README.md 11/29/2018

DjangoAPI

Download

Download zip file

or

• git clone https://github.com/phi-grib/RDTExtractorWeb

Install

Install and activate the environment

```
conda env create -f environment.yml source activate RDTExtractor
```

You will need to put the data files we will provide in the API/static/data/ folder.

Run

In the root folder execute:

```
python manage.py runserver
```

and then navigate to http://127.0.0.1:8000.

Introduction

This tool is designed to extract data from the *in vivo* repeat-dose toxicity (RDT) studies' database generated within the context of the eTOX project. These data are expanded using an histopathological observation and an anatomical entity ontologies. The histopathological ontology is obtained from Novartis and can be used under the Apache License 2.0. The anatomical entities ontology is extracted from the following paper:

• Hayamizu TF, Mangan M, Corradi JP, Kadin JA, Ringwald M. Genome Biol. 2005; 6(3): R29

In order to be able to aggregate the data by parent compound, some pre-processing has to be done to data as they exist in the database. Each substance is standardised according to the following protocol:

- From this repository use the *process_smiles.std* method to standardize, discard mixtures, discard compound with metal ions, and remove all salts. Also use the *neutralise.run* method to neutralise all charges when possible.
- Using molVS, get the canonical tautomer.

This project is an extension of the work published in the following paper:

 López-Massaguer O, Pinto-Gil K, Sanz F, Amberg A, Anger LT, Stolte M, Ravagli C, Marc P, Pastor M. Toxicol Sci. 2018 Mar; 162(1): 287–300.

Manual

README.md 11/29/2018

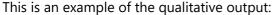
Exract studies' findings based on the given filtering and the organs' and morphological changes' ontologies-based expansions of these findings.

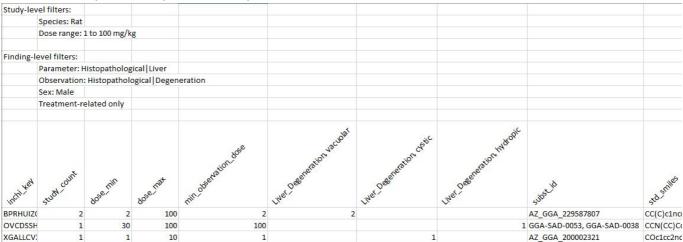
Output example

On clicking the 'Extract' button, two output files are generated, one with quantitative and the other with qualitative data. Both have a caption summarising the filtering criteria applied. After this caption, they both have a table with the data aggergated by parent compound. The table contains several fixed columns, namely 6 at the beginning:

- inchi_key: Parent compound's InChlKey.
- study_count: Number of relevant studies (according to the current filtering scheme) in which the compound appears.
- dose_min: Minimum dose at which the compound has been tested among the relevant studies.
- dose_max: Maximum dose at which the compound has been tested among the relevant studies.
- min_observation_dose: Minimum dose for which a relevant finding (according to the current filtering scheme) has been reported for the compound.
- is_active: Boolean indicating whether the substance has been found to have any toxicity according to the current finding-related filtering criteria. And two at the end:
- subst_id: All substance IDs corresponding to the parent compound.
- std_smiles: Smiles string corresponding to the standardised parent compound.

Between these two groups, there is a column for each relevant finding. In these columns a value is provided if the finding is reported for the given substance, and it is empty otherwise. The value will be the number of studies that report the finding in the qualitative file, and the minimum dose at which the finding is reported in the quantitative file.





README.md 11/29/2018

This is an example of the quantitative output:

Study-lev	el filters:								
	Species: Rat								
	Dose range: 1	l to 100 mg/kg	3						
Finding-le	evel filters:								
	Parameter: H	listopatho <mark>l</mark> ogi	cal Liver						
	Observation:	Histopatholo	gical Degen	eration					
	Sex: Male								
	Treatment-re	elated only							
inchixed	Study Count	dose min	bose not	nin observation	Liver Degenera	Liver Degenera	Liver Degener	tion hydropic	std strikes
BPRHUIZO		2	100	2	2		•	AZ GGA 22	9 CC(C)c1nc(
OVCDSSH	1	30	100	100			100	GGA-SAD-0	
XGALLCV	1		10	1		1		AZ_GGA_20	72 32