

@kochi0603

2017.07.08

KNIMEのBATCH MODEを使ってみた

最近KNIMEをよく使ってます。

- ▶ 起動は遅いけど、**PP**よりもワークフロー(=プロトコル)を作りやすい。
- ▶ 例) **PP**って列名を手書きすることが多いけど、KNIMEはほとんどない。
- ▶ 今日は、いい点、悪い点の紹介と、batch mode(コマンドライン)でどこまで使えるかを紹介します。

2017.07.08

いい点

タダ

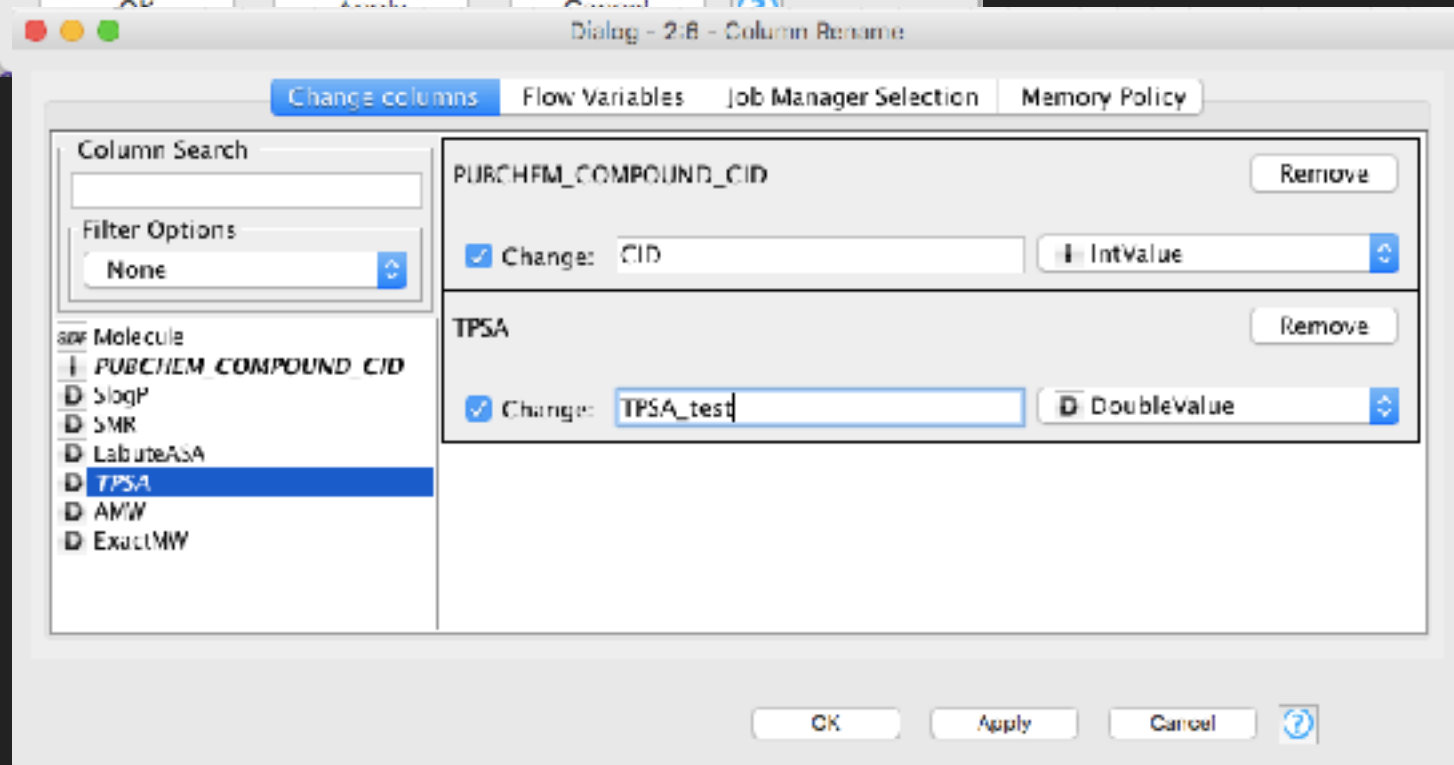
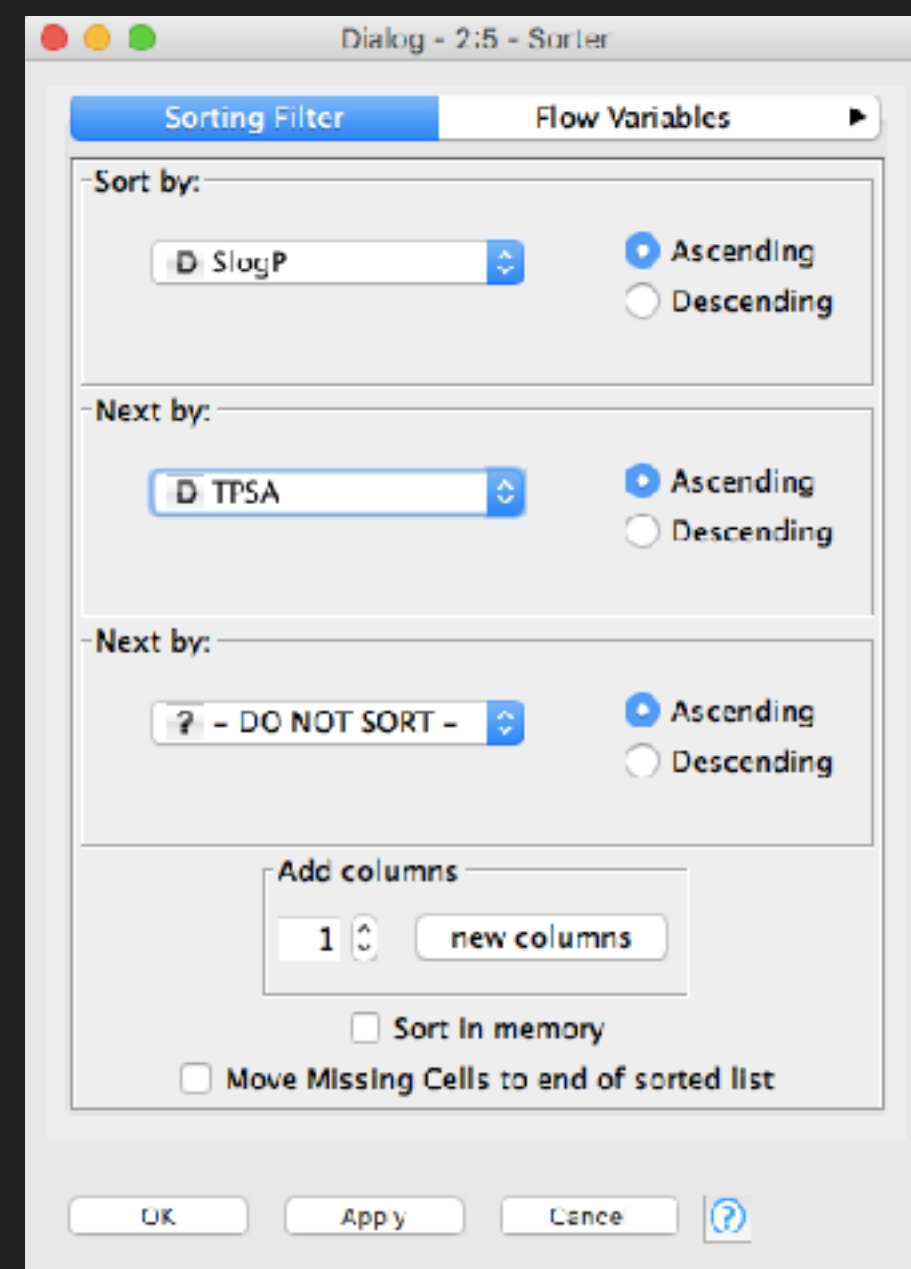
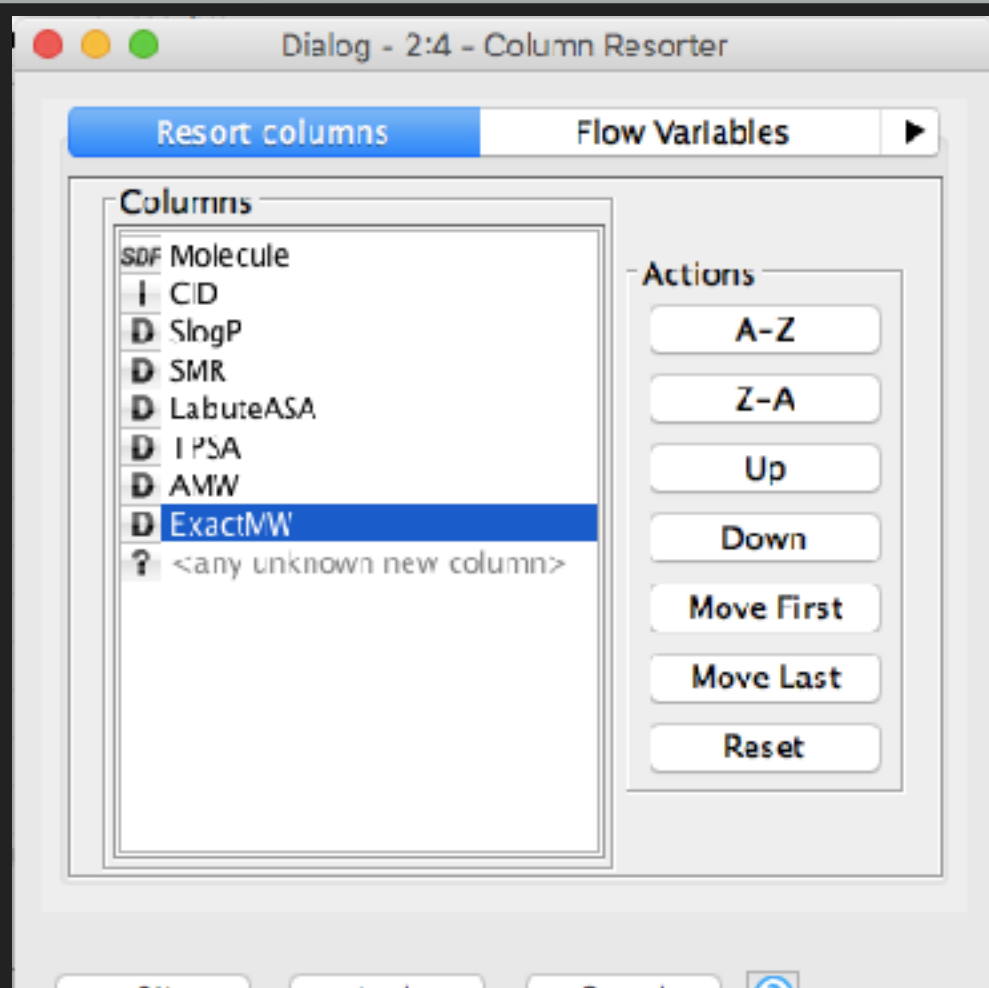
2017.07.08

直前のノードが保持するテーブルの列がリストアップされるのでマウス操作だけでOK

The screenshot displays the KNIME Analytics Platform interface. On the left, the 'KNIME Explorer' and 'Node Repository' panels are visible. The main workspace shows a workflow with two nodes: 'SDF Reader' (Node 1) and 'RDKit Descriptor Calculation' (Node 2). A dialog box titled 'Dialog - 2:2 - RDKit Descriptor Calculation' is open in the foreground. The dialog has four tabs: 'Options', 'Flow Variables', 'Job Manager Selection', and 'Memory Policy'. The 'Options' tab is active, showing the 'RDKit Mol column' set to 'SDF Molecule'. Below this, there are three sections: 'Exclude', 'Select', and 'Include'. The 'Exclude' section has a red border and contains a list of descriptors: NumLipinskiHBA, NumLipinskiHBD, NumRotatableBond, NumHBD, NumHBA, and NumAmideBonds. The 'Select' section has a grey border and contains buttons for 'Add >>', 'Add All >>', '<< Remove', and '<< Remove All'. The 'Include' section has a green border and contains a list of descriptors: SlogP, SMR, LabuteASA, TPSA, AMW, and ExactMW. At the bottom of the dialog are 'OK', 'Apply', and 'Cancel' buttons.

