comparative trials against lamivudine or adefovir, telbivudine demonstrated greater virolo"
      }
    ]
  }
}
```

<https://library.linkeddata.es/solr/#/covid-sentences/core-overview>

# Task 1

## Bio-API

- /articles?keywords=paracetamol+fever
- /paragraphs?keywords=paracetamol+fever
- /drugs?keywords=paracetamol+fever
- /diseases?keywords=paracetamol+fever

*on progress....*

*(<http://library.linkeddata.es/bio-api>)*



# Task 1

## Bio-API



### Most frequent drugs..

- ..used in the experiments: <https://library.linkeddata.es/bio-api/drugs>
- ..grouped by therapeutic group (ATC-4): <https://library.linkeddata.es/bio-api/drugs?level=4>
- ..considered together with lopinavir: <https://library.linkeddata.es/bio-api/drugs?keywords=lopinavir>
- ..used as viral vaccines: <https://library.linkeddata.es/bio-api/drugs?keywords=viral%20vaccines&level=5>
- ..used to handle fever: <https://library.linkeddata.es/bio-api/drugs?keywords=fever&level=5>

### Most frequent diseases..

- ..considered in the corpus: <https://library.linkeddata.es/bio-api/diseases>
- ..grouped by symptoms : <https://library.linkeddata.es/bio-api/diseases?level=2>
- .. treated by chloroquine: <https://library.linkeddata.es/bio-api/diseases?keywords=chloroquine>
- ..appearing along with hallucination: <https://library.linkeddata.es/bio-api/diseases?keywords=hallucination>

### Texts about...

- ... covid-19 and inflammation difficulties: <https://library.linkeddata.es/bio-api/texts?keywords=covid19,inflammation>
- ... hydroxychloroquine: <https://library.linkeddata.es/bio-api/texts?keywordshydroxychloroquine>

### Interactions with..

- ...paracetamol: <https://library.linkeddata.es/bio-api/interactions?keywords=paracetamol&size=5>

# Task 2

## Drugs Description

- [https://library.linkeddata.es/solr/atc/select?q=label\\_t%3Aapixaban](https://library.linkeddata.es/solr/atc/select?q=label_t%3Aapixaban)
- add a textual description of the drug based on the official overview

```
"response": {
  "numFound": 1,
  "start": 0,
  "docs": [
    {
      "id": "B01AF02",
      "label_t": "apixaban",
      "cui_s": "C1831808",
      "parent_s": "B01AF",
      "level_i": 5,
      "cima_codes": [
        "11691003",
        "111691014",
        "111691010",
        "11691002",
        "11691005",
        "111691009"
      ],
      "category": [
        "Human",
        "Human"
      ],
      "medicines": [
        "Eliquis",
        "Eliquis"
      ],
      "therapeutic_area": [
        "Arthroplasty, Venous Thromboembolism",
        "Arthroplasty, Venous Thromboembolism"
      ],
      "emea_codes": [
        "EMA/H/C/002148",
        "EMA/H/C/002148"
      ],
      "overviews": [
        "https://www.ema.europa.eu/en/medicines/human/EPAR/eliquis",
        "https://www.ema.europa.eu/en/medicines/human/EPAR/eliquis"
      ],
      "_version_": 1664814464300482600
    }
  ]
}
```



# Task 3

## Main concepts

- Identification of the main concepts that are representative enough of all the diseases present in the corpus

### Main problems:

- Most frequent words are not the most important ones: stopwords or words with low semantic value
- Automatic Terminology Identification:
  - Huge amount of data
  - Could not rely in POS taggers or chunkers
  - Uniword and multiword terms

```
"responseHeader": {
  "status": 0,
  "QTime": 1598
},
"terms": {
  "text_t": [
    "the",
    1889444,
    "of",
    1841684,
    "and",
    1810805,
    "in",
    1698381,
    "to",
    1587742,
    "a",
    1462115,
    "with",
    1205303,
    "for",
    1205048,
    "is",
    1024048,
    "that",
    985209
  ]
}
```

*Results obtained for the most frequent words in covid-19 texts*

[https://librairy.linkeddata.es/data/covid-paragraphs/terms?q=bionlp\\_diseases\\_N5:covid-19&terms.fl=text\\_t](https://librairy.linkeddata.es/data/covid-paragraphs/terms?q=bionlp_diseases_N5:covid-19&terms.fl=text_t)

# Task 3

## Main concepts

Solution: Rake algorithm

- Removes stopwords and punctuation marks
- Based on the degree of the words  $\rightarrow n^{\circ}$  of relations to other words that are term candidates
- Can handle big volumes of data
- Semantic analysis is not needed

**Words that comprise the obtained terms have a true semantic value and can be used in next tasks**



'cells control viral infection'  
'cells following viral infection'  
'virus protein 1'  
'subsequent viral protein 1'  
'1 upon virus infection'  
'length virus protein 2'  
'viral protein 1'  
'1 viral protein'  
'influenza virus infected cells'  
'virus infected cells compared'  
'1 virus infection'  
'cell motility protein 1'  
'tim 1 cell protein'  
'human influenza virus infection'  
...

