DjangoAPI

Install

Install the enviroment

In the root folder execute "python manage.py runserver"

Run

```
% python manage.py runserver
```

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

The script can work with version 2016.1 or with later versions. For the former, you need to request access to the data files from us and place these files in the data folder. For the latter, you need to have the Oracle database provided by Lhasa.

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:

- Use the process_smiles.std method from https://github.com/bet-gregori/standardiser to standardize, discard mixtures, discard compound with metal ions, and remove all salts.
- Neutralize all charges.
- Differentiate stereoisomers.
- Merge all tautomers.

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

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 begining:

- inchi_key: Parent compound's InChIKey.
- 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 1 in the qualitative file and the minimum dose at which the finding is reported in the quantitative file.

This is an example of the qualitative output:

subst_id	study_count	dose_max	dose_min	is_active	liver_basophilic focus	liver_benign tumor	liver_bile duct hyperplasia	liver_cell adaptation/injury/death	liver_cell injury/death	liver_cellular adaptation	liver_cellular adaptation of growth	liver_chronic inflammatory/proliferative/metaplastic chang	liver_clear cell focus	liver_compartmental cell accumulation	liver_congestion	liver_cyst with squamous/ciliated epithelium	liver_cytoplasmic alteration	liver_decreased
X	9	650	1	True			1	1	1	1	1							1
X	1	504	97.9	False														
X	1	25	8	False														
X	2	400	400	False														
X	1	10	1	False														
X	1	1000	1000	False														
X	8	2000	10	True				1			1							
X	1	110	64	False														
X	1	140	120	True				1	1	1								1
X	8	2000	1	True	1	1	1	1	1	1	1							1
X	4	150	18.9	False														
X	1	2000	2000	False														
X	1	30	30	False														
X	1	3	0.2	False														
X	1	2	1	False														
X	4	120.5	0.5	True				1	1		1							
X	2	650	60	True				1		1	1							1
X	5	500	5	True				1	1	1	1				1			
X	1	50	13	False														
X	5	12	0.04	True				1	1		1							
X	1	100	100	False														
X	2	1000	50	True				1										
X	1	350	71	False														
X	6	500	50	True				1	1	1	1							1
X	1	150	10	True				1	1									
X	5	450	45	True				1		1				1				

This is an example of the quantitative output:

subst_id	study_count	dose_max	dose_min	is_active	erythropoiesis	liver_increased mitoses	liver_increased, cellular content	liver_increased, histiocyte number	liver_increased, lipid content	liver_infarct	liver_infectious diseases	liver_inflammation	liver_inflammation, granulomatous	liver_inflammatory cell infiltration	liver_inflammatory processes	liver_intracellular accumulation	liver_intracellula
X	9	650	1	True		5	1		1			650			650	650	
X	1	504	97.9	False													
X	1	25	8	False													
X	2	400	400	False													
X	1	10	1	False													
X	1	1000	1000	False													
X	8	2000	10	True													
X	1	110	64	False													
X	1	140	120	True												120	
X	8	2000	1	True			1		1			10		5	5	5	
X	4	150	18.9	False													
X	1	2000	2000	False													
X	1	30	30	False													
X	1	3	0.2	False													
X	1	2	1	False													
X	4	120.5	0.5	True		0.5											
X	2	650	60	True													
X	5	500	5	True			5		5			5		5	5	5	
X	1	50	13	False													
X	5	12	0.04	True		1.2										0.12	
X	1	100	100	False													
X	2	1000	50	True												50	