2017.07.08

列の順番を変えるときか、
列名変更とか、
データのソートとか。

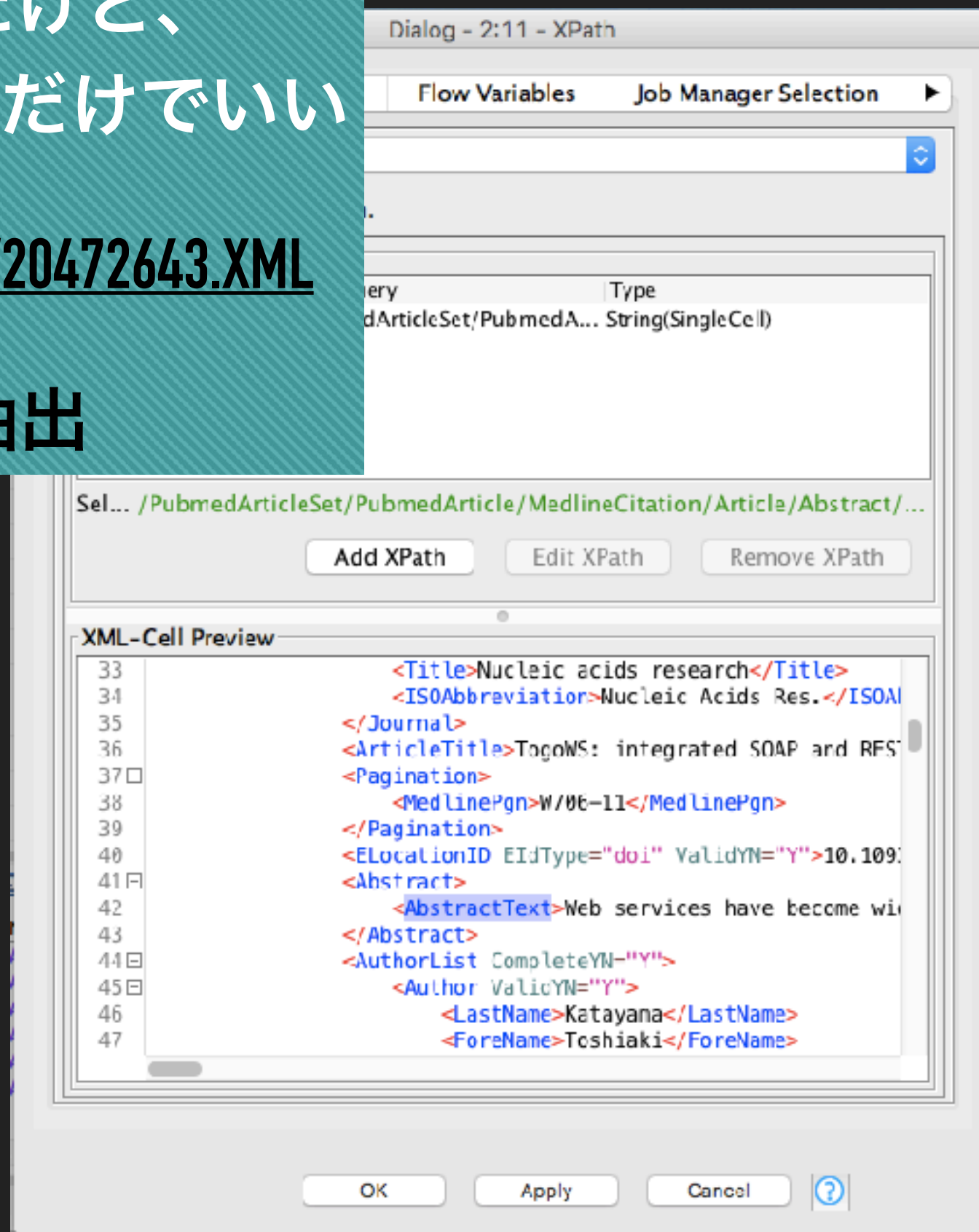
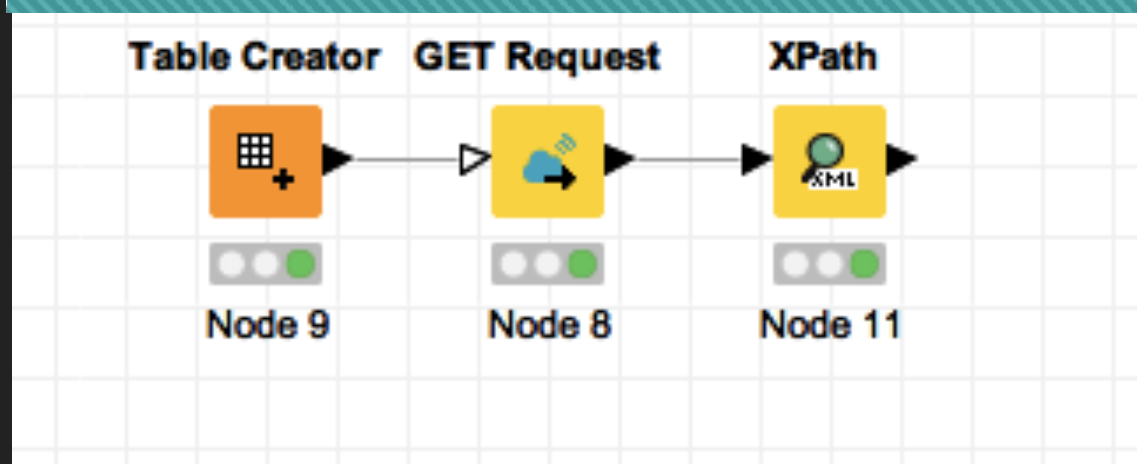


