# Sander Su

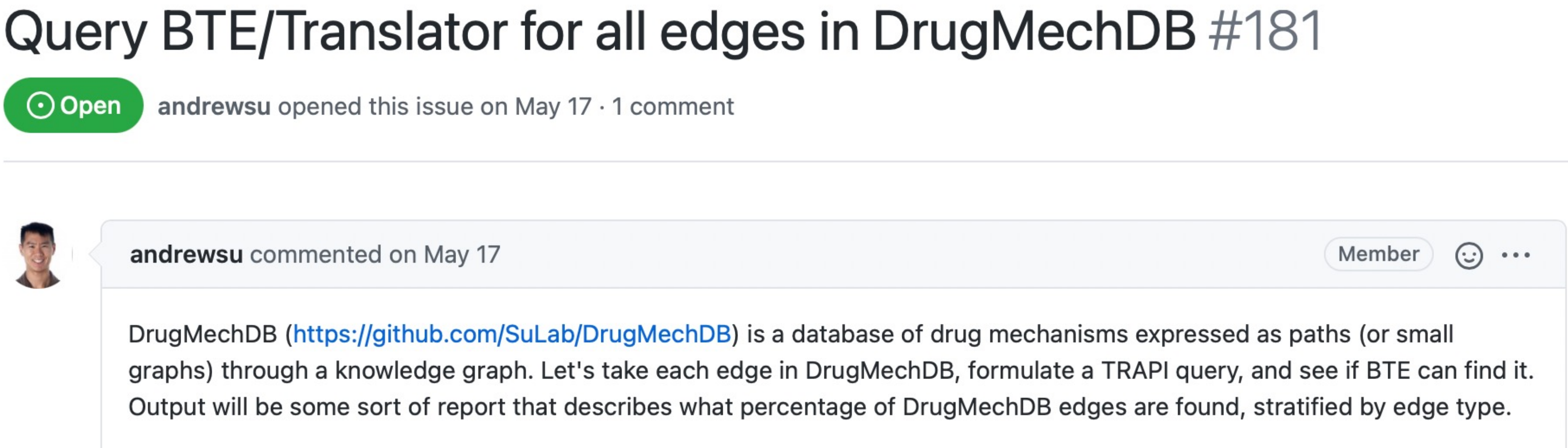
BTE TRAPI Project Report

06-23-2021

# Agenda

- Background (Issue #181)
- Method
- Result
- Summary and Discussion
- Appendix (#201)

# Background (Issue #181)

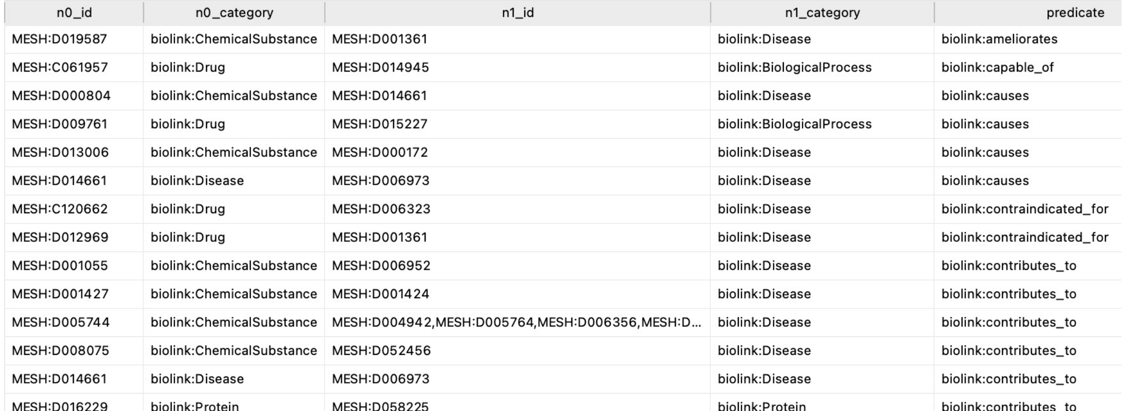


## Query BTE/Translator for all edges in DrugMechDB #181

andrewsu opened this issue on May 17 · 1 comment

andrewsu commented on May 17

DrugMechDB (https://github.com/SuLab/DrugMechDB) is a database of drug mechanisms expressed as paths (or small graphs) through a knowledge graph. Let's take each edge in DrugMechDB, formulate a TRAPI query, and see if BTE can find it. Output will be some sort of report that describes what percentage of DrugMechDB edges are found, stratified by edge type.



| n0_id | n0_category | n1_id | n1_category | predicate |
|---|---|---|---|---|
| MESH:D019587 | biolink:ChemicalSubstance | MESH:D001361 | biolink:Disease | biolink:ameliorates |
| MESH:C061957 | biolink:Drug | MESH:D014945 | biolink:BiologicalProcess | biolink:capable_of |
| MESH:D000804 | biolink:ChemicalSubstance | MESH:D014661 | biolink:Disease | biolink:causes |
| MESH:D009761 | biolink:Drug | MESH:D015227 | biolink:BiologicalProcess | biolink:causes |
| MESH:D013006 | biolink:ChemicalSubstance | MESH:D000172 | biolink:Disease | biolink:causes |
| MESH:D014661 | biolink:Disease | MESH:D006973 | biolink:Disease | biolink:causes |
| MESH:C120662 | biolink:Drug | MESH:D006323 | biolink:Disease | biolink:contraindicated_for |
| MESH:D012969 | biolink:Drug | MESH:D001361 | biolink:Disease | biolink:contraindicated_for |
| MESH:D001055 | biolink:ChemicalSubstance | MESH:D006952 | biolink:Disease | biolink:contributes_to |
| MESH:D001427 | biolink:ChemicalSubstance | MESH:D001424 | biolink:Disease | biolink:contributes_to |
| MESH:D005744 | biolink:ChemicalSubstance | MESH:D004942,MESH:D005764,MESH:D006356,MESH:D... | biolink:Disease | biolink:contributes_to |
| MESH:D008075 | biolink:ChemicalSubstance | MESH:D052456 | biolink:Disease | biolink:contributes_to |
| MESH:D014661 | biolink:Disease | MESH:D006973 | biolink:Disease | biolink:contributes_to |
| MESH:D016229 | biolink:Protein | MESH:D058225 | biolink:Protein | biolink:contributes_to |

