VEGA Command Line Interface [beta]

INTRODUCTION

The VEGA Command Line Interface is a Java application that provides easy access to all VEGA models for a command line based batch execution. It uses a script as input, where it should be specified what molecule files should be read and what models should be executed, and provides its output as single or multiple plain text files.

USE

The application should be run as a normal Java command line application, with the command:

java -jar VEGA-CLI.jar [-help] [-script SCRIPT_NAME]

If a “-help” parameter is used, a short help and the list of available models are reported.

The script parameter specifies the input script file that should be used for the batch processing.

SCRIPT SYNTAX

The script is a plain text XML file that should be written with the following syntax:

```
<vega>
     <input>
          <type>FileSMI</type>
          <source>[SOURCE_FILE]</source>
     </input>
     <models>
          <model>[MODEL_TAG]</model>
          <model>[MODEL_TAG]</model>
          ...
     </models>
     <output>
          <multipleTXT>[OUTPUT_DIR]</multipleTXT>
          <singleTXT>[OUTPUT_FILE]</singleTXT>
     </output>
</vega>
```

In the Input section, it should be specified the format of the input that will be read in the <type> tag (currently, only the FileSMI identifier for SMILES format is available), and the file that should be read (with its complete path) should be specified in the <source> tag.

In the Models section, the list of models to be run is reported: each <model> tag should contain the identifier for a VEGA model as reported in the list available below.

In the Output section, it is possible to set the type of the output.
The <multipleTXT> tag allows to save multiple text files, one for each calculated model, where all the outputs are reported. Inside the tag, the directory where the files will be saved should be reported (note: the files will have automatic names)
The <singleTXT> tag allows to save a single summary text file, where the output of all calculated models is reported with the main assessment, the prediction, the experimental value (if found) and the ADI value. Inside the tag, the filename (with its complete path) to be used should be reported.

AVAILABLE MODELS

Following, the list of currently available models, with the identifier that should be used in the script file:

Identifier	Model description
muta_caesar	QSAR classification model for Mutagenicity (from CAESAR project) ver.2.1.13
muta_sarpy	QSAR classification model for Mutagenicity (SarPy/IRFMN) ver.1.0.7
muta_iss	Classification model for Mutagenicity (Ames test) based on Benigni-Bossa (Istituto Superiore di Sanità) rule set ver.1.0.2
muta_knn	KNN (Read-Across) model for Mutagenicity (Ames test) ver.1.0.0
carc_caesar	QSAR classification model for Carcinogenicity (from CAESAR project) ver.2.1.9
carc_iss	Classification model for Carcinogenicity based on Benigni-Bossa (Istituto Superiore di Sanità) rule set ver.1.0.2
carc_antares	QSAR classification model for Carcinogenicity (IRFMN/Antares) based on the Antares dataset ver.1.0.0
carc_isscan	QSAR classification model for Carcinogenicity (IRFMN/ISSCAN-CGX) based on the ISSCAN-CGX dataset ver.1.0.0
devtox_caesar	QSAR classification model for Developmental Toxicity (from CAESAR project) ver.2.1.7
devtox_pg	Virtual compound library for Developmental/Reproductive (developed by Procter And Gamble) ver.1.0.0
rba_irfmn	Classification model for Relative Binding Affinity (IRFMN) ver.1.0.1
skin_caesar	QSAR classification model for Skin sensitization (from CAESAR project) ver.2.1.6
fish_irfmn	QSAR classification model for fish acute (LC50) toxicity ver.1.0.2
fish_knn	KNN (Read-Across) model for fish acute (LC50) toxicity ver.1.0.0
fathead_epa	QSAR model for Fathead Minnow LC50 96h (from EPA T.E.S.T software) ver.1.0.7
daphnia_epa	QSAR model for Daphnia Magna LC50 48h (from EPA T.E.S.T software) ver.1.0.7
daphnia_demetra	QSAR model for Daphnia Magna LC50 48h (from EPA T.E.S.T software) ver.1.0.7
bcf_caesar	QSAR regression model for fish BCF (from CAESAR project) ver.2.1.14
bcf_meyla	QSAR regression model for fish BCF (based on Meylan model) ver.1.0.3
bcf_knn	KNN (Read-Across) model for fish BCF ver.1.1.0
rb_irfmn	QSAR classification model for Ready Biodegradability ver.1.0.9
p_sed_irfmn	Persistence (sediment) model ver.1.0.0
p_soil_irmfn	Persistence (soil) model ver.1.0.0
p_water_irfmn	Persistence (water) model ver.1.0.0
logp_meylan	LogP model based on Meylan work (and implemented in EPI Suite software as KowWin). ver.1.1.4
logp_mlogp	LogP model based on MLogP descriptor. ver.1.0.0
logp_alogp	LogP model based on ALogP descriptor. ver.1.0.0