

gmod_load_cvterms.pl is a perl script for loading and more importantly, updating controlled vocabulary and ontology terms in the cvterm table. This script was contributed by Naama Menda and the developers at the Sol Genomics Network (SGN) lead by Lukas Mueller.

Where to find it

gmod 1.0

In the 1.0 release of gmod, the script is called load_cvterms.pl and is not installed. It can be found in the distribution folder in the bin/cxgn directory.

gmod 1.1

In the 1.1 release, gmod_load_cvterms.pl is installed with other scripts with the distribution and will typically go in /usr/bin or /usr/local/bin.

Command line options

- -H hostname for database [required if -p isn't used]
- -D database name [required if -p isn't used]
- -p password (if you need to provide a password to connect to your db)
- -r username (if you need to provide a username to connect to your database)
- -d driver name (e.g. ?Pg? for postgres). Driver name can be provided in gmod_config
- -g GMOD database profile name (can provide host, DB name, password, username, and driver)
Default:?default?
- -s database name for linking (must be in db table)
- -n controlled vocabulary name (e.g ?biological_process?). optional. If not given, terms of all namespaces related with database name will be handled.
- -F File format. Can be obo or go_flat and others supported by Bio::OntologyIO. Default: obo
- -u update all the terms. Without -u, the terms in the database won't be updated to the contents of the file, in terms of definitions, etc. New terms will still be added.
- -v verbose output
- -o outfile for writing errors and verbose messages (optional)
- -t trial mode. Don't perform any store operations at all. (trial mode cannot test inserting associated data for

new terms)

The script parses the ontology in the file and the corresponding ontology in the database, if present. It compares which terms are new in the file compared to the database and inserts them, and compares all the relationships that are new and inserts them. It removes the relationships that were not specified in the file from the database. It never removes a term entry from the database.

Terms that are in the database but not in the file are set to `is_obsolete=1`. All the terms that are present in the database are updated (if using `-u` option) to reflect the term definitions that are in the file. New terms that are in the file but not in the database are stored.