

## **Pick 2D Descriptors**

Shaillay Kumar Dogra  
Scientific Editor – QSAR World  
[editor@qsarworld.com](mailto:editor@qsarworld.com)

### **Notes:**

1. This Jython script works in Sarchitect Designer version 2.2
2. Learn about Sarchitect Designer – <http://www.strandls.com/sarchitect/index.html>
3. Get Sarchitect – <http://www.strandls.com/sarchitect/freetrial.php>

*The actual script follows this discussion. It is also accessible directly from the webpage in .py format.*

### **Discussion:**

A small but useful script that allows one to pick constitutional and topological descriptors (non 3D descriptors) from a set wherein 'all' descriptors have been computed. This is of help if, say, one wants to build models on 2D descriptors only, then compare against models obtained on the full set (or the 3D descriptors only). If comparable models can be obtained across different descriptor sets, then models obtained with 2D descriptors are preferable since they are intuitive to understand and require lesser computation (no need for structure optimization etc.).

Running this script will create a subset with only the 2D descriptors. Input is a superset of descriptors that also contains these 2D descriptors, in particular, "numAtoms" and "Lop" as these are used in the script to mark the start and end of 2D descriptor columns. If these are not found, a message gets prompted to that effect and the subset containing 2D descriptors is not created.

Another sister script does the complementary. It picks the 3D descriptors and creates a subset.

### **Cite this as:**

Dogra, Shaillay K., "Script for picking 2D descriptors" from QSARWorld – free online resource for QSAR modeling.  
<http://www.qsarworld.com/virtual-workshop.php>

---

```

##
##
## sarchitect designer 2.2 script to pick 2-D descriptors from
## a dataset that contains the full set of descriptors and
## carry them into a child-dataset labelled as "2D-descriptors"
##
## Note - "Marked" columns are also carried on into the child-dataset.
## This script doesn't control that.
##
##
## Author: Shaillay Kumar Dogra
## Date: July 06, 2007
## editor@qsarworld.com
##
##
## There could be various ways of picking 2D descriptors, like:
##
## a) based on names; will have to type out all the 444 2-D descriptor ## names and handle
## exceptions if some not found (say, some subset ## was not computed by ticking it off)
##
## b) look for first 2D descriptor ("numAtoms"), find its index; find ## index of last 2D descriptor
## ("Lop") and thus get handle of all
## the 2D descriptor indices assuming them to
## be in b/w these two - "numAtoms" & "Lop"
##
## (ticking on/off the options that compute "numAtoms" & "Lop" will
## stop the script)
##
## Implementing latter option(b) in this script
##
## Will give an error message if start-point ('numAtoms') and/or stop-## point ('Lop') not found.
## Will not create a child dataset in such cases and give a message to ## that effect.
##
##
##

import script
from script.dataset import *
from script.algorithm import *
from script.project import *
from script.omega import createComponent, showDialog
from javax.swing import *
from math import *

node = getActiveProject().getActiveDatasetNode()
dataset = node.getDataset()

startcol = dataset.getColumn('numAtoms')
startindex = dataset.index(startcol)
#print startcol, startindex
if (startindex == -1): # Happens if 'numAtoms' not found
    parent=script.tool.getTool().getFrame()
    msg = "Start point 'numAtoms' not found!"
    JOptionPane.showMessageDialog(parent,msg,"STATUS!",JOptionPane.INFORMATIO
N_MESSAGE)

```

```
stopcol = dataset.getColumn('Lop')
stopindex = dataset.index(stopcol)
#print stopcol, stopindex
if (stopindex == -1): # Happens if 'Lop' not found
    parent=script.tool.getTool().getFrame()
    mesg = "Stop point 'Lop' not found!"
    JOptionPane.showMessageDialog(parent,mesg,"STATUS!",JOptionPane.INFORMATIO
N_MESSAGE)

colindex = [ ]

i = startindex
if (startindex != -1 and stopindex != -1): # Happens if start-point "numAtoms" and stop-point "Lop"
was not found
    while i < (stopindex+1):
        colindex.append(i)
        i = i + 1

rowIndices=[i for i in range(dataset.getRowCount())]
colIndices= colindex
if (startindex != -1 and stopindex != -1): # Happens if start-point "numAtoms" and stop-point "Lop"
was not found
    node.addChildDatasetNode("2D-Descriptors", rowIndices, colIndices)
    script.view.Table(rowHeight=80).show()
else:
    parent=script.tool.getTool().getFrame()
    mesg = "No child-dataset created!"
    JOptionPane.showMessageDialog(parent,mesg,"STATUS!",JOptionPane.INFORMATIO
N_MESSAGE)

##
## END
##
```

---

End of Document