
Bio2BEL KEGG Documentation

Release 0.2.6-dev

Daniel Domingo-Fernández and Charles Tapley Hoyt

Nov 20, 2019

CONTENTS:

1	Citation	3
2	Command Line Interface	5
3	Manager	7
4	Models	9
5	Constants	11
6	Web	13
7	Indices and tables	15
	Python Module Index	17
	Index	19

Bio2BEL KEGG is a package to build KEGG gene sets in the ComPath environment.

Bio2BEL KEGG is a package for enriching BEL networks with KEGG information by wrapping its RESTful API. KEGG. This package downloads pathway information from KEGG's API and store it in template data model relating genes to pathways. Furthermore, it is integrated in the [ComPath environment](#) for pathway database comparison.

CITATION

- Kanehisa, Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K.; KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353-D361 (2017).
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457-D462 (2016).
- Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27-30 (2000).

COMMAND LINE INTERFACE

The command line interface allows you to communicate with the package and perform basic functions such as:

- Populate the database: `python3 -m bio2bel_kegg populate`. By default the database is reset every time is populated. However, another optional parameter “`--reset-db=False`”, allows you to avoid the reset. More logging can be activated by added “`-vv`” or “`-v`” as an argument.
- Drop the database: `python3 -m bio2bel_kegg drop`. More logging can be activated by added “`-vv`” or “`-v`” as an argument.
- Export gene sets as an excel file: `python3 -m bio2bel_kegg export`.

MANAGER

Manager for Bio2BEL KEGG.

```
class bio2bel_kegg.manager.Manager (*args, **kwargs)
    Protein-pathway memberships.

    Doesn't let this class get instantiated if the pathway_model.

    namespace_model
        alias of bio2bel_kegg.models.Pathway

    pathway_model
        alias of bio2bel_kegg.models.Pathway

    protein_model
        alias of bio2bel_kegg.models.Protein

    pathway_model_identifier_column
        KEGG id of the pathway

    get_or_create_pathway (kegg_id: str, name: Optional[str] = None) →
        bio2bel_kegg.models.Pathway
        Get an pathway from the database or creates it.

        Parameters
        • kegg_id – A KEGG pathway identifier
        • name – name of the pathway

    get_protein_by_kegg_id (kegg_id: str) → Optional[bio2bel_kegg.models.Protein]
        Get a protein by its kegg id.

        Parameters kegg_id – A KEGG identifier

    get_protein_by_hgnc_id (hgnc_id: str) → Optional[bio2bel_kegg.models.Protein]
        Get a protein by its hgnc_id.

        Parameters hgnc_id – hgnc_id

    get_protein_by_hgnc_symbol (hgnc_symbol: str) → Optional[bio2bel_kegg.models.Protein]
        Get a protein by its hgnc symbol.

        Parameters hgnc_id – hgnc identifier

    populate (pathways_url=None, protein_pathway_url=None, metadata_existing=False)
        Populate all tables.

    count_pathways () → int
        Count the pathways in the database.
```

list_pathways () → List[bio2bel_kegg.models.Pathway]
List the pathways in the database.

count_proteins () → int
Count the pathways in the database.

summarize () → Mapping[str, int]
Summarize the database.

to_bel () → pybel.struct.graph.BELGraph
Serialize KEGG to BEL.

get_pathway_graph (*kegg_id: str*) → Optional[pybel.struct.graph.BELGraph]
Return a new graph corresponding to the pathway.

Parameters *kegg_id* – A KEGG pathway identifier (prefixed by “path:”)

enrich_kegg_pathway (*graph: pybel.struct.graph.BELGraph*) → None
Enrich all proteins belonging to KEGG pathway nodes in the graph.

enrich_kegg_protein (*graph: pybel.struct.graph.BELGraph*) → None
Enrich all KEGG pathways associated with proteins in the graph.

MODELS

KEGG database models.

class `bio2bel_kegg.models.Pathway` (***kwargs*)

Pathway Table.

A simple constructor that allows initialization from *kwargs*.

Sets attributes on the constructed instance using the names and values in *kwargs*.

Only keys that are present as attributes of the instance's class are allowed. These could be, for example, any mapped columns or relationships.

kegg_id

KEGG id of the pathway

name

pathway name

serialize_to_pathway_node () → `pybel.dsl.node_classes.BiologicalProcess`

Serialize to PyBEL node data dictionary.

get_gene_set () → `Set[bio2bel_kegg.models.Protein]`

Return the genes associated with the pathway (gene set).

