

Duplicate Node Remover:

How to Use

1. Unzip the tool and open the folder.
2. Open the appsettings.json file and configure the tool as needed.
 1. DbConnectionConfigurationData is the connection string to access the database where the user's data is kept. By default, this is the connection string that points to localhost:5432 with the default username and password. This should be correct for most local setups.
 2. AccountID is the account which contains the duplicate data that needs to be removed.
 3. FirstRootNode and SecondRootNode are optional values.
 - If both are set to null, then the tool will run from the account's root node and cover the entire structure.
 - If just one of them is set to the ID of a node, then the tool will start at that part of the tree and prune all of its children.
 - If both are set to the ID of a node, then the tool will bring all of the contents of the SecondRootNode into the FirstRootNode, removing duplicates along the way. The second node will still be left in the tree structure, but it will be empty.
 4. WriteToFile will tell the tool to write all of the nodes to be edited to file. There will be two files produced: moveRename.xml and remove.xml. This option is available because the tool may run for a very long time on large datasets, and this allows you to break the process into gathering all of the nodes to update and executing changes on them.
 5. LoadFromFile will circumvent the FirstRootNode and SecondRootNode and instead load all of the files from the moveRename.xml and remove.xml files.
 6. DoPruningStep determines whether the tool should attempt to make all of the changes or not. If you want to break up the process of cleaning a user's data you could do it in two passes with the following configurations:
 1. WriteToFile is true, LoadFromFile is false, DoPruningStep is false.
This will do the work of crawling through the user's data, extract the nodes to be updated, write them to file, and then terminate.
 2. WriteToFile is false, LoadFromFile is true, DoPruningStep is true. This will load the needed changes from file without having to crawl through the data again, then execute those changes.
3. **Back up the user's data before running the tool.**
4. Run DuplicateNodeRemover.exe. It will proceed in two main steps:
 1. Finding all of the nodes to update. This might take considerable time depending on the size of the data to be pruned. Eventually it will report how many nodes are set to be removed, moved, and renamed-then-moved.

2. Executing all of the changes. If there are more than a thousand nodes to update, you will have the option to periodically decide if the tool should continue or if it should save its progress and resume later.
 5. Done!
-

Revision #3

Created 15 August 2023 17:22:33 by Abe Austin

Updated 29 September 2023 22:29:33 by Abe Austin