2017.07.08

XMLを扱う機会は減ったけど、
XPathの指定もマウスで選ぶだけでいい

[HTTP://TOGOWS.ORG/ENTRY/PUBMED/20472643.XML](http://togows.org/entry/pubmed/20472643.xml)

からXML→ABSTRACT抽出



スクリプト言語

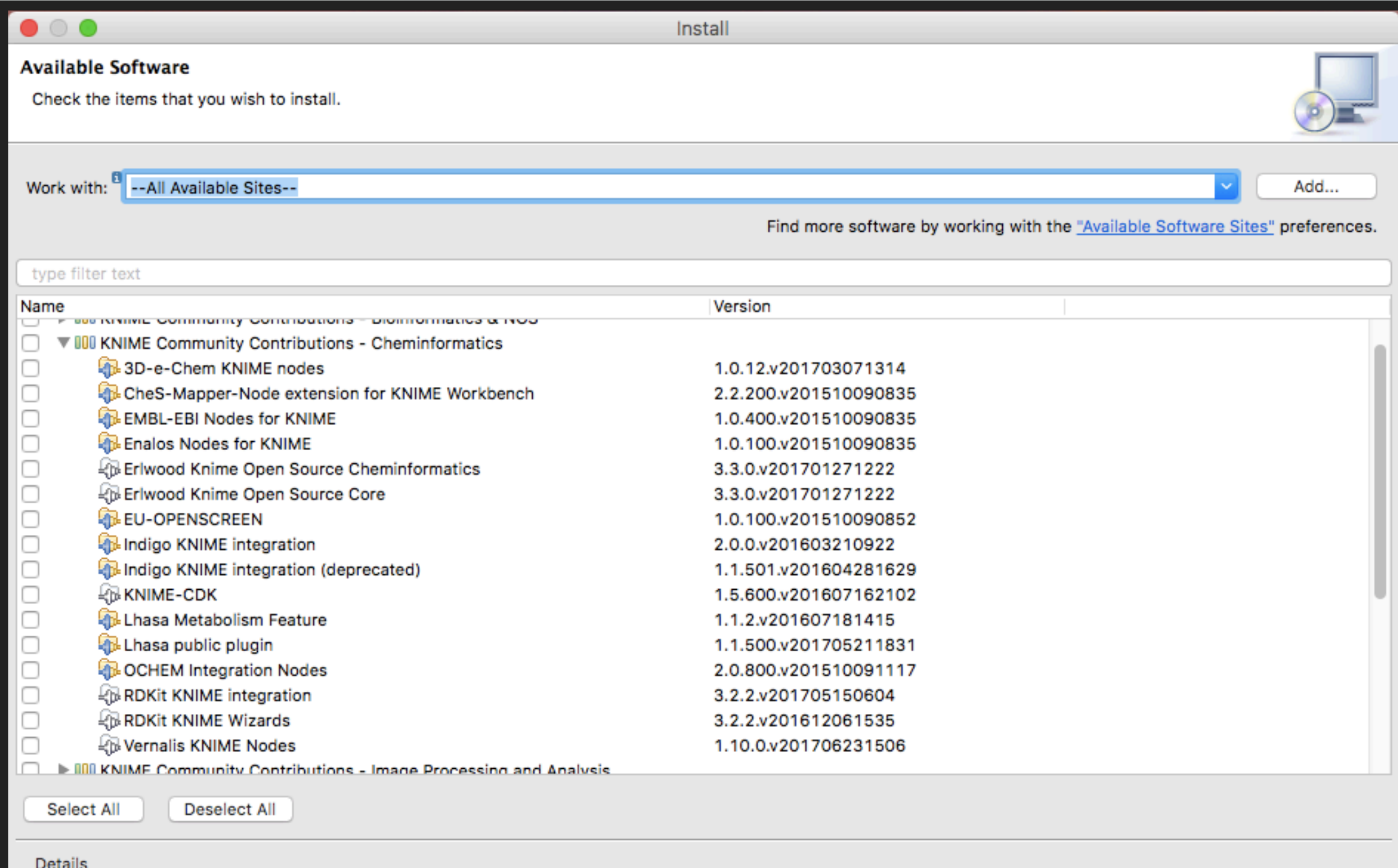
Java KNIMEはJavaなの 入力画面が使いにくい。
で動作は早い 型変換とか面倒。

Python pandasが大好きな 1行ずつ処理されず、全件
(2.7) 人には良い 一括処理しかできない

Perl 比較的書きやすい 絶賛するほどでもない

PPのパイロットスクリプトのような存在がない
から、なんとか標準ノードだけを組み合わせで処
理させようとする。(いい意味で)

2017.07.08



悪い点

- ▶ 起動が遅い。
 - ▶ extensionを入れれば入れるほど遅くなるので、all free extensionsは避け、まっさらでインストールして、必要なものだけ選ぶ。
- ▶ freeなAnalytics版にはPPのWebPort機能がない(有償のKNIME Server版にはある)
- ▶ 複数人でシェアするにはxmlでimport,exportするしかない
- ▶ 国内にコミュニティがない。教えてくれる人がいない

KNIMEは誰向けのソフトか？

- ▶ ITじゃない人でちょっとしたデータ処理を自力でやりたい人？
 - ▶ 初心者にKNIMEの基本機能を2h教えると、XLSやSDFをマージ等できるようになった
 - ▶ PPやWebPort化するほどでない処理は結構ある
- ▶ ITな人は遅さやスクリプトの不自由さに満足できない??