Note this function restricts to HGNC symbols genes.

property resource_id

Return kegg identifier.

property url

Return url pointing to kegg pathway.

class `bio2bel_kegg.models.Protein` (***kwargs*)

Genes Table.

A simple constructor that allows initialization from *kwargs*.

Sets attributes on the constructed instance using the names and values in *kwargs*.

Only keys that are present as attributes of the instance's class are allowed. These could be, for example, any mapped columns or relationships.

kegg_id

KEGG id of the protein

uniprot_id

uniprot id of the protein (there could be more than one)

hgnc_id

hgnc id of the protein

hgnc_symbol

hgnc symbol of the protein

to_pybel () → pybel.dsl.node_classes.Protein

Serialize to PyBEL node data dictionary.

get_uniprot_ids () → Optional[List[str]]

Return a list of uniprot ids.

get_pathways_ids () → Set[str]

Return the pathways associated with the protein.

CONSTANTS

This module contains all the constants used in Bio2bel Kegg project.

This module contains the flask application to visualize the db.

INDICES AND TABLES

- `genindex`
- `modindex`
- `search`

PYTHON MODULE INDEX

b

- `bio2bel_kegg`, ??
- `bio2bel_kegg.constants`, [11](#)
- `bio2bel_kegg.manager`, [7](#)
- `bio2bel_kegg.models`, [9](#)
- `bio2bel_kegg.web`, [13](#)

INDEX

B

`bio2bel_kegg` (module), 1
`bio2bel_kegg.constants` (module), 11
`bio2bel_kegg.manager` (module), 7
`bio2bel_kegg.models` (module), 9
`bio2bel_kegg.web` (module), 13

C

`count_pathways()` (*bio2bel_kegg.manager.Manager* method), 7
`count_proteins()` (*bio2bel_kegg.manager.Manager* method), 8

E

`enrich_kegg_pathway()`
(*bio2bel_kegg.manager.Manager* method), 8
`enrich_kegg_protein()`
(*bio2bel_kegg.manager.Manager* method), 8

G

`get_gene_set()` (*bio2bel_kegg.models.Pathway* method), 9
`get_or_create_pathway()`
(*bio2bel_kegg.manager.Manager* method), 7
`get_pathway_graph()`
(*bio2bel_kegg.manager.Manager* method), 8
`get_pathways_ids()` (*bio2bel_kegg.models.Protein* method), 10
`get_protein_by_hgnc_id()`
(*bio2bel_kegg.manager.Manager* method), 7
`get_protein_by_hgnc_symbol()`
(*bio2bel_kegg.manager.Manager* method), 7
`get_protein_by_kegg_id()`
(*bio2bel_kegg.manager.Manager* method), 7

`get_uniprot_ids()` (*bio2bel_kegg.models.Protein* method), 10

H

`hgnc_id` (*bio2bel_kegg.models.Protein* attribute), 9
`hgnc_symbol` (*bio2bel_kegg.models.Protein* attribute), 9

K

`kegg_id` (*bio2bel_kegg.models.Pathway* attribute), 9
`kegg_id` (*bio2bel_kegg.models.Protein* attribute), 9

L

`list_pathways()` (*bio2bel_kegg.manager.Manager* method), 7

M

`Manager` (class in *bio2bel_kegg.manager*), 7

N

`name` (*bio2bel_kegg.models.Pathway* attribute), 9
`namespace_model` (*bio2bel_kegg.manager.Manager* attribute), 7

P

`Pathway` (class in *bio2bel_kegg.models*), 9
`pathway_model` (*bio2bel_kegg.manager.Manager* attribute), 7
`pathway_model_identifier_column`
(*bio2bel_kegg.manager.Manager* attribute), 7
`populate()` (*bio2bel_kegg.manager.Manager* method), 7
`Protein` (class in *bio2bel_kegg.models*), 9
`protein_model` (*bio2bel_kegg.manager.Manager* attribute), 7

R

`resource_id()` (*bio2bel_kegg.models.Pathway* property), 9

S

`serialize_to_pathway_node()`
 (*bio2bel_kegg.models.Pathway* *method*),
 9
`summarize()` (*bio2bel_kegg.manager.Manager*
 method), 8

T

`to_bel()` (*bio2bel_kegg.manager.Manager* *method*), 8
`to_pybel()` (*bio2bel_kegg.models.Protein* *method*),
 10

U

`uniprot_id` (*bio2bel_kegg.models.Protein* *attribute*),
 9
`url()` (*bio2bel_kegg.models.Pathway* *property*), 9