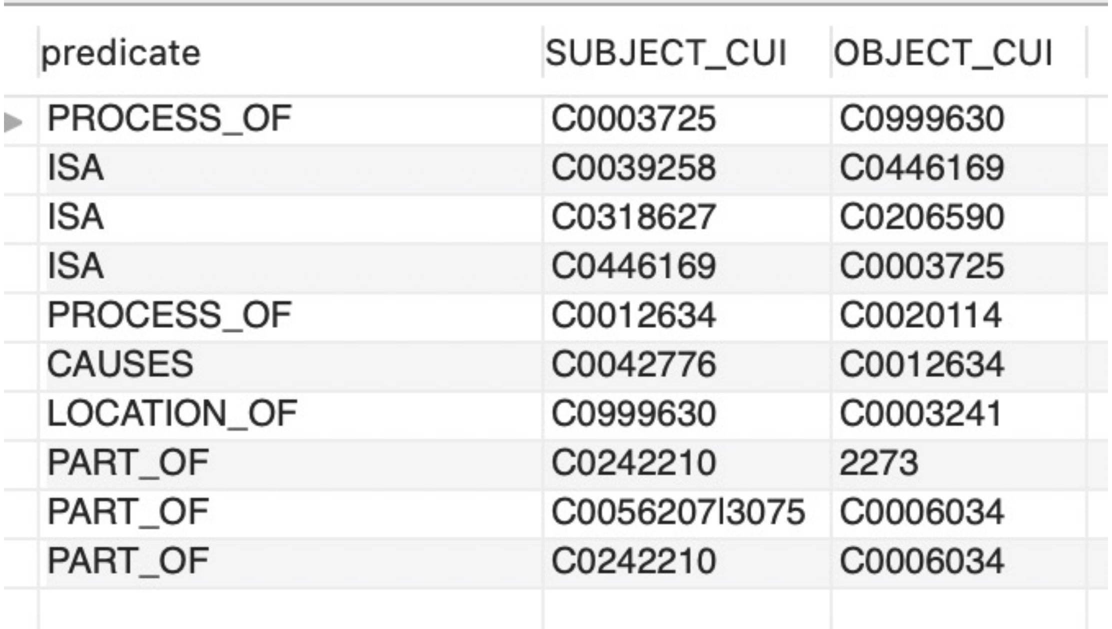
# Background (Issue #181)

- Older version of predicate filter: 67/5,005 (1.33%)
- Current version: 109/5,665 (1.9%)

?

# Method

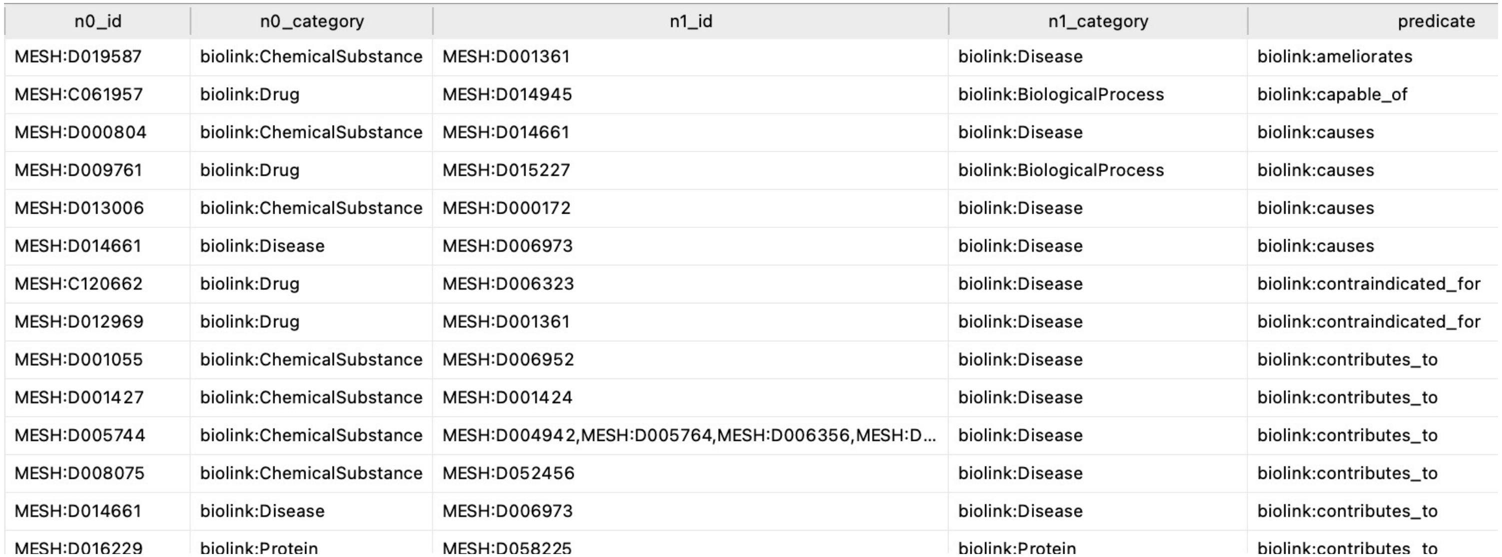
- Step 1:
  - Download SemMedDB and use PREDICATION table for verification



| predicate | SUBJECT_CUI | OBJECT_CUI |
|---|---|---|
| PROCESS_OF | C0003725 | C0999630 |
| ISA | C0039258 | C0446169 |
| ISA | C0318627 | C0206590 |
| ISA | C0446169 | C0003725 |
| PROCESS_OF | C0012634 | C0020114 |
| CAUSES | C0042776 | C0012634 |
| LOCATION_OF | C0999630 | C0003241 |
| PART_OF | C0242210 | 2273 |
| PART_OF | C0056207\|3075 | C0006034 |
| PART_OF | C0242210 | C0006034 |