*Sample of most important terms recognized*



# Word-Embeddings4Diseases

- Build a word embeddings from paragraphs filtered by disease:
  - bionlp\_diseases\_N5:covid-19
  - bionlp\_diseases\_N4:dengue
  - bionlp\_diseases\_N2:malaria



# Task 4

**Word-Embeddings4Diseases**



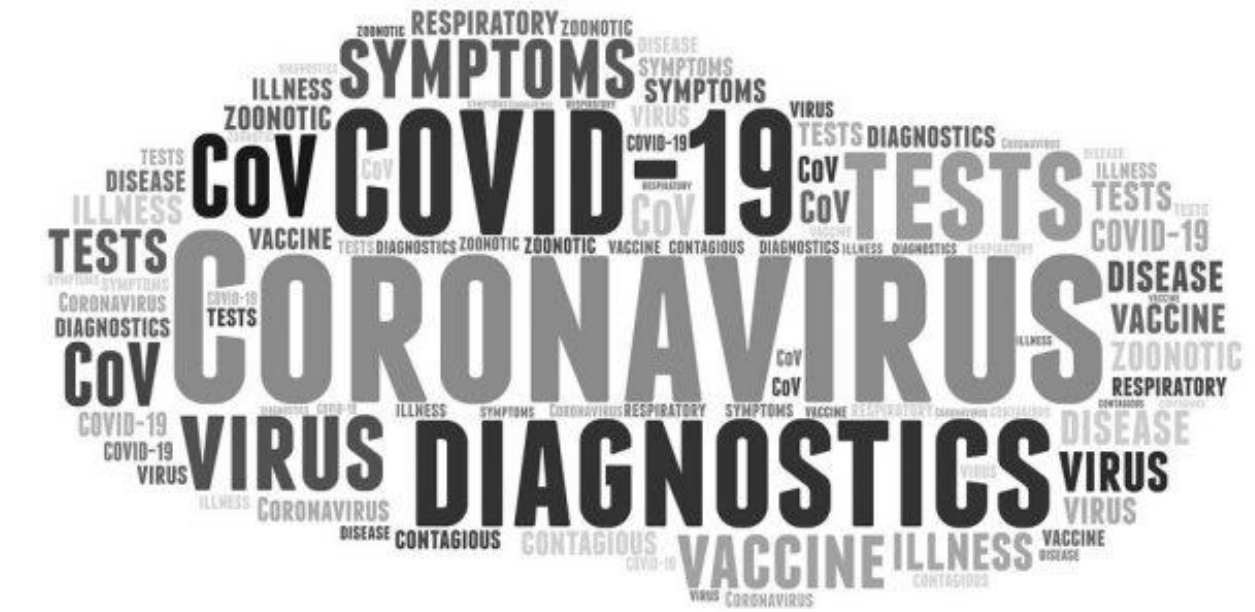


# Topics4Diseases



-

# Topics4Diseases



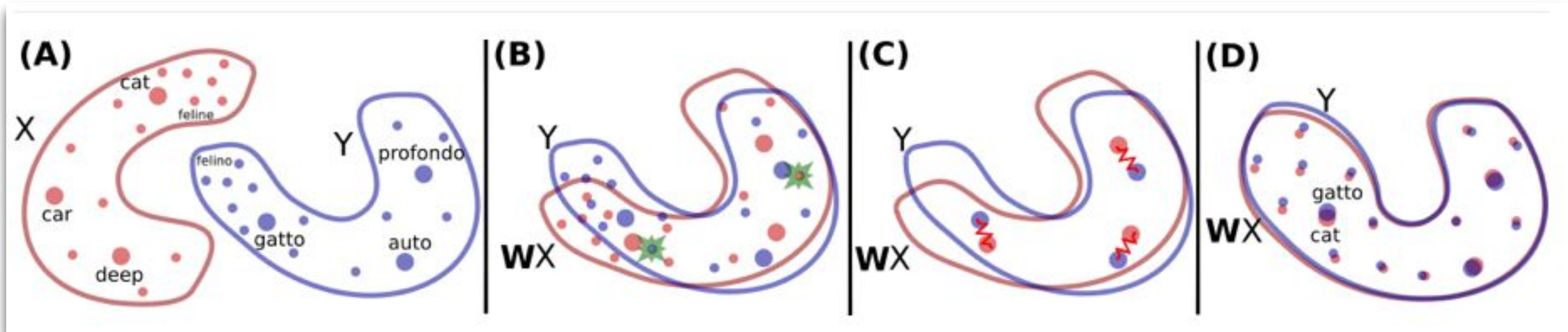
- Topic Model for **Malaria**:  
<http://library.linkeddata.es/malaria-disease-model>
- Topic Model for **Dengue**:  
<https://library.linkeddata.es/dengue-disease-model>
- Topic Model for **Covid19**:  
<https://library.linkeddata.es/covid19-disease-model>

JSON requests for **librAlry**: <https://github.com/oeg-upm/drugs4covid19-nlp/tree/master/code/json>



# Task 6

## Cord-19 Alignment



- How is a drug located in a disease-defined space?
