

# Open knowledge graph on clinical trials

Sandeep Kunkunuru<sup>\*1,2</sup>

<sup>1</sup>IIPH Hyderabad

<sup>2</sup>VaidhyaMegha Private Limited

<sup>\*</sup> Correspondence: Sandeep Kunkunuru <hi18sk@iiphh.org>

## Abstract

Using any clinical trial id from across the globe find the associated diseases, interventions, research articles and genes. Also discover relationships between various medical topics through co-occurrences in articles. Query the graph using SparQL from cli or GraphQL using any API client tool ex: Postman or curl

## Introduction

### Sources

- WHO's ICTRP
  - Registries covered in ICTRP include :
- AACT Clinicaltrials.gov
- NLM MeSH
- NLM MRCOC
- NLM PubMed
- NLM PheGenI

## Methods

### Trial to condition

### Trial to intervention

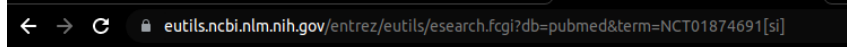
## Clinical trials to research articles

- The NLM (The world's largest medical library, the U.S. National Library of Medicine is part of the National Institutes of Health) extracts trail ids from an article and places them into the article's metadata in secondary id field.

- To retrieve journal articles related to a clinical trial id ex: NCT00000419, use PubMed's API called e-Utils with clinical trial id as shown below:

[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=NCT01874691\[si\]](https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=pubmed&term=NCT01874691[si])

- In the above URL “[si]” refers to Secondary ID which can be used to search within article's metadata.
- All the journal articles related to the clinical trial id will be returned
- Output contains PMIDs (pubmed records) of respective clinical trials.



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```

<?xml version="1.0">
<eSearchResult>
  <Count>9</Count>
  <RetMax>9</RetMax>
  <RetStart>0</RetStart>
  <IdList>
    <Id>34631809</Id>
    <Id>34493500</Id>
    <Id>31567475</Id>
    <Id>31515430</Id>
    <Id>31471442</Id>
    <Id>31255895</Id>
    <Id>30807351</Id>
    <Id>28052755</Id>
    <Id>27530939</Id>
  </IdList>
  <TranslationSet/>
  <TranslationStack>
    <TermSet>
      <Term>NCT01874691[si]</Term>
      <Field>si</Field>
      <Count>9</Count>
      <Explode>N</Explode>
    </TermSet>
    <OP>GROUP</OP>
  </TranslationStack>
  <QueryTranslation>NCT01874691[si]</QueryTranslation>
</eSearchResult>

```

- Using Spring WebClient, JAXB, Jackson and Lambok response XML is automatically parsed and PMID list is constructed in-memory.
- The PMID list is then written into RDF along with trial id using Apache Jena.

## Article to MeSH DUIs

## Gene id to MeSH DUIs

## Results

### Feature list

- Using GraphQL API knowledge graph can be queried using any API client tool ex: curl or Postman.
- Graph includes trials from across the globe. Data is sourced from WHO's ICTRP and clinicaltrials.gov
- Links from trial to MeSH vocabulary are added for conditions and interventions employed in the trial.

- Links from trial to PubMed articles are added. PubMed's experts curate this metadata information for each article.
- Added MRCOC to the graph for the selected articles linked to clinical trials.
- Added PheGenI links i.e. links from phenotype to genotype as links between MeSH DUI and GeneID.
- Added SparQL query execution feature. Adding CLI mode. Adding a count SparQL query for demo.
- 5 co-existing bi-partite graphs together comprise this knowledge graph. Bi-partite graphs are between
  - trial-> condition
  - trial-> intervention
  - trial-> articles
  - article-> MeSH DUIs
  - gene id-> MeSH DUIs

## Demonstration

### Querying knowledge graph using SparQL

```
$ java -jar -Xms4096M -Xmx8144M \
    target/vaidhyamegha-knowledge-graphs-1.0-SNAPSHOT-jar-with-dependencies.jar \
    -m cli -q src/main/sparql/1_count_of_records.rq
...
Results:
-----
4766048~http://www.w3.org/2001/XMLSchema#integer
```

### Querying knowledge graph using GraphQL (via HyperGraphQL)

#### Start server

```
java -Dorg.slf4j.simpleLogger.defaultLogLevel=debug -jar lib/hypergraphql-3.0.1-exe.jar \
    --config src/main/resources/hql-config.json
```

**Start client** In a separate terminal execute GraphQL query using curl (alternatively use Postman)

```
$ curl --location --request POST 'http://localhost:8080/graphql' \
    --header 'Accept: application/ntriples' \
    --header 'Accept-Language: en-GB,en-US;q=0.9,en;q=0.8,kn;q=0.7' \
    --header 'Content-Type: application/json' \
    --data-raw \
    '{"query": "{\n trial_GET(limit: 30, offset: 1) {\n label\n }\n \n}", "variables": {}}'

<https://www.who.int/clinical-trials-registry-platform/EUCTR2007-006072-11-SE>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
```