# Method

- Step 2:
  - Took 325 not found records from DMDB for test (only includes MESH-MESH id pairs for #201)



| n0_id | n0_category | n1_id | n1_category | predicate |
|---|---|---|---|---|
| MESH:D019587 | biolink:ChemicalSubstance | MESH:D001361 | biolink:Disease | biolink:ameliorates |
| MESH:C061957 | biolink:Drug | MESH:D014945 | biolink:BiologicalProcess | biolink:capable_of |
| MESH:D000804 | biolink:ChemicalSubstance | MESH:D014661 | biolink:Disease | biolink:causes |
| MESH:D009761 | biolink:Drug | MESH:D015227 | biolink:BiologicalProcess | biolink:causes |
| MESH:D013006 | biolink:ChemicalSubstance | MESH:D000172 | biolink:Disease | biolink:causes |
| MESH:D014661 | biolink:Disease | MESH:D006973 | biolink:Disease | biolink:causes |
| MESH:C120662 | biolink:Drug | MESH:D006323 | biolink:Disease | biolink:contraindicated_for |
| MESH:D012969 | biolink:Drug | MESH:D001361 | biolink:Disease | biolink:contraindicated_for |
| MESH:D001055 | biolink:ChemicalSubstance | MESH:D006952 | biolink:Disease | biolink:contributes_to |
| MESH:D001427 | biolink:ChemicalSubstance | MESH:D001424 | biolink:Disease | biolink:contributes_to |
| MESH:D005744 | biolink:ChemicalSubstance | MESH:D004942,MESH:D005764,MESH:D006356,MESH:D... | biolink:Disease | biolink:contributes_to |
| MESH:D008075 | biolink:ChemicalSubstance | MESH:D052456 | biolink:Disease | biolink:contributes_to |
| MESH:D014661 | biolink:Disease | MESH:D006973 | biolink:Disease | biolink:contributes_to |
| MESH:D016229 | biolink:Protein | MESH:D058225 | biolink:Protein | biolink:contributes_to |

# Result

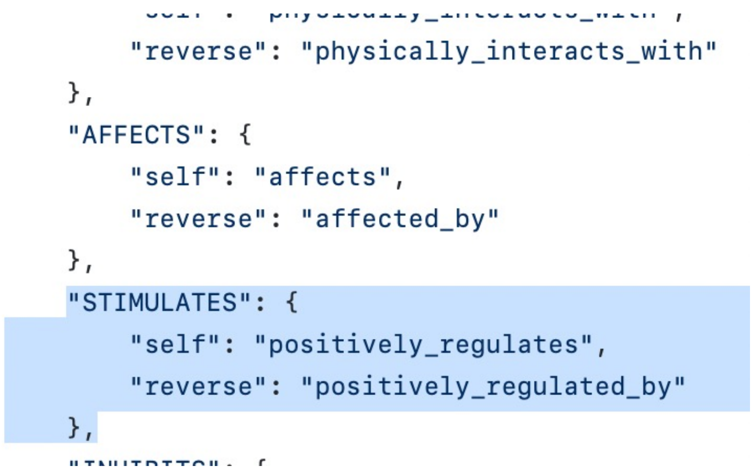
- Found 16/325 records from the table 325 (~ 5%)

# Result

- Why did Translator find nothing?
  - No (updated) records in SEMMED-related APIs in our system. E.g., C0002563 isa C0162713
  - Did not (trigger to) find any edges
    - E.g., C0023566 treats C0029458
  - Inconsistent (missing) mappings
    - E.g., C0033554 augments C0042130

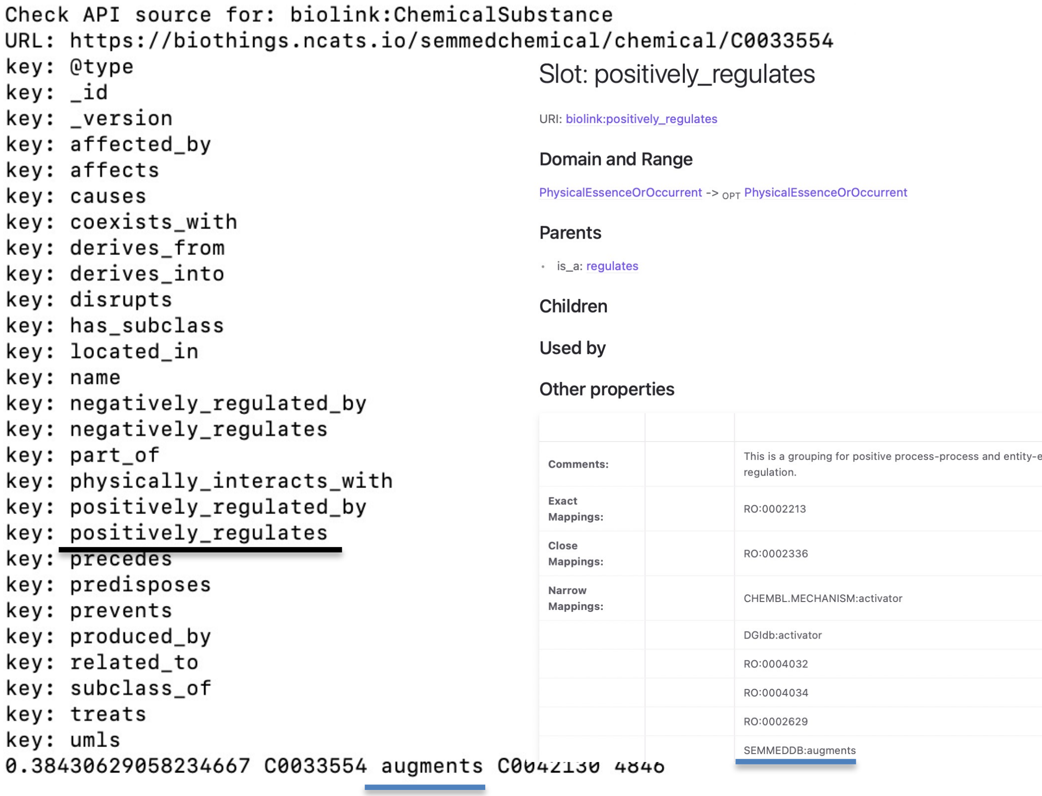
# Result

- Inconsistent (missing) mappings:



```
        "self": "physically_interacts_with",
        "reverse": "physically_interacts_with"
    },
    "AFFECTS": {
        "self": "affects",
        "reverse": "affected_by"
    },
    "STIMULATES": {
        "self": "positively_regulates",
        "reverse": "positively_regulated_by"
    },
    "INHIBITS": {
```