- What drugs are nearby?
- What symptoms does it group together?

# Task 6.1

**Cord-19 WordEmbeddings-based Alignment**



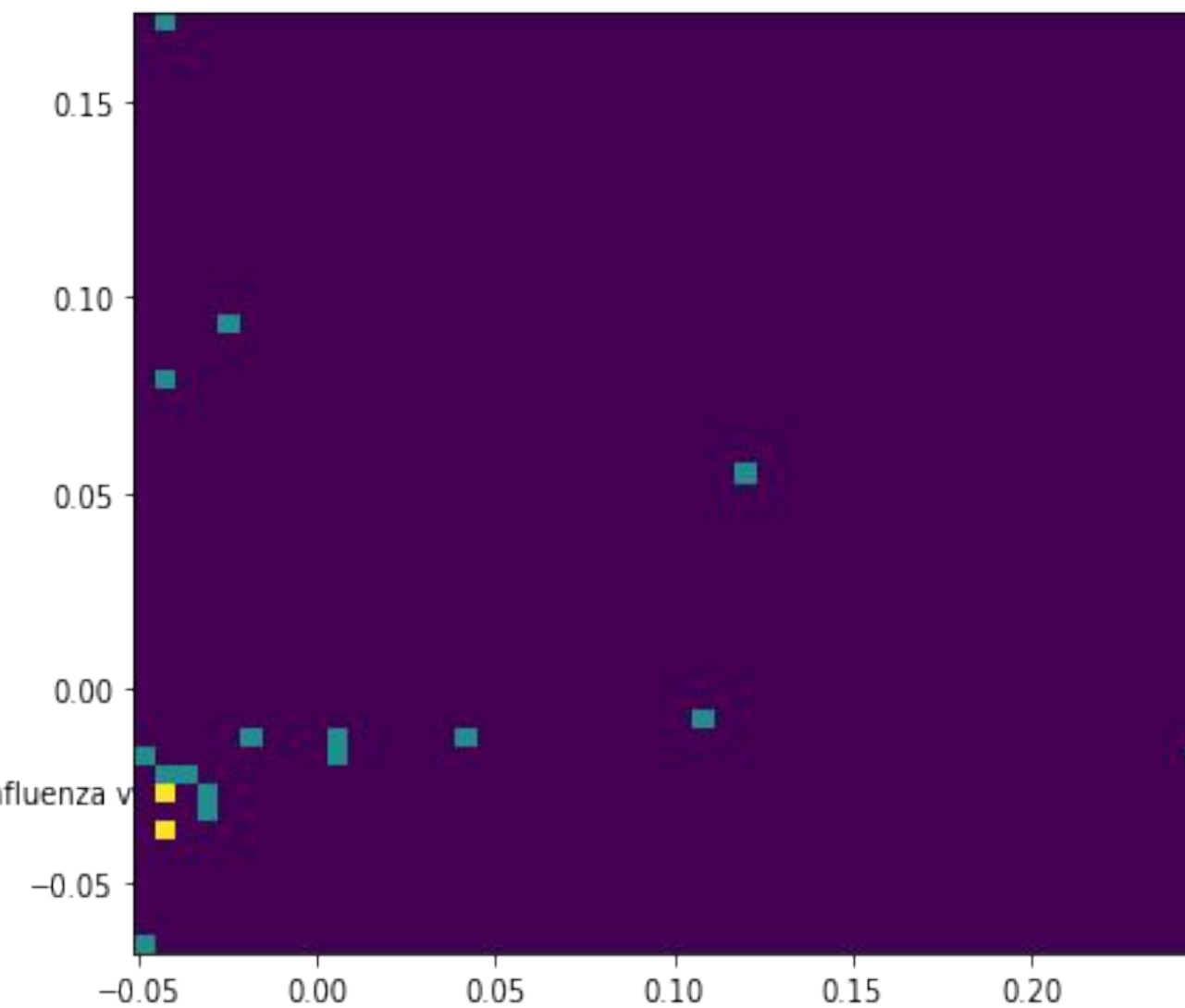
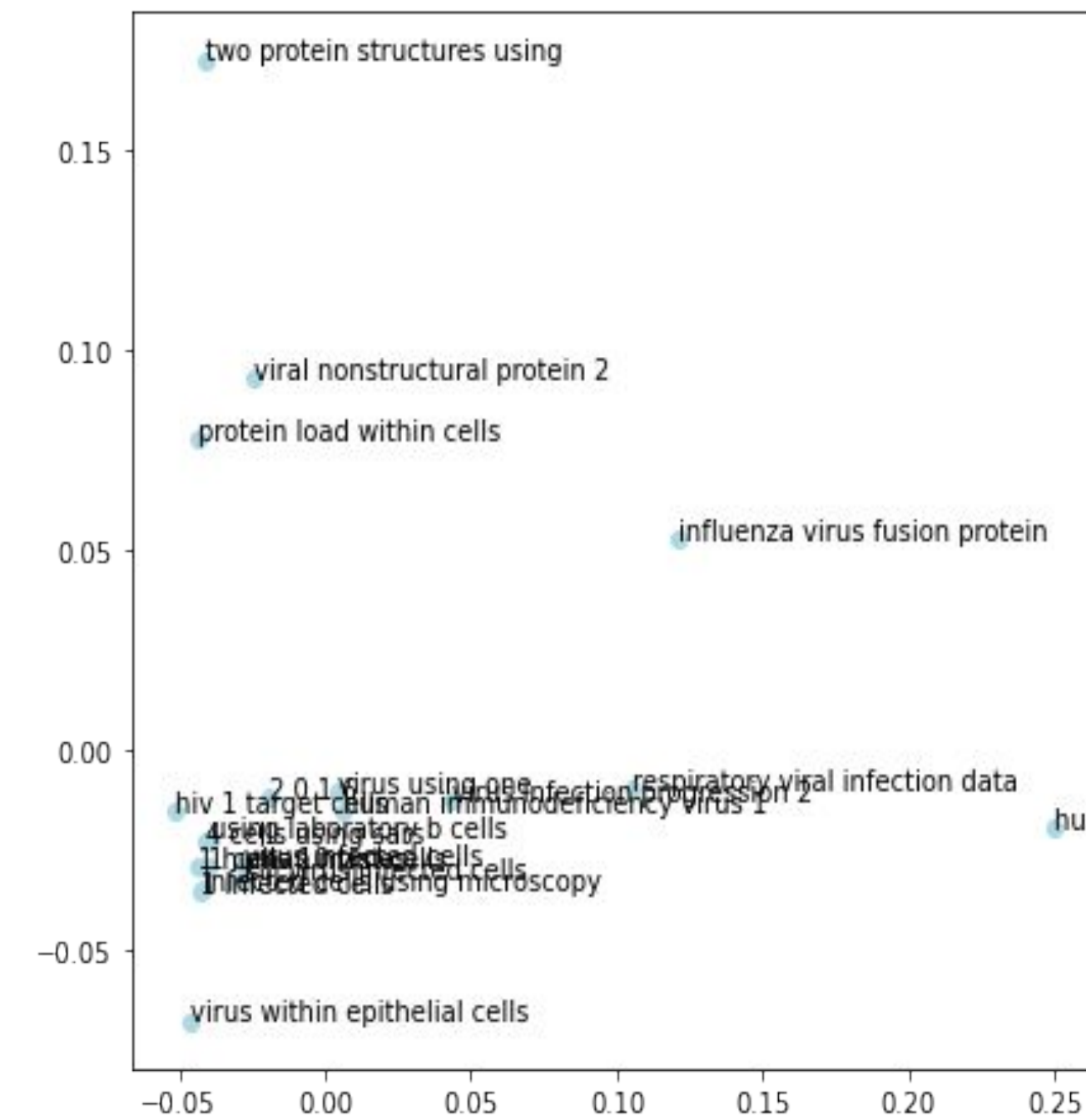


# Task 6.2

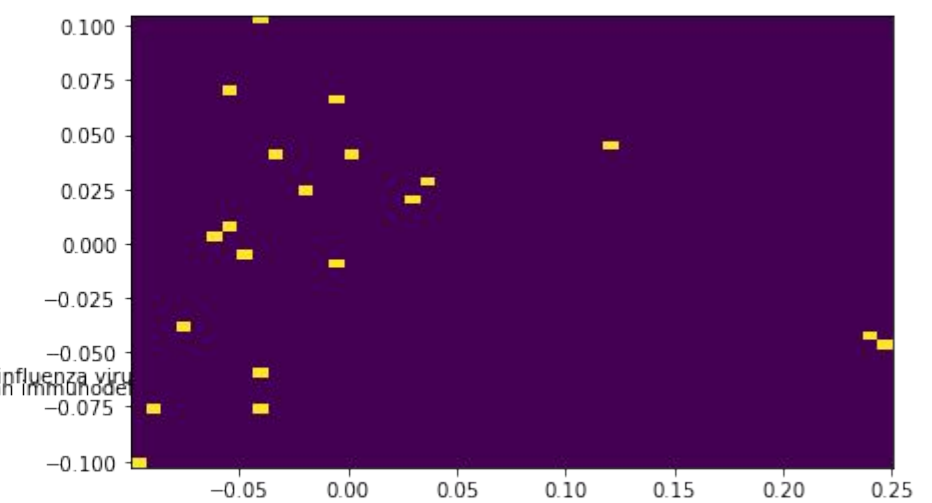
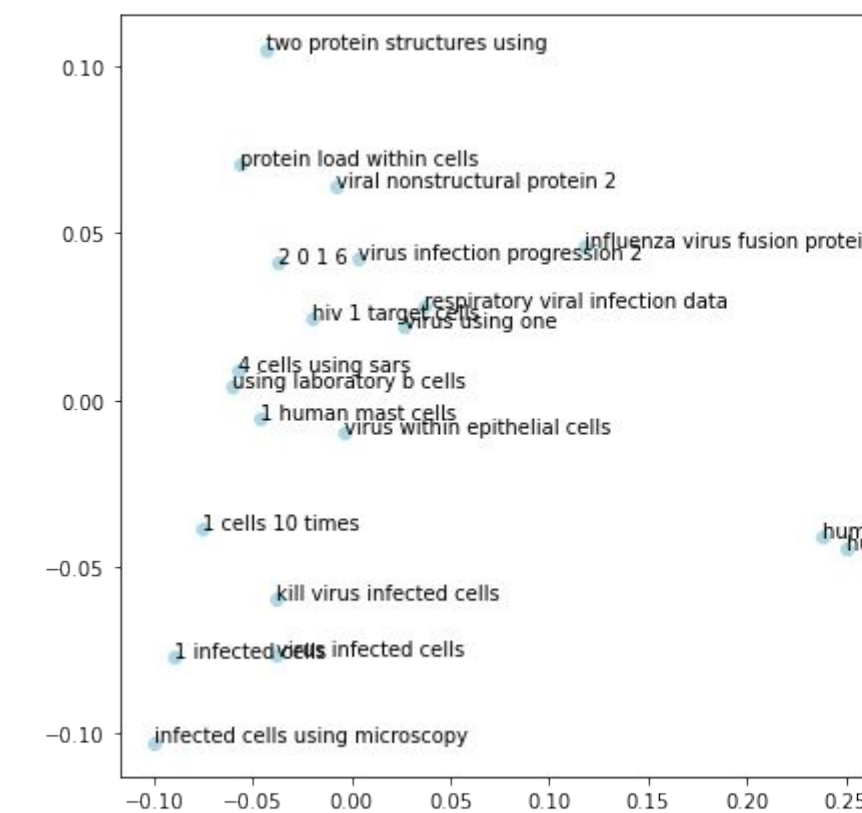
## Cord-19 Topic-based Alignment



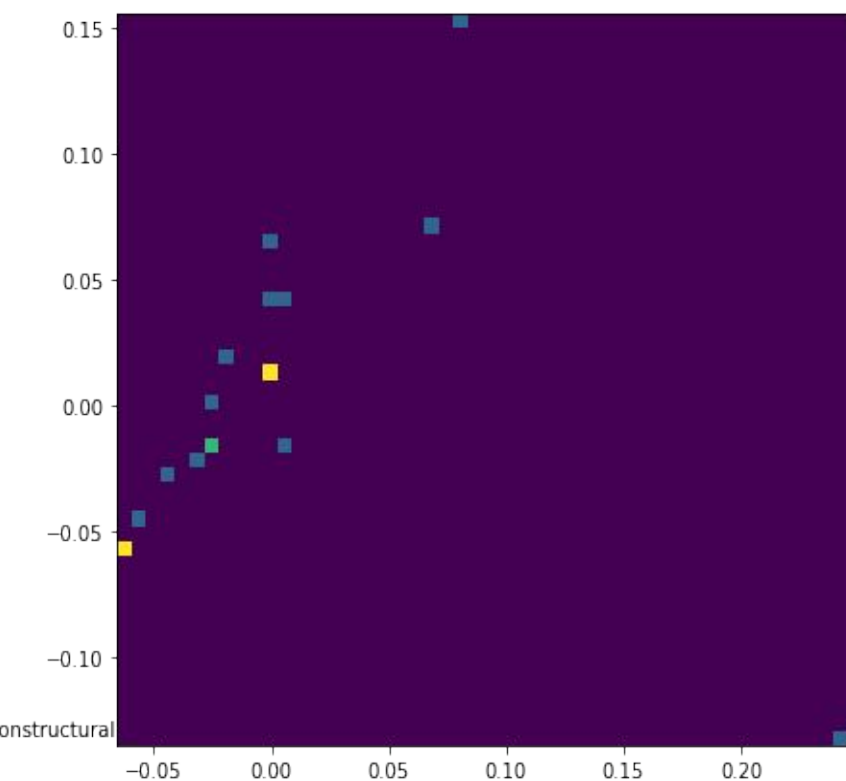
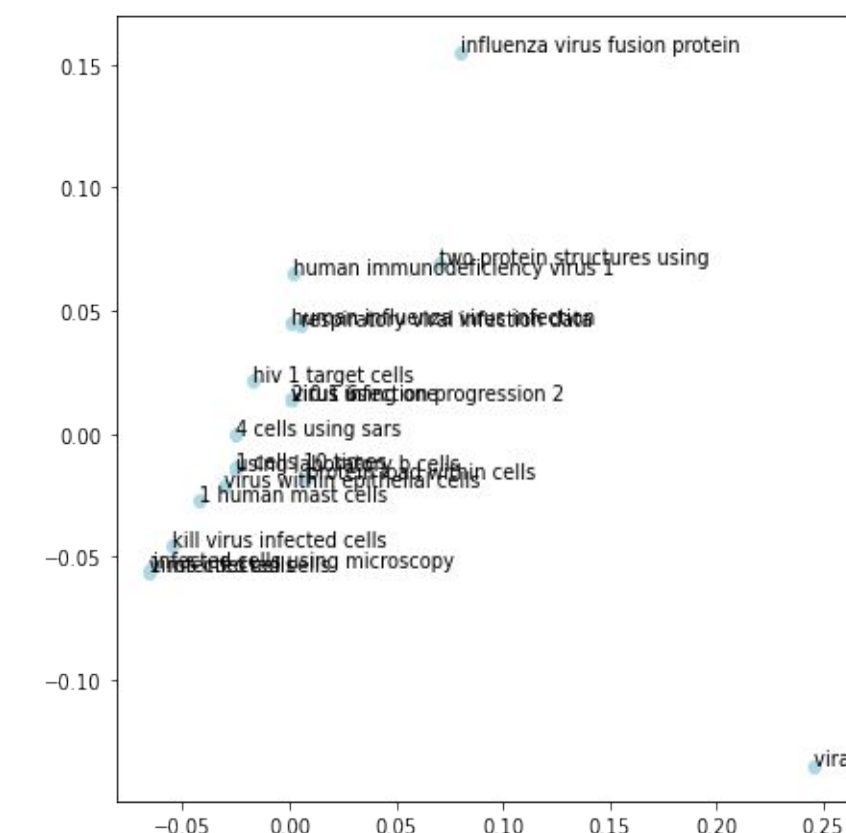
