==========
*
* BioCreative VI
*
* Text mining chemical-protein interactions (CHEMPROT) track
*
*
* Test set - version 1.0 - September 12th
*
*
* URL: http://www.biocreative.org/tasks/biocreative-vi/track-5/
*
*
* contact e-mail: krallinger.martin@gmail.com
*
==========

This directory contains the BioCreative VI CHEMPROT track test set abstracts and manual annotation entity mention annotations.

Important: Do revise the ChemProt Sample set for additional details on the used annotation guidelines and example predictions/format. It is available at:

http://www.biocreative.org/media/store/files/2017/chemprot_sample.zip

Test set submission instructions will be placed on the ChemProt tack webpage soon.

1. Test set abstracts

- File: *chemprot_test_abstracts.tsv*

This file contains plain-text, UTF8-encoded CHEMPROT test set PubMed record in a tab-separated format with the following three columns:

1- Article identifier (PMID, PubMed identifier)
2- Title of the article
3- Abstract of the article

In total 3399 PubMed records are included in the ChemProt test set. In this file, each line contains a single PMID, title and abstract separated by tabulators.

Important: from these records a subset of **800 abstracts correspond to the Gold Standard test set**. We have included an additional collection of 2599 to make sure that participating systems do scale and avoid manual corrections of the results. A similar setting was used for previous BioCreative evaluation efforts (e.g. CHEMDNER tracks).

Participants need send predictions for the entire test set data collection of 3399 records.

2. Entity mention annotations

- File: *chemprot_test_entities.tsv*

This file contains the manually labeled mention annotations of chemical compounds and genes/proteins (so-called gene and protein related objects – GPRO as defined during BioCreative V) generated for the test set records.

This file consists of tab-separated fields containing:

1- Article identifier (PMID)
2- Entity or term number (for this record)
3- Type of entity mention (CHEMICAL, GENE-Y, GENE-N)*
4- Start character offset of the entity mention
5- End character offset of the entity mention
6- Text string of the entity mention

* CHEMICAL: Chemical entity mention type; GENE-Y: gene/protein mention type that can be normalized or associated to a biological database identifier; GENE-N: gene/protein mention type that cannot be normalized to a database identifier. (See ChemProt sample set for additional details).

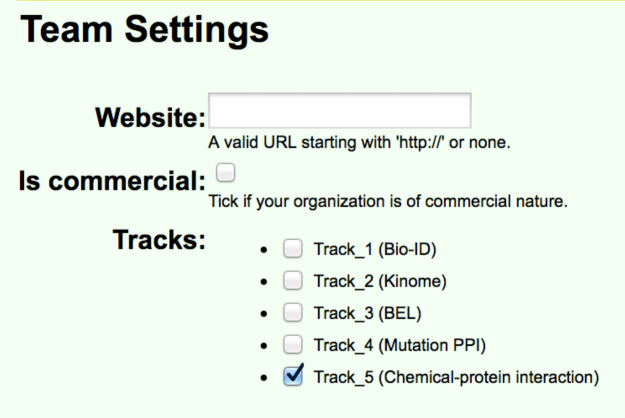
Example CHEMPROT entity mention annotations:

10027835 T10 CHEMICAL 1782 1790 esuprone
10027835 T11 CHEMICAL 1796 1804 LU 53439
10027835 T12 CHEMICAL 1944 1952 esuprone
10027835 T13 CHEMICAL 194 203 monoamine
10027835 T14 CHEMICAL 737 747 L-deprenyl
10027835 T15 CHEMICAL 749 759 selegiline
10027835 T16 CHEMICAL 897 905 LU 53439
10027835 T17 CHEMICAL 907 975 3,4-dimethyl-7-(2-isopropyl-1,3, 4-thiadiazol-5-yl)-methoxy-coumarin
10027835 T18 CHEMICAL 1022 1032 L-deprenyl
10027835 T19 CHEMICAL 1095 1103 LU 43839
10027835 T1 CHEMICAL 1216 1231 tranylcypromine
10027835 T20 CHEMICAL 1105 1113 esuprone
10027835 T21 CHEMICAL 1115 1162 7-hydroxy-3, 4-dimethylcoumarin ethanesulfonate
10027835 T22 CHEMICAL 14 23 monoamine
10027835 T23 GENE-N 1202 1205 MAO
10027835 T24 GENE-Y 1397 1402 MAO-B
10027835 T25 GENE-N 275 278 MAO
10027835 T26 GENE-Y 1567 1572 MAO-B
10027835 T27 GENE-Y 1642 1647 MAO-A
10027835 T28 GENE-Y 1671 1676 MAO-A

3. CHEMPROT team registration

In order to participate as a team, you need to register for Track 5 at:

http://www.biocreative.org/events/biocreative-vi/team/



The BioCreative mailing list offers the possibility to discuss-task and workshop related aspects:

https://sourceforge.net/projects/biocreative/lists/biocreative-participant