Hard-coded mappings in SEMMED APIs' parsers:



```
Check API source for: biolink:ChemicalSubstance
URL: https://biothings.ncats.io/semmedchemical/chemical/C0033554
key: @type
key: _id
key: _version
key: affected_by
key: affects
key: causes
key: coexists_with
key: derives_from
key: derives_into
key: disrupts
key: has_subclass
key: located_in
key: name
key: negatively_regulated_by
key: negatively_regulates
key: part_of
key: physically_interacts_with
key: positively_regulated_by
key: positively_regulates
key: precedes
key: predisposes
key: prevents
key: produced_by
key: related_to
key: subclass_of
key: treats
key: umls
0.38430629058234667 C0033554 augments C0042130 4846
```

Slot: positively_regulates

URI: biolink:positively_regulates

Domain and Range

PhysicalEssenceOrOccurrent -> OPT PhysicalEssenceOrOccurrent

Parents

- is_a: regulates

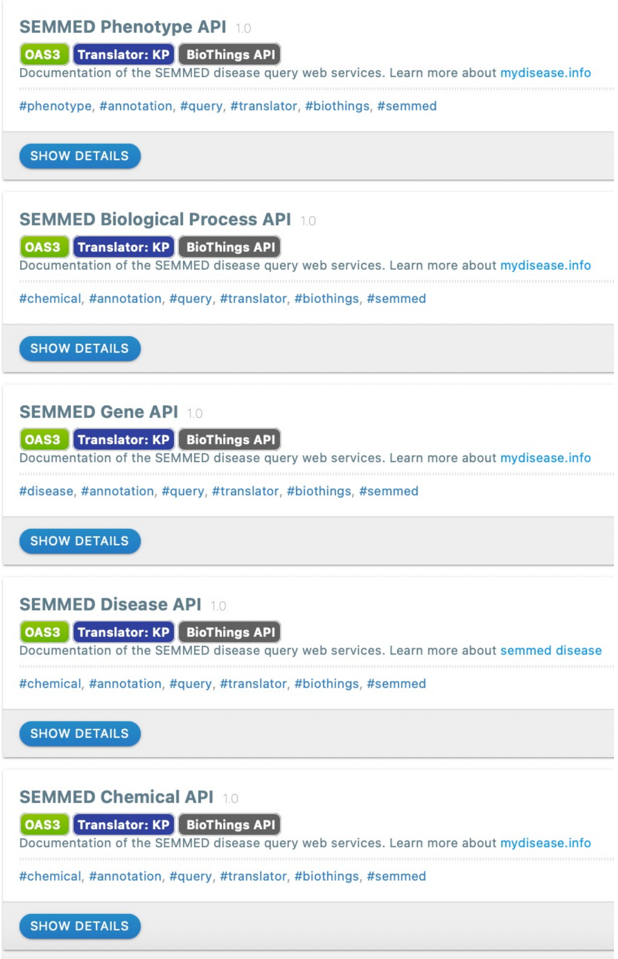
Children

Used by

Other properties

| | | |
|---|---|---|
| Comments: | | This is a grouping for positive process-process and entity-e regulation. |
| Exact Mappings: | | RO:0002213 |
| Close Mappings: | | RO:0002336 |
| Narrow Mappings: | | CHEMBL.MECHANISM:activator |
| | | DGIdb:activator |
| | | RO:0004032 |
| | | RO:0004034 |
| | | RO:0002629 |
| | | SEMMEDDB:augments |

# Summary and Discussion



SEMMED Phenotype API 1.0

Documentation of the SEMMED disease query web services. Learn more about mydisease.info

#phenotype, #annotation, #query, #translator, #biothings, #semmed

SEMMED Biological Process API 1.0

Documentation of the SEMMED disease query web services. Learn more about mydisease.info

#chemical, #annotation, #query, #translator, #biothings, #semmed

SEMMED Gene API 1.0

Documentation of the SEMMED disease query web services. Learn more about mydisease.info

#disease, #annotation, #query, #translator, #biothings, #semmed

SEMMED Disease API 1.0

Documentation of the SEMMED disease query web services. Learn more about semmed disease

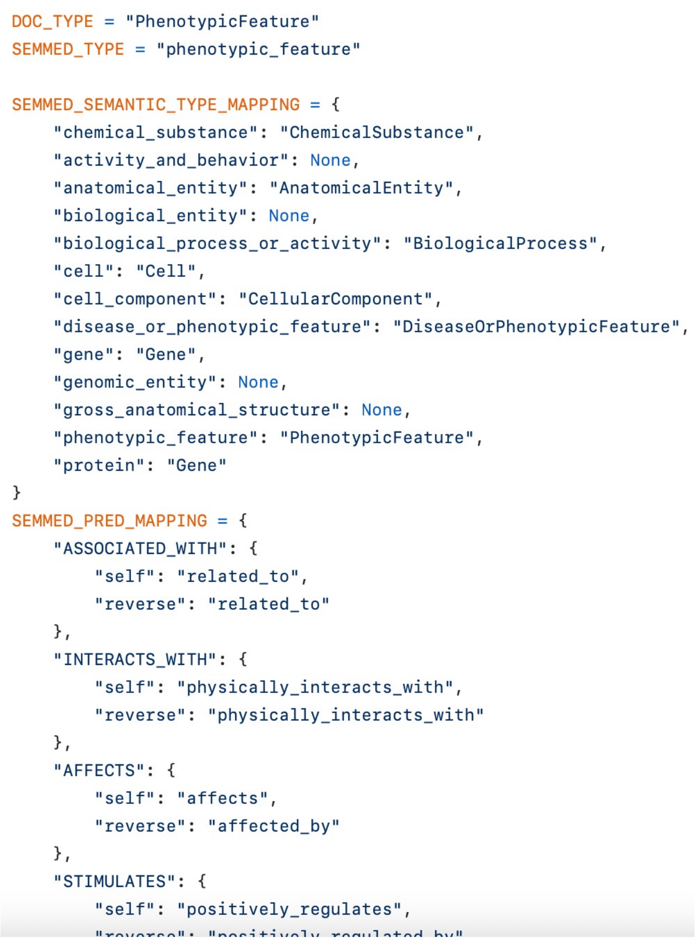
#chemical, #annotation, #query, #translator, #biothings, #semmed