Covid19



Malaria



Dengue



Notebook:

<https://github.com/oeg-upm/drugs4covid19-nlp/tree/master/notebooks/diseases-as-topics>

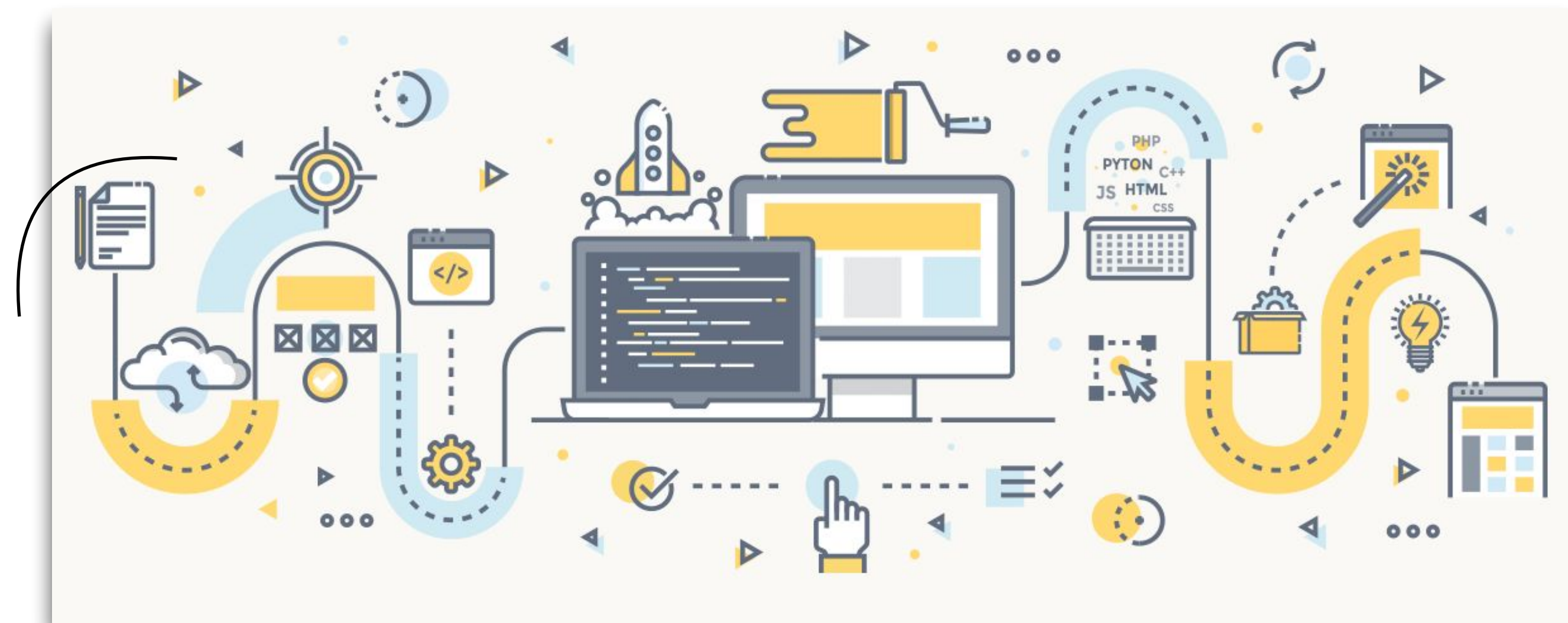


# Task 7

## KG Continuous Integration

- Ontology describing the domain:  
<https://w3id.org/def/DRUGS4COVID19>
- Script for downloading and generating rdf based on the data in solr and the ontology: YES (1 execution per day)