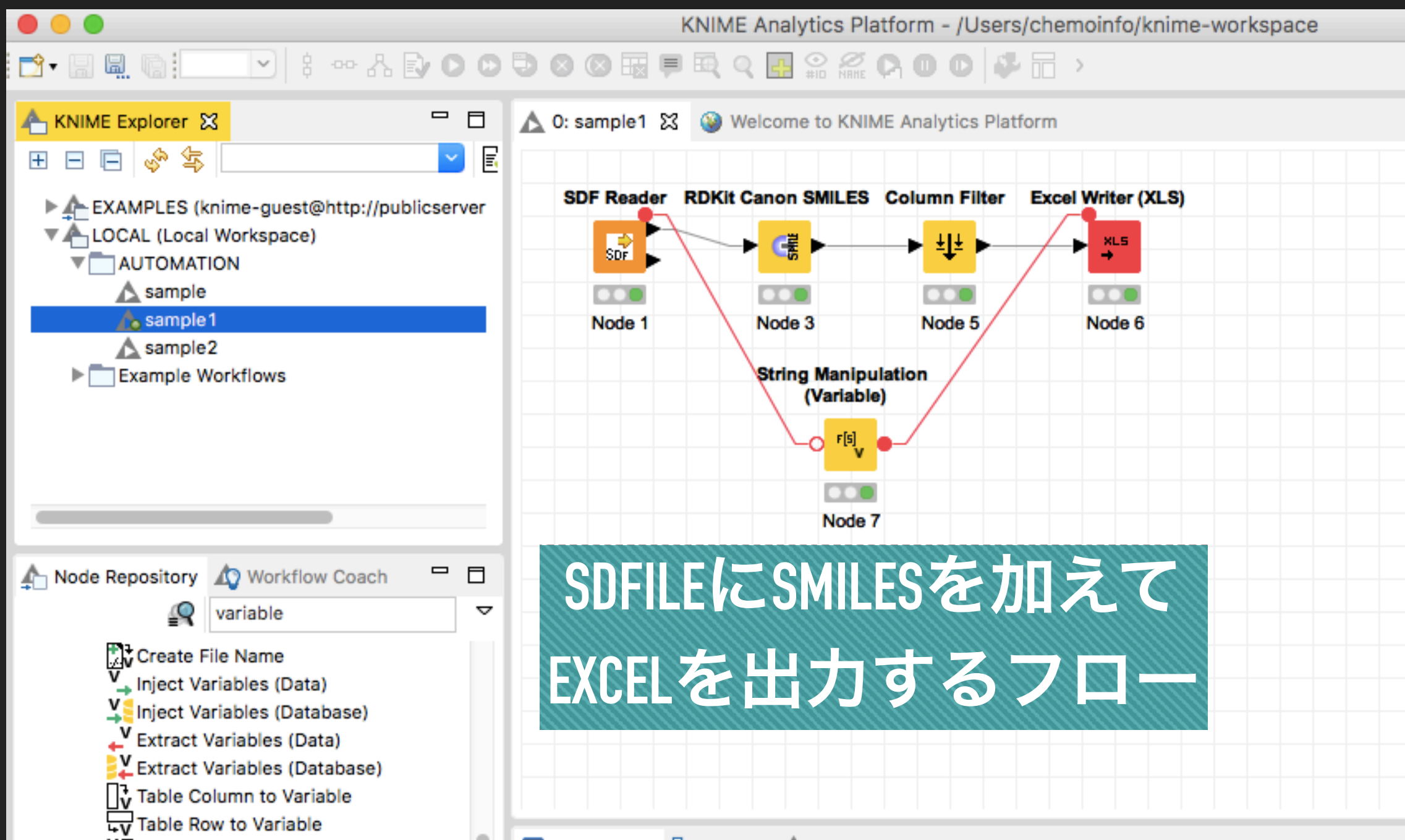
BATCH MODEを
使ってみる

BATCH MODE(コマンドライン)で実行できるよ

- ▶ コンソールで、↓のように実行
 - ▶ `Knime -consoleLog -nosplash -reset -noexit -application org.knime.product.KNIME_BATCH_APPLICATION -workflowDir="/Users/chemoinfo/knime-workspace/AUTOMATION/sample1"`
 - ▶ 対話型ではないので、指定したワークフローを実行して終了
 - ▶ 処理内容を都度変更したい時には、FlowVariableを使う