```

    <https://vaidhyamegha.com/open_kg/ct> .
<https://www.who.int/clinical-trials-registry-platform/EUCTR2007-006072-11-SE>
    <http://www.w3.org/2000/01/rdf-schema#label>
        "EUCTR2007-006072-11-SE"^^<http://www.w3.org/2001/XMLSchema#string> .
<https://clinicaltrials.gov/ct2/show/NCT02954757>
    <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
        <https://vaidhyamegha.com/open_kg/ct> .
<https://clinicaltrials.gov/ct2/show/NCT02954757>
    <http://www.w3.org/2000/01/rdf-schema#label>
        "NCT02954757"^^<http://www.w3.org/2001/XMLSchema#string> .
<https://www.who.int/clinical-trials-registry-platform/EUCTR2014-005525-13-FI>
    <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
        <https://vaidhyamegha.com/open_kg/ct> .
<https://www.who.int/clinical-trials-registry-platform/EUCTR2014-005525-13-FI>
    <http://www.w3.org/2000/01/rdf-schema#label>
        "EUCTR2014-005525-13-FI"^^<http://www.w3.org/2001/XMLSchema#string> .
<https://clinicaltrials.gov/ct2/show/NCT02721914>
    <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
        <https://vaidhyamegha.com/open_kg/ct> .
<https://clinicaltrials.gov/ct2/show/NCT02721914>
    <http://www.w3.org/2000/01/rdf-schema#label>
        "NCT02721914"^^<http://www.w3.org/2001/XMLSchema#string> .
...
<http://hypergraphql.org/query> <http://hypergraphql.org/query/trial_GET>
    <https://www.who.int/clinical-trials-registry-platform/EUCTR2016-002461-66-IT> .
<http://hypergraphql.org/query> <http://hypergraphql.org/query/trial_GET>
    <https://www.who.int/clinical-trials-registry-platform/CTRI/2020/08/027368> .
<http://hypergraphql.org/query> <http://hypergraphql.org/query/trial_GET>
    <https://www.who.int/clinical-trials-registry-platform/EUCTR2013-001294-24-DE> .

```

## Discussions

### Further reading

For more information please read

[1]. [2].

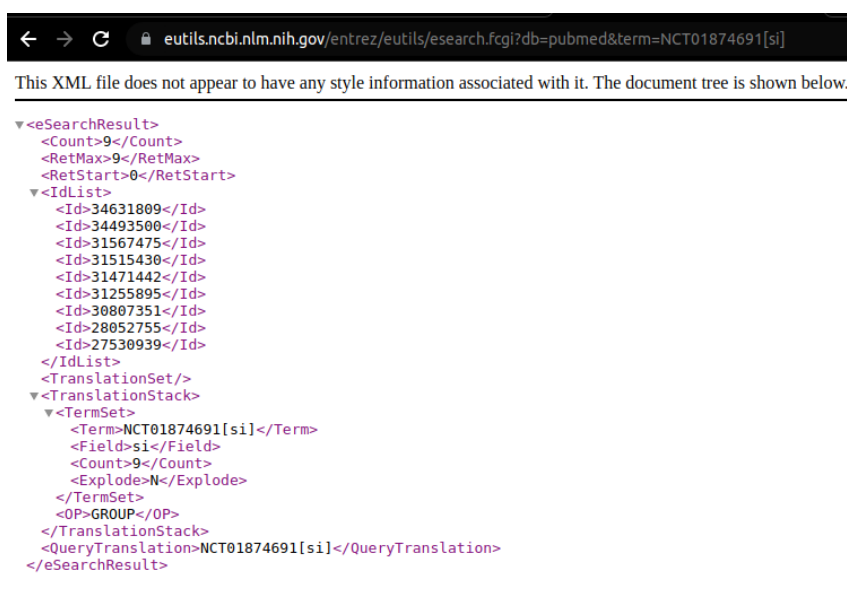
## Acknowledgements

## Declarations

## Appendix

## Tables

## Figures



```
<?xml version="1.0">
<eSearchResult>
  <Count>9</Count>
  <RetMax>9</RetMax>
  <RetStart>0</RetStart>
  <IdList>
    <Id>34631809</Id>
    <Id>34493500</Id>
    <Id>31567475</Id>
    <Id>31515430</Id>
    <Id>31471442</Id>
    <Id>31255895</Id>
    <Id>30807351</Id>
    <Id>28052755</Id>
    <Id>27530939</Id>
  </IdList>
  <TranslationSet/>
  <TranslationStack>
    <TermSet>
      <Term>NCT01874691[si]</Term>
      <Field>si</Field>
      <Count>9</Count>
      <Explode>N</Explode>
    </TermSet>
    <OP>GROUP</OP>
  </TranslationStack>
  <QueryTranslation>NCT01874691[si]</QueryTranslation>
</eSearchResult>
```

## References

- [1] Carroll, J.J. et al. 2004. Jena: Implementing the semantic web recommendations. *Proceedings of the 13th international world wide web conference on alternate track papers & posters* (New York, NY, USA, 2004), 74–83.
- [2] Grobe, M. 2009. RDF, jena, sparql and the 'semantic web'. *Proceedings of the 37th annual acm siguics fall conference: Communication and collaboration* (New York, NY, USA, 2009), 131–138.