- RDF:<https://delicias.dia.fi.upm.es/nextcloud/index.php/s/BnWJGEArnQSoRaG>
- Mappings and scripts to generate more data/annotations:  
<https://github.com/oeg-upm/drugs4covid19-kg>
- Virtuoso with 44 millions of facts  
[kg.drugs4covid.oeg-upm.net/sparql](http://kg.drugs4covid.oeg-upm.net/sparql)

**MISSION COMPLETE!**



Website with some queries to illustrate the use of the KG:

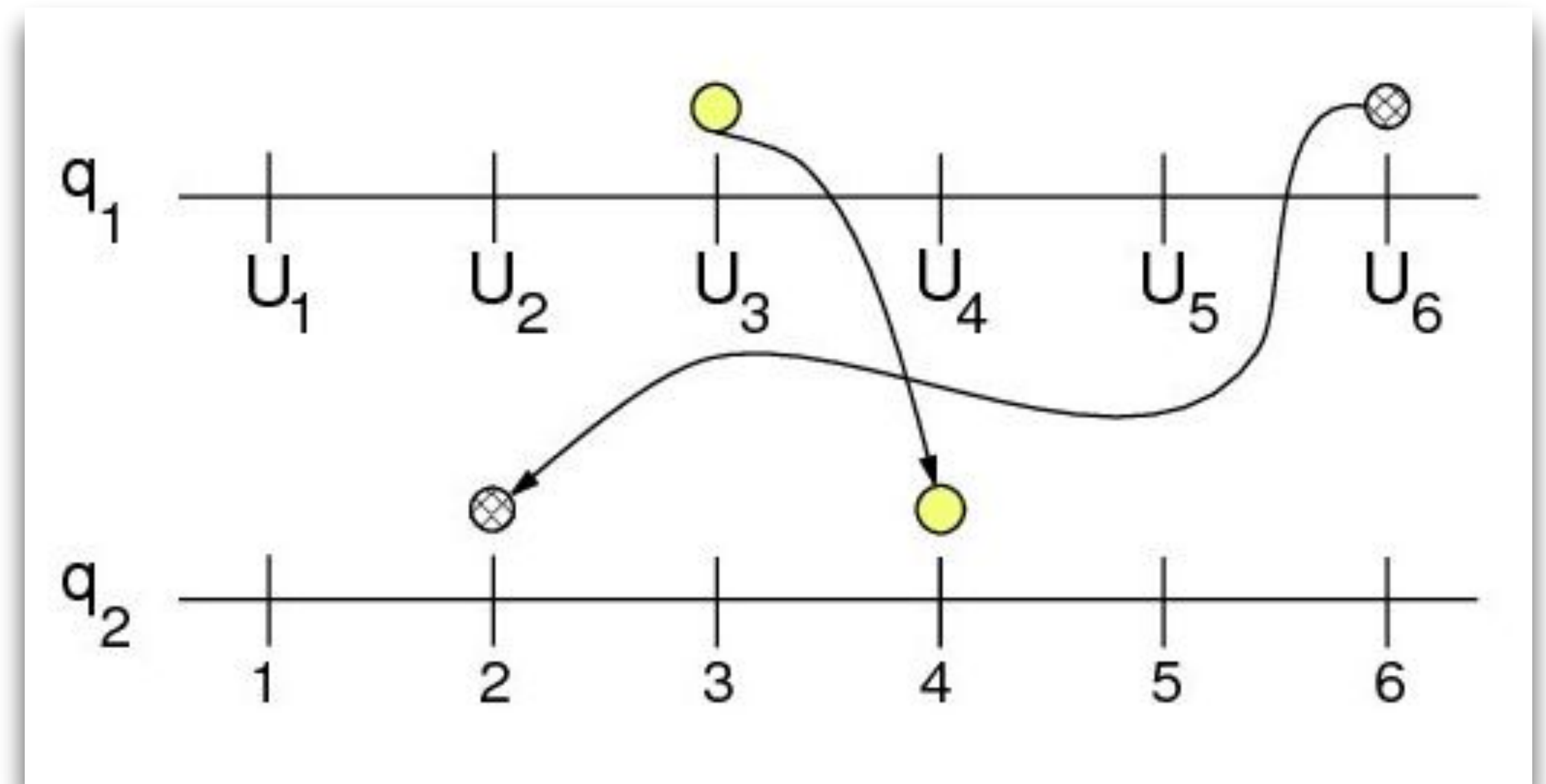
<https://bio-nlp-dashboard.netlify.app/kg>



# Task 8

## SPARQL Queries with Ranking

- Listing resources sorted by Boolean queries based on weights/scores
  - Weights assigned to each bounded attribute and a score computed. Results ordered according to this score
  - Useful for ranking answers according to how specific is the disease or the drug that is mentioned
  - See example in <https://bio-nlp-dashboard.netlify.app/kg>



# Task 9

## KG Rest API

- publish data following REST principles
- e.g. R4R, grlc.io....(using sparql queries)
- deploy on a server

### Restful-API for RDF data (r4r)

If you have data described in [RDF](#) format (e.g. a knowledge base or an ontology) and you want to publish them on the web following the REST principles via API over HTTP, this is your site!