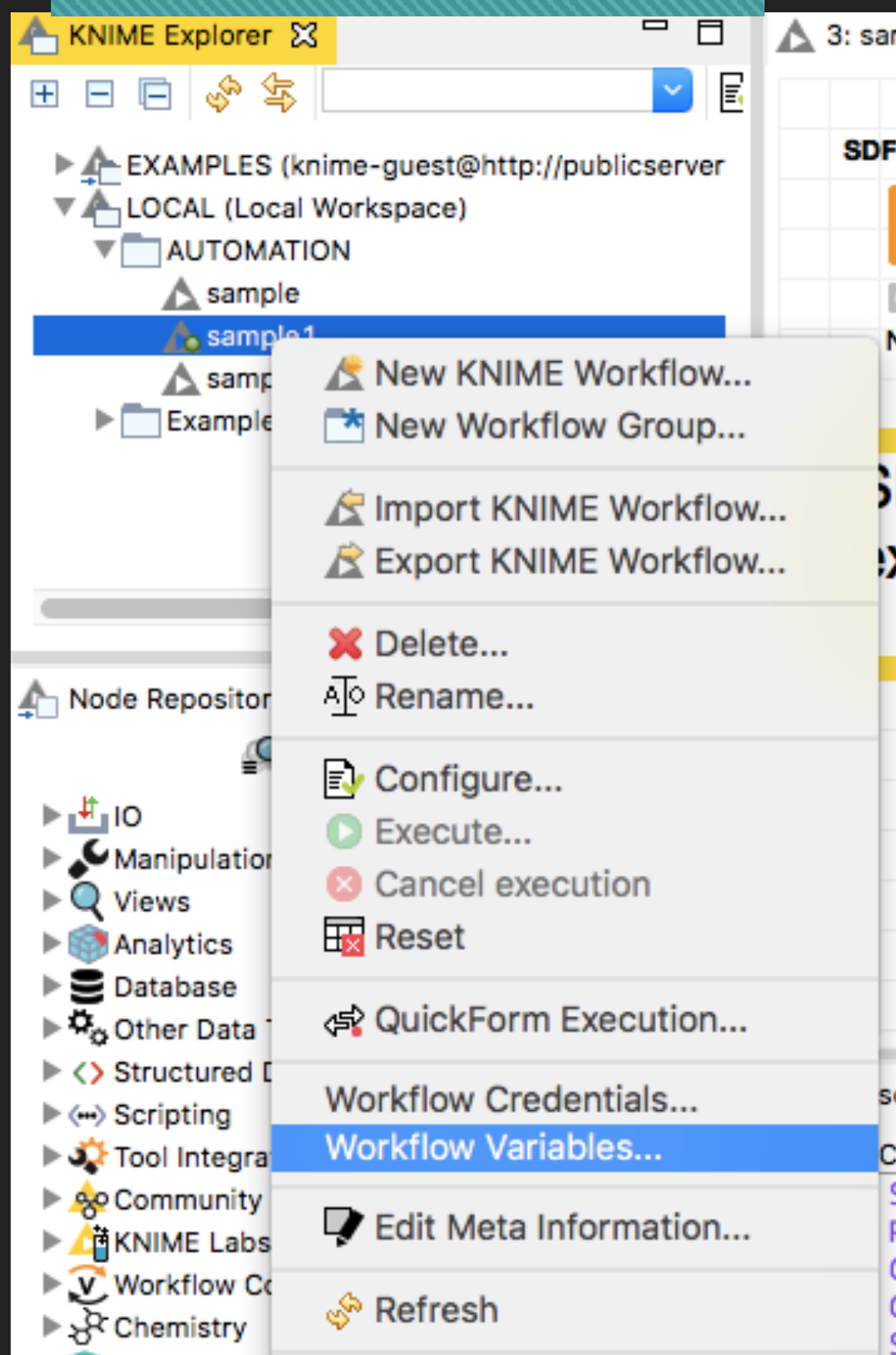
2017.07.08

こんな感じで作ります。

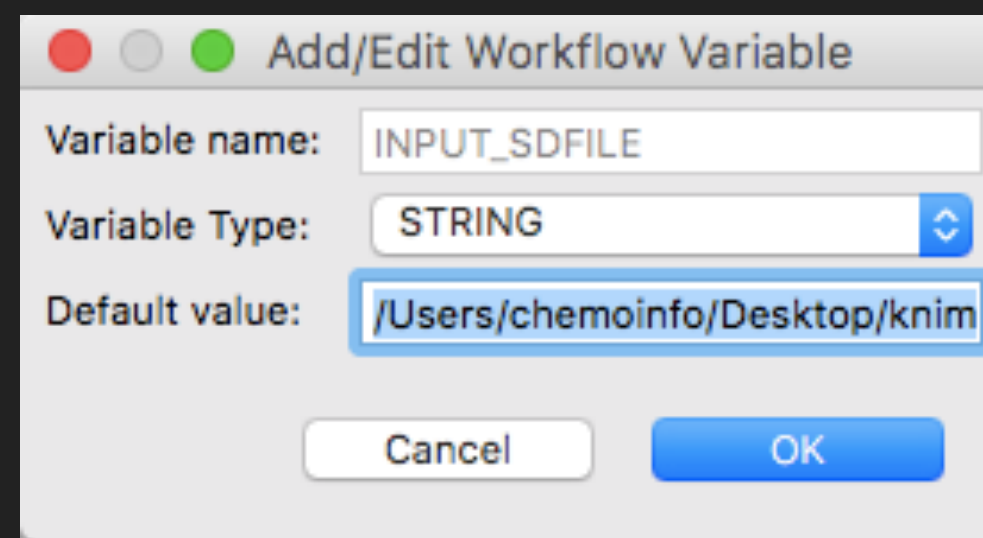
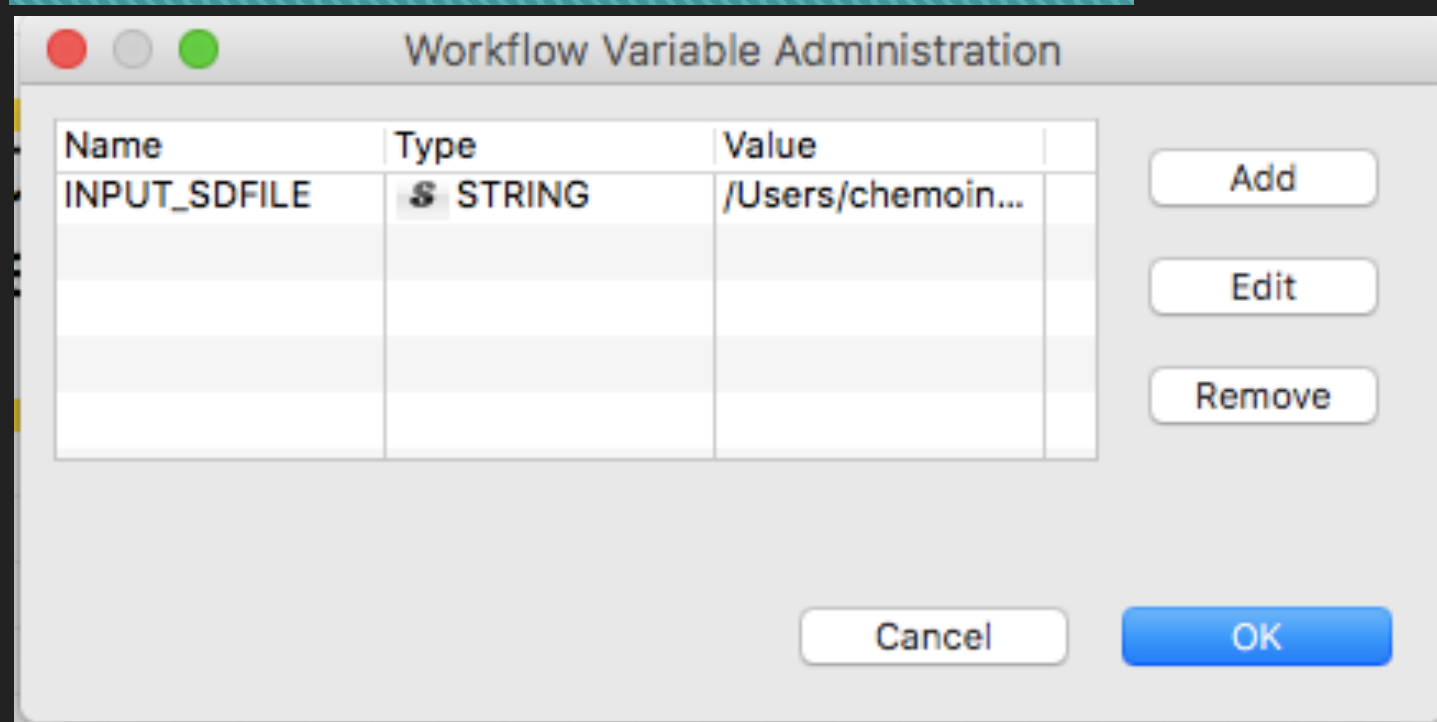