SEMMED Chemical API 1.0

Documentation of the SEMMED disease query web services. Learn more about mydisease.info

#chemical, #annotation, #query, #translator, #biothings, #semmed

Copy and paste

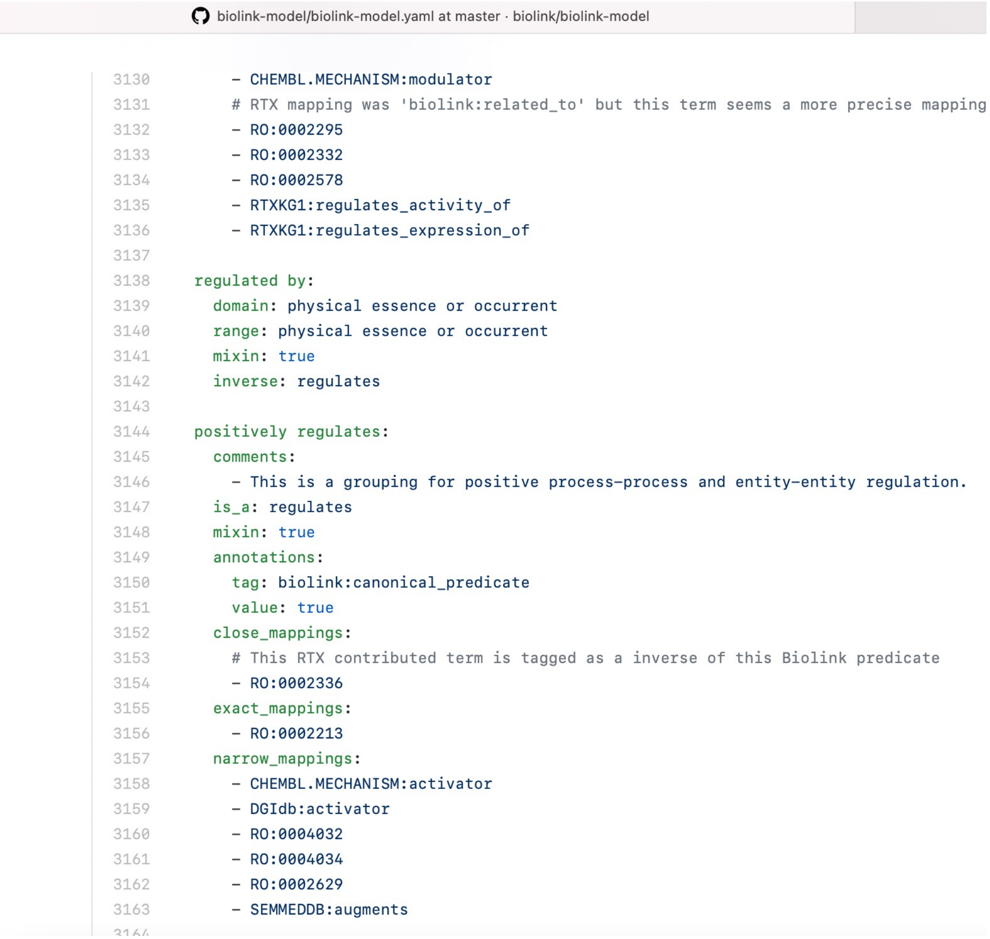


```
DOC_TYPE = "PhenotypicFeature"
SEMMED_TYPE = "phenotypic_feature"

SEMMED_SEMANTIC_TYPE_MAPPING = {
    "chemical_substance": "ChemicalSubstance",
    "activity_and_behavior": None,
    "anatomical_entity": "AnatomicalEntity",
    "biological_entity": None,
    "biological_process_or_activity": "BiologicalProcess",
    "cell": "Cell",
    "cell_component": "CellularComponent",
    "disease_or_phenotypic_feature": "DiseaseOrPhenotypicFeature",
    "gene": "Gene",
    "genomic_entity": None,
    "gross_anatomical_structure": None,
    "phenotypic_feature": "PhenotypicFeature",
    "protein": "Gene"
}
SEMMED_PRED_MAPPING = {
    "ASSOCIATED_WITH": {
        "self": "related_to",
        "reverse": "related_to"
    },
    "INTERACTS_WITH": {
        "self": "physically_interacts_with",
        "reverse": "physically_interacts_with"
    },
    "AFFECTS": {
        "self": "affects",
        "reverse": "affected_by"
    },
    "STIMULATES": {
        "self": "positively_regulates",
        "reverse": "positively_regulated_by"
```

