GenePattern

Join.Files Documentation

Description: Uses R's merge() function to perform a database style join on two files.

Author: Wrapped as a module by Anthony S. Castanza, Mesirov Lab, UCSD School of Medicine.

Contact: genepattern.org/help

Summary: Simple file joining options using the R base merge functions.

File [X] Parameters (options are the same for both Input File 1 and File 2):

Name	Description
File [X]	File 1 or File 2 to perform the join operation on.
File [X] Delimiter	Each file option supports tab delimited (.txt) or comma delimited (.csv) files. This need not be the same for both files.
File [X] Header	If the first line of the respective file is a header line (TRUE) or the file does not have a header (FALSE).
File [X] Column	Which column of the input file to join on the basis of. By default the first column is used.
File [X] Unmatched	What to do with unmatched lines from the specific file. If kept, unmatched lines will be filled with "NA" for empty cells. By default, options for both File 1 and File 2 are set to "Keep", this is equivalent to a "full outer join". Setting both File 1 and File 2 to "Discard" is equivalent to an "inner join". File 1: Keep; File 2: Discard = Left outer join. File 1: Discard; File 2: Keep = Right outer join.

Output:

Name	Description
Output Type	Output File type (txt = tab delimited .txt, csv = comma separated .csv)

Output file will be named <File.1.basename>.<File.2.basename>.merged.<Output.Type>
Where "File.1.basename" is the base name without the file extension of the file provided to the File 1 parameter, "File.2.basename" is the base name without the file extension of the file provided to the File 2 parameter, and "Output.Type" is the option selected from the Output Type Parameter.

Module Language: R
Source Repository: https://github.com/genepattern/Join.Files/releases/tag/v1
Docker image: jupyter/datascience-notebook:r-3.6.3

Version	Comment
1	Initial release.