2017.07.08

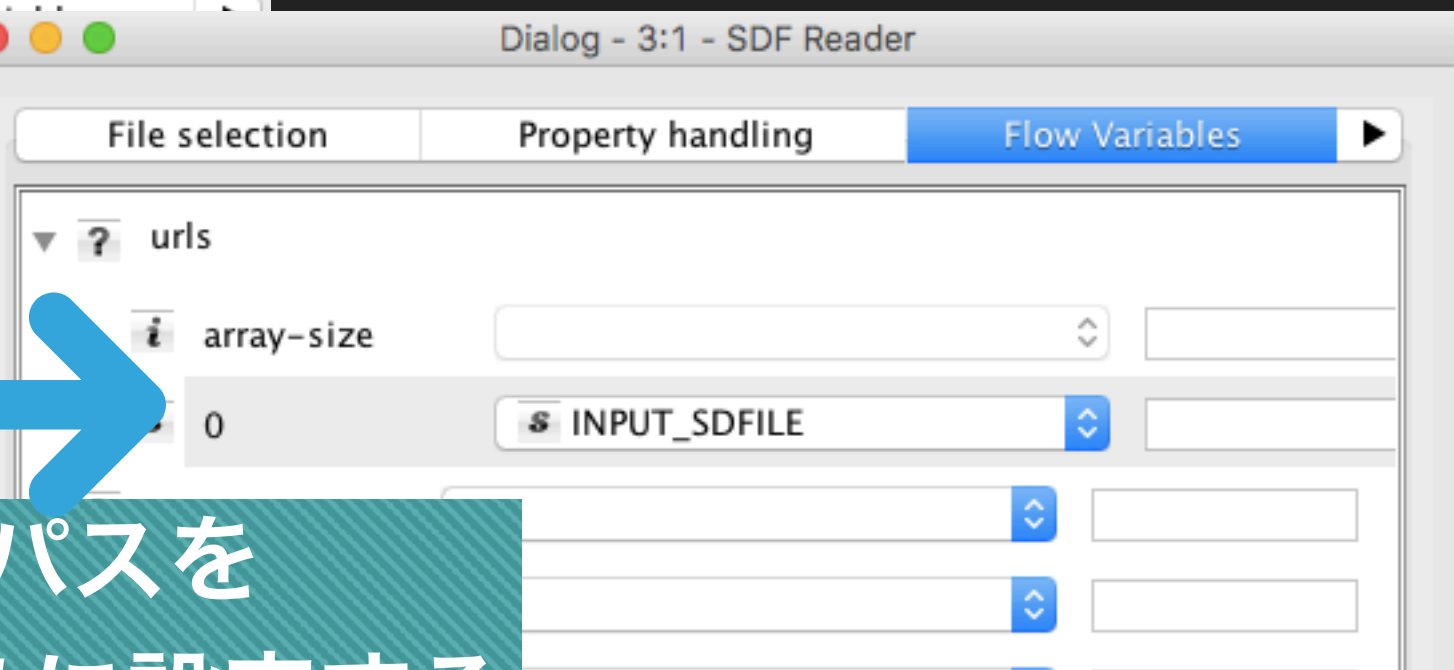
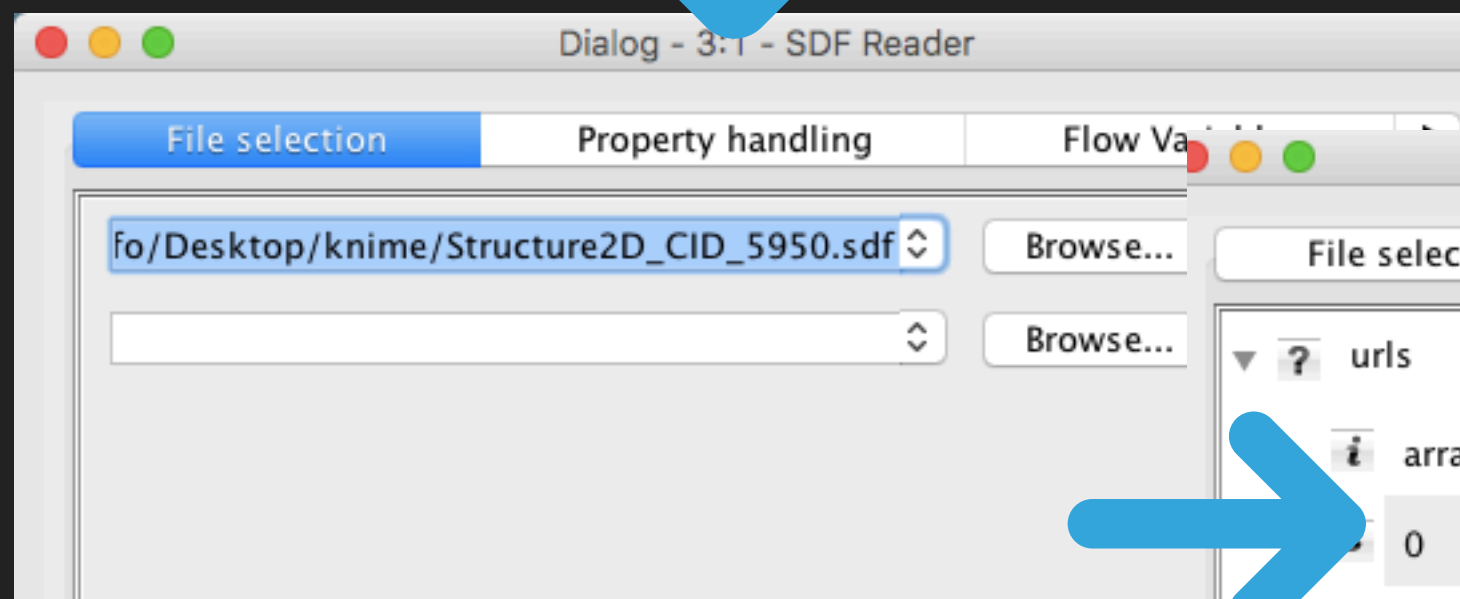