# Summary and Discussion

https://github.com/biolink/biolink-model/blob/master/biolink-model.yaml



biolink-model/biolink-model.yaml at master · biolink/biolink-model

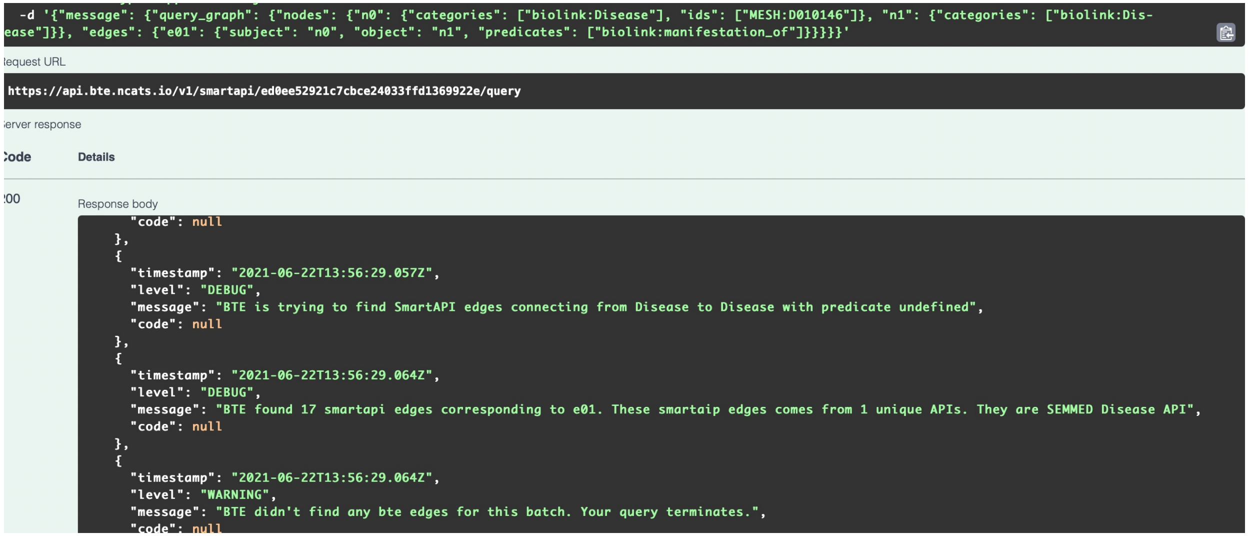
```yaml
      - CHEMBL.MECHANISM:modulator
      # RTX mapping was 'biolink:related_to' but this term seems a more precise mapping
      - RO:0002295
      - RO:0002332
      - RO:0002578
      - RTXKG1:regulates_activity_of
      - RTXKG1:regulates_expression_of

  regulated by:
    domain: physical essence or occurrent
    range: physical essence or occurrent
    mixin: true
    inverse: regulates

  positively regulates:
    comments:
      - This is a grouping for positive process-process and entity-entity regulation.
    is_a: regulates
    mixin: true
    annotations:
      tag: biolink:canonical_predicate
      value: true
    close_mappings:
      # This RTX contributed term is tagged as a inverse of this Biolink predicate
      - RO:0002336
    exact_mappings:
      - RO:0002213
    narrow_mappings:
      - CHEMBL.MECHANISM:activator
      - DGIdb:activator
      - RO:0004032
      - RO:0004034
      - RO:0002629
      - SEMMEDDB:augments
```

# Summary and Discussion

- Did not find mappings in the API’s knowledge graph



```
  -d '{"message": {"query_graph": {"nodes": {"n0": {"categories": ["biolink:Disease"], "ids": ["MESH:D010146"]}, "n1": {"categories": ["biolink:Dis-
ease"]}}, "edges": {"e01": {"subject": "n0", "object": "n1", "predicates": ["biolink:manifestation_of"]}}}}}'
```

Request URL

```
https://api.bte.ncats.io/v1/smartapi/ed0ee52921c7cbce24033ffd1369922e/query
```