You only need [Docker](#) installed in your system.

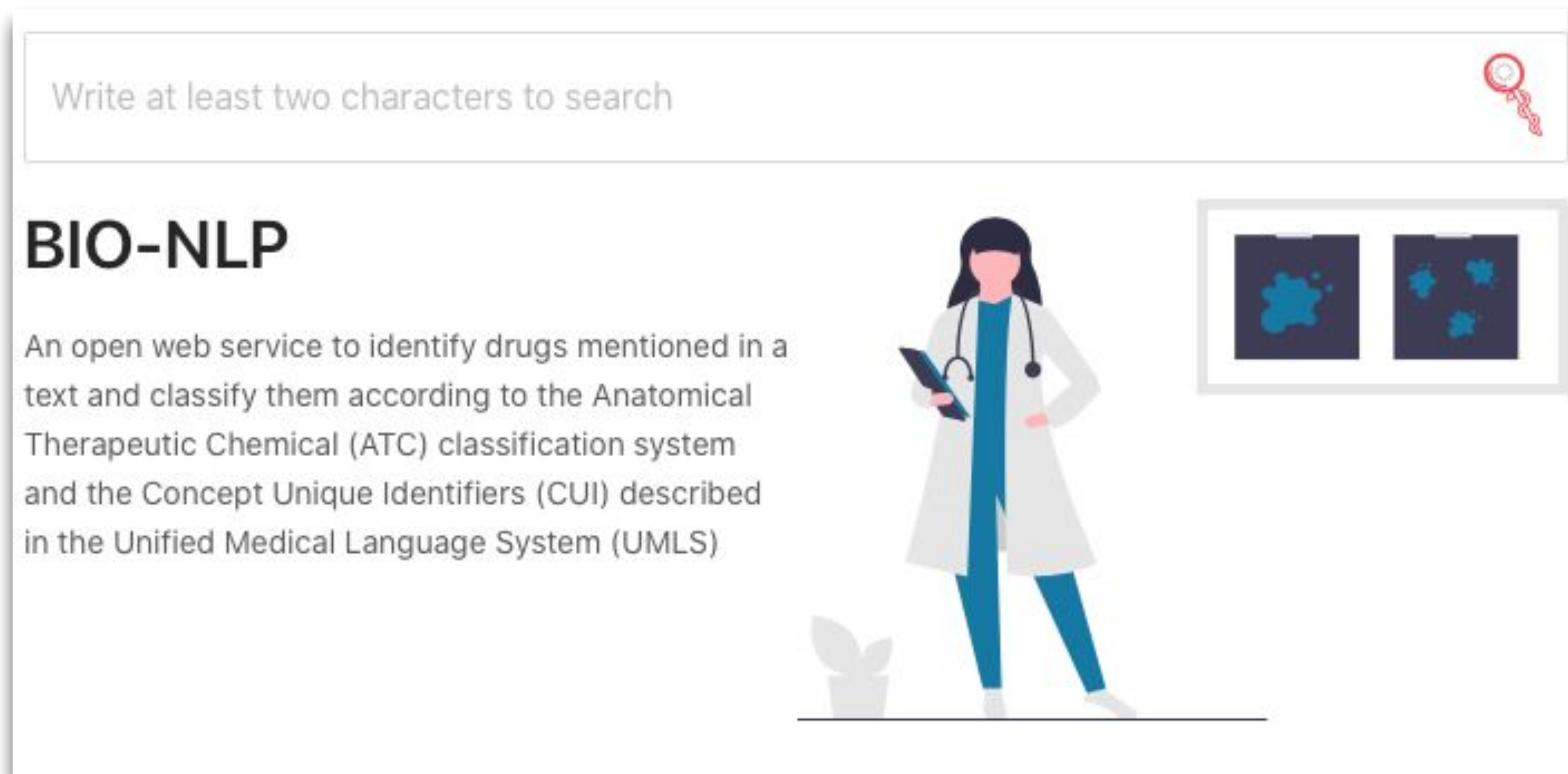
Once Docker is installed, you should be able to run the R4R container by:

```
docker run -it --rm \
  -p 8080:7777 \
  -v "$(pwd)/resources:/resources" \
  -e "SPARQL_ENDPOINT=http://dbpedia.org/sparql" \
  -e "RESOURCE_NAMESPACE=http://dbpedia.org/resource/" \
  cbadenes/r4r:latest
```



# Task 10

## Keyword Search Engine + Web Development



# Task 11

## Citizen Science