



Sander Su

Integrated Dietary Supplement Knowledge Base

Data Analysis

06-01-2021

New API for iDISK

– Background

The first step would be to create a data parser, following the instructions here:

<https://docs.biothings.io/en/latest/doc/studio.html#part-1-single-datasource>. In this case, the primary key for the document should be a concept identifier for a dietary ingredient, and the document should contain all the relationships stated in the database.

For example, here is a set of lines from MRREL.RRF, which contains these relationships:

```
DR1378627|DC0477352|is_effective_for|DC0478476|NMCD
DR1378629|DC0477352|has_adverse_effect_on|DC0478449|NMCD
DR1378613|DC0477352|is_effective_for|DC0478466|NMCD
```

this snippet declares that DC0477352 has three relationships, that it is `is_effective_for` - DC0478476, that it `has_adverse_effect_on` - DC0478449, and that it `is_effective_for` - DC0478466. All of those facts were originally sourced from the "NMCD" database. Those concept identifiers can be resolved in the MRCONSO.RRF file using either UMLS or MEDDRA identifiers:

```
$ egrep 'DC0477352|DC0478476|DC0478466|DC0478449' MRCONSO.RRF | egrep 'UMLS|MEDDRA'
DC0477352|DA0137592|7-oxodehydroepiandrosterone|SY|UMLS|C0525091|Y
DC0477352|DA0137591|7-oxodehydroepiandrosterone|SY|UMLS|C0525091|Y
DC0478449|DA0170196|Nervous system disorders|SY|MEDDRA|10029205|Y
DC0478466|DA0170257|Obesity, NOS|SY|UMLS|C0028754|Y
DC0478476|DA0170285|RAYNAUD PHENOMENON|SY|UMLS|C0034735|Y
DC0478476|DA0170283|RAYNAUD PHENOMENON|SY|UMLS|C0034735|Y
```

So the dietary ingredient 7-oxodehydroepiandrosterone (UMLS: C0525091) is related to RAYNAUD PHENOMENON (UMLS: C0034735) because 7-oxodehydroepiandrosterone `is_effective_for` (treating) RAYNAUD PHENOMENON. Similarly, that ingredient `has_adverse_effect_on` - Nervous systems disorders, and it `is_effective_for` - Obesity, NOS.

New API for iDISK

– Background

```
{
  "UMLS:C0525091": {
    "name": "7-oxodehydroepiandrosterone",
    "umls": "UMLS:C0525091",
    "is_effective_for": [
      {
        "name": "RAYNAUD PHENOMENON",
        "umls": "UMLS:C0034735",
        "source": [
          {
            "name": "iDISK",
            "record": "DR1378627"
          },
          {
            "name": "NMCD"
          }
        ]
      }
    ],
    {
      "name": "Obesity, NOS",
      "umls": "UMLS:C0028754",
      "source": [
        {
          "name": "iDISK",
          "record": "DR1378613"
        },
        {
          "name": "NMCD"
        }
      ]
    }
  ],
  "has_adverse_effect_on": [
    {
      "name": "Nervous systems disorders",
      "meddra": "MEDDRA:0029205",
      ..
    }
  ]
}
```

New API for iDISK

- **Source:** <https://conservancy.umn.edu/handle/11299/204783>

File View/Open	Description	Size	Format
README.md	Description of the data file formats (TXT)	4.712Kb	text/markdown
README.htm	Description of the data file formats (HTML)	81.73Kb	HTML
idisk-rrf-1.0.1-2020-01-28.zip	iDISK RRF	9.565Mb	application/zip

– Main files

- MRREL: Relation file of concepts (708,164 records)
 - 919 concepts
- MRCONSO: Detail of concept (209,499 records)
 - Sources (UMLS or MEDDRA): 10,527 records

New API for iDISK

- MRREL (similar to UMLS concept)

rui	cui1	rel	cui2	sab
DR0000007	DC0001355	has_ingredient	DC0478167	DSL
DR0000013	DC0001355	has_ingredient	DC0478268	DSL
DR0000015	DC0001355	has_ingredient	DC0477973	DSL
DR0000019	DC0001355	has_ingredient	DC0483034	DSL
DR0000009	DC0001355	has_ingredient	DC0483298	DSL
DR0000011	DC0001355	has_ingredient	DC0482716	DSL
DR0000003	DC0001355	has_ingredient	DC0483261	DSL
DR0000017	DC0001355	has_ingredient	DC0483608	DSL
DR0000005	DC0001355	has_ingredient	DC0482541	DSL
DR0000023	DC0001357	has_ingredient	DC0478191	DSL
DR0000025	DC0001358	has_ingredient	DC0477963	DSL

New API for iDISK

– MRCONSO

cui	au	str	tty	sab	scode	ispref
DC0478122	DA0159849	Rose Hip	SY	DSL	118	Y
DC0478122	DA0159871	ROSE HIPS	SY	UMLS	C0772037	Y
DC0478122	DA0159872	ROSE HIPS	SY	UMLS	C0772037	Y
DC0478122	DA0159873	ROSE HIPS	SY	UMLS	C0772037	Y
DC0478122	DA0159898	Rose hips	SN	NHPID	3972	Y
DC0478122	DA0159843	Rose Hip	SY	NMCD	839	Y
DC0478122	DA0159920	Rose hip	SN	NHPID	0432	Y
DC0478122	DA0159904	Rosehips	SY	DSL	118	N
DC0478122	DA0159858	Rose Hips fruit powder	SY	DSL	118	N
DC0478122	DA0159891	Hop Fruit	SY	NMCD	839	N
DC0478122	DA0159881	Satapatrika	SY	NMCD	839	N

Steps:

1. Find unique cui1s and their cui2s from MRREL file
2. Find cuis with UMLS or MEDDRA sources only from MRCONSO file
3. Loop cui1s and cui2s from step 1 to get the detail of cui2 from step 2
4. Combine findings from step 1 and step 3 as JSON objects

New API for iDISK

– JSON Output

```
"_id": "DSLID:30552",
"name": "Douglas Laboratories - L-Lysine 500 mg",
"dsld": "DSLID:30552",
"has_ingredient": [
  {
    "name": "AjiPure pharmaceutical grade L-Lysine HCl",
    "ispref": "N",
    "dsld": "DSLID:172",
    "source": [
      {
        "name": "iDisk",
        "record": "DR0000001"
      },
      {
        "name": "DSLID"
      }
    ]
  },
  {
    "name": "Hydrochlorure de L-Lysine",
    "ispref": "N",
    "nmcd": "NMCD:237",
    "source": [
      {
        "name": "iDisk",
        "record": "DR0000001"
      },
      {
        "name": "DSLID"
      }
    ]
  }
],
```



Thank you