Server response

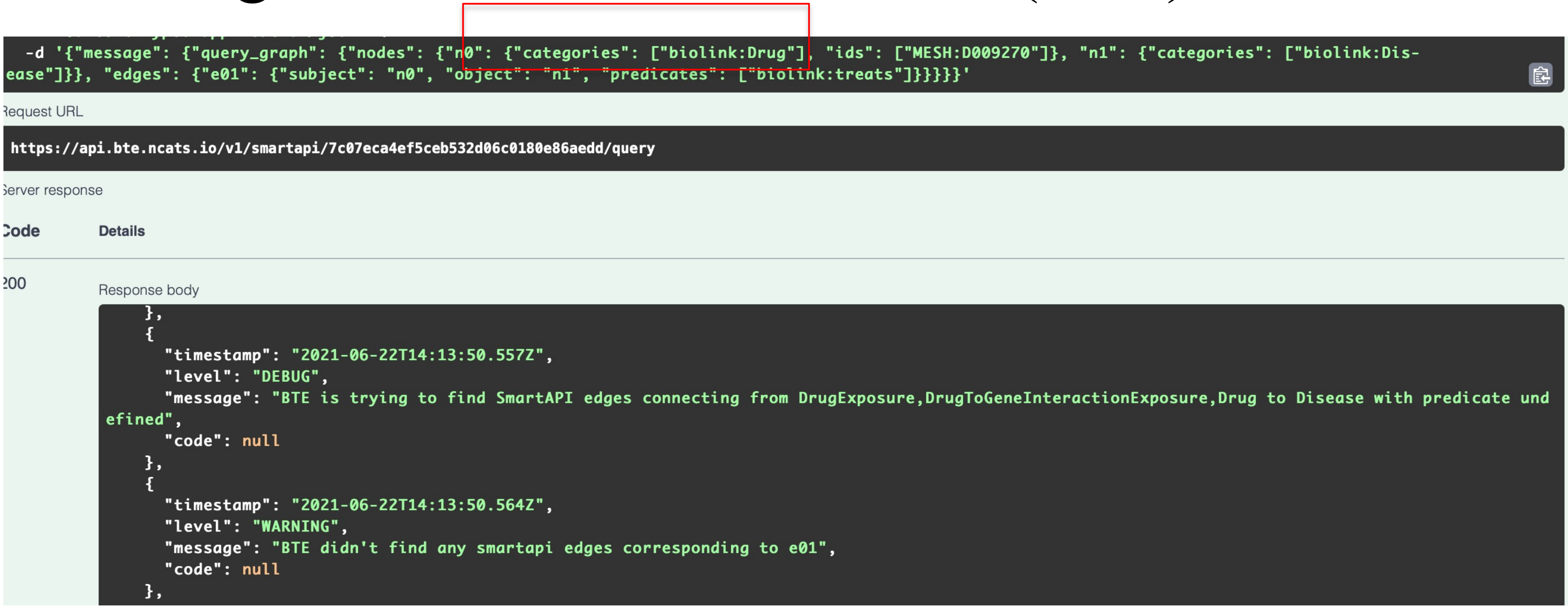
| Code | Details |
|---|---|
| 200 | Response body |

```
      "code": null
    },
    {
      "timestamp": "2021-06-22T13:56:29.057Z",
      "level": "DEBUG",
      "message": "BTE is trying to find SmartAPI edges connecting from Disease to Disease with predicate undefined",
      "code": null
    },
    {
      "timestamp": "2021-06-22T13:56:29.064Z",
      "level": "DEBUG",
      "message": "BTE found 17 smartapi edges corresponding to e01. These smartaip edges comes from 1 unique APIs. They are SEMMED Disease API",
      "code": null
    },
    {
      "timestamp": "2021-06-22T13:56:29.064Z",
      "level": "WARNING",
      "message": "BTE didn't find any bte edges for this batch. Your query terminates.",
      "code": null
```

```
    "subject": "biolink:Disease",
    "predicate": "biolink:related_to",
    "object": "biolink:BiologicalProcess",
    "relation": null
  },
  {
    "subject": "biolink:Disease",
    "predicate": "biolink:treated_by",
    "object": "biolink:BiologicalProcess",
    "relation": null
  },
  {
    "subject": "biolink:Disease",
    "predicate": "biolink:affected_by",
    "object": "biolink:Disease",
    "relation": null
  },
  {
    "subject": "biolink:Disease",
    "predicate": "biolink:affects",
    "object": "biolink:Disease",
    "relation": null
  },
  {
    "subject": "biolink:Disease",
    "predicate": "biolink:caused_by",
    "object": "biolink:Disease",
    "relation": null
  },
  {
    "subject": "biolink:Disease",
    "predicate": "biolink:causes",
    "object": "biolink:Disease",
    "relation": null
  },
  {
    "subject": "biolink:Disease",
    "predicate": "biolink:coexists_with",
    "object": "biolink:Disease",
    "relation": null
  },
  {
    "subject": "biolink:Disease",
    "predicate": "biolink:disrupted_by",
    "object": "biolink:Disease",
    "relation": null
  },
  {
    "subject": "biolink:Disease",
    "predicate": "biolink:disrupts",
    "object": "biolink:Disease",
    "relation": null
  },
  {
    "subject": "biolink:Disease",
    "predicate": "biolink:entity_negatively_regulated_by_entity",
    "object": "biolink:Disease",
    "relation": null
  },
  {
    "subject": "biolink:Disease",
    "predicate": "biolink:entity_negatively_regulates_entity",
    "object": "biolink:Disease",
    "relation": null
```

# Summary and Discussion

- Drug <-> ChemicalSubstance (v1.1)



```
  -d '{"message": {"query_graph": {"nodes": {"n0": {"categories": ["biolink:Drug"], "ids": ["MESH:D009270"]}, "n1": {"categories": ["biolink:Dis-
ease"]}}, "edges": {"e01": {"subject": "n0", "object": "n1", "predicates": ["biolink:treats"]}}}}}'
```

Request URL

```
https://api.bte.ncats.io/v1/smartapi/7c07eca4ef5ceb532d06c0180e86aedd/query
```

Server response

| Code | Details |
|---|---|
| 200 | Response body |

```
    },
    {
      "timestamp": "2021-06-22T14:13:50.557Z",
      "level": "DEBUG",
      "message": "BTE is trying to find SmartAPI edges connecting from DrugExposure,DrugToGeneInteractionExposure,Drug to Disease with predicate und
efined",
      "code": null
    },
    {
      "timestamp": "2021-06-22T14:13:50.564Z",
      "level": "WARNING",
      "message": "BTE didn't find any smartapi edges corresponding to e01",
      "code": null
    },
```