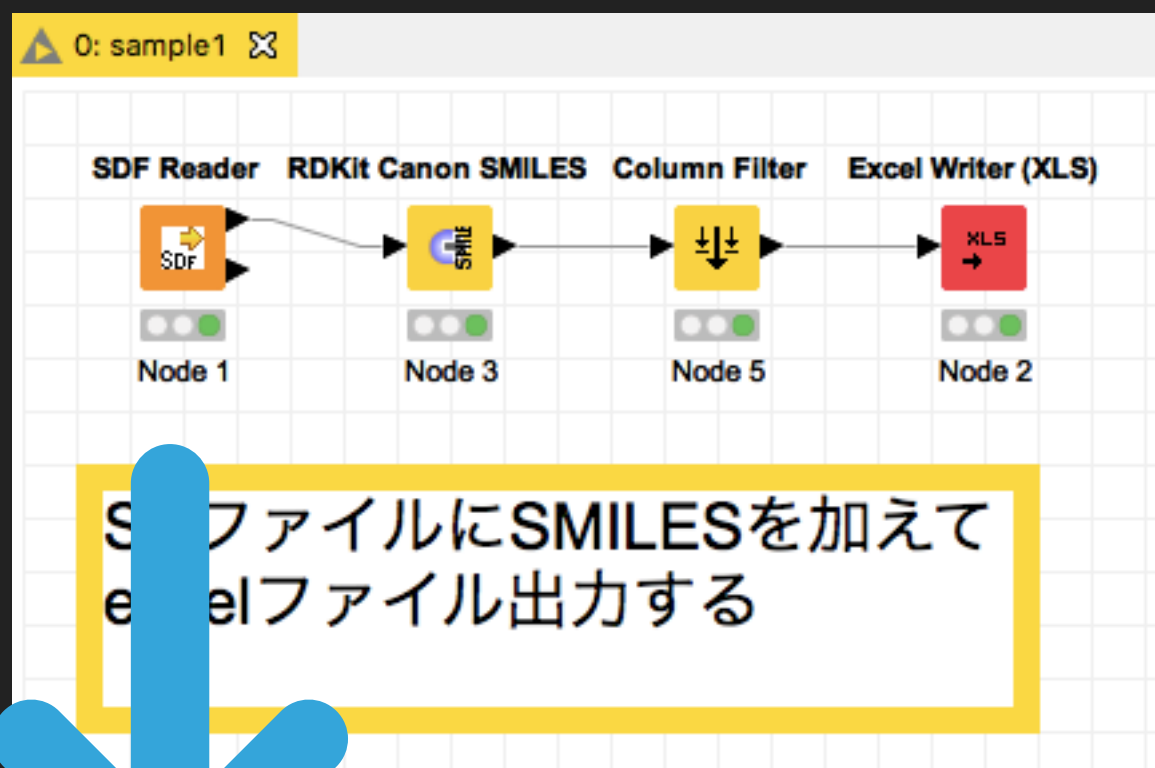
作成したWORKFLOW に



変数INPUT_SDFILEを追加する

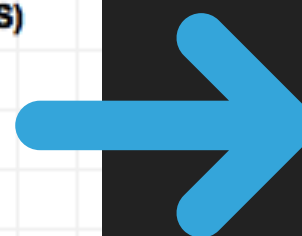