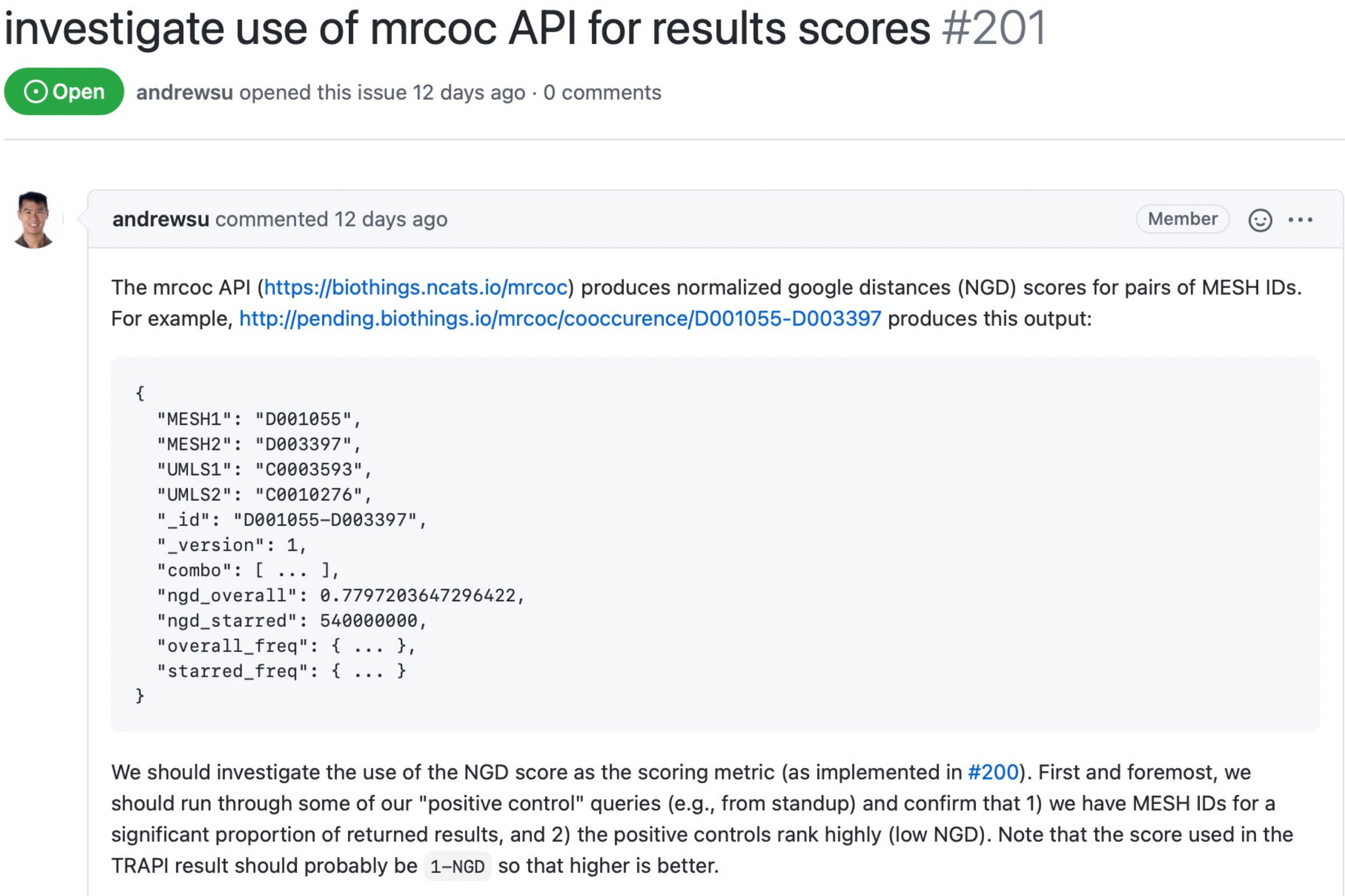
# Summary and Discussion

- #181 will be addressed partially until fixing possible issues
- Review KPs and fix related issues

Thank you

# Appendix

# Background (Issue # 201)



## investigate use of mrcoc API for results scores #201

andrewsu opened this issue 12 days ago · 0 comments

andrewsu commented 12 days ago

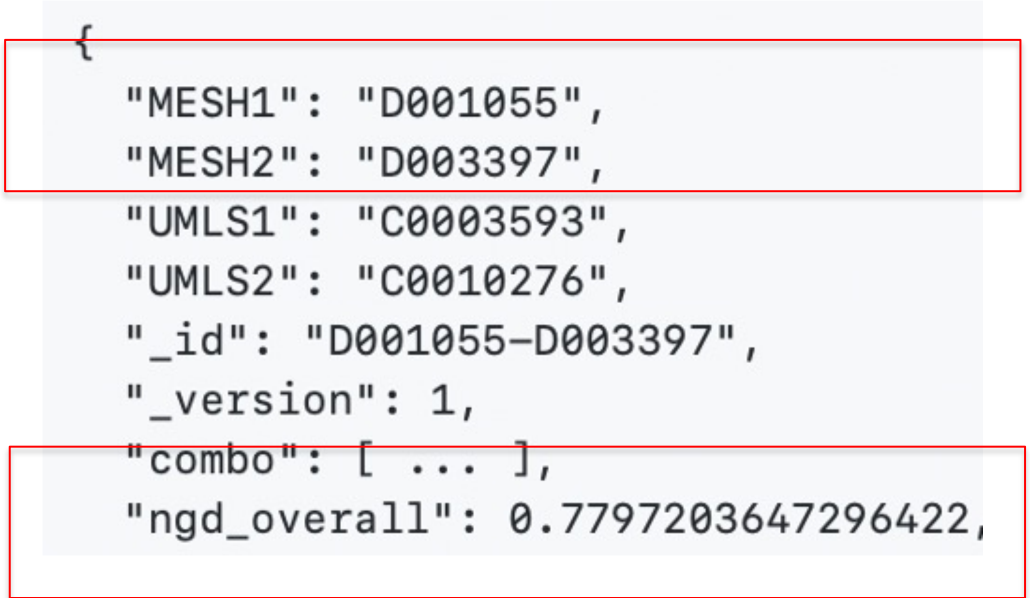
The mrcoc API (https://biothings.ncats.io/mrcoc) produces normalized google distances (NGD) scores for pairs of MESH IDs. For example, http://pending.biothings.io/mrcoc/cooccurence/D001055-D003397 produces this output:

```
{
  "MESH1": "D001055",
  "MESH2": "D003397",
  "UMLS1": "C0003593",
  "UMLS2": "C0010276",
  "_id": "D001055-D003397",
  "_version": 1,
  "combo": [ ... ],
  "ngd_overall": 0.7797203647296422,
  "ngd_starred": 540000000,
  "overall_freq": { ... },
  "starred_freq": { ... }
}
```

We should investigate the use of the NGD score as the scoring metric (as implemented in #200). First and foremost, we should run through some of our "positive control" queries (e.g., from standup) and confirm that 1) we have MESH IDs for a significant proportion of returned results, and 2) the positive controls rank highly (low NGD). Note that the score used in the TRAPI result should probably be `1-NGD` so that higher is better.

# Method

- Step:
  - Found 325 records from PREDICATION table with NGD scores from #201



```
{
   "MESH1": "D001055",
   "MESH2": "D003397",
   "UMLS1": "C0003593",
   "UMLS2": "C0010276",
   "_id": "D001055-D003397",
   "_version": 1,
   "combo": [ ... ],
   "ngd_overall": 0.7797203647296422,
```

# Result

- Found 16 records from 325 submissions (~ 5%)
- NGD score distribution (0~1):

| | 0~0.25 | 0.26-0.5 | 0.51-0.75 | 0.76-1 |
|---|---|---|---|---|
| Not found records | 10% | 47% | 30% | 13% |
| Found records | 25% | 63% | 6% | 6% |

# Summary and Discussion

- The outcome of #201 can be used for sorting/filtering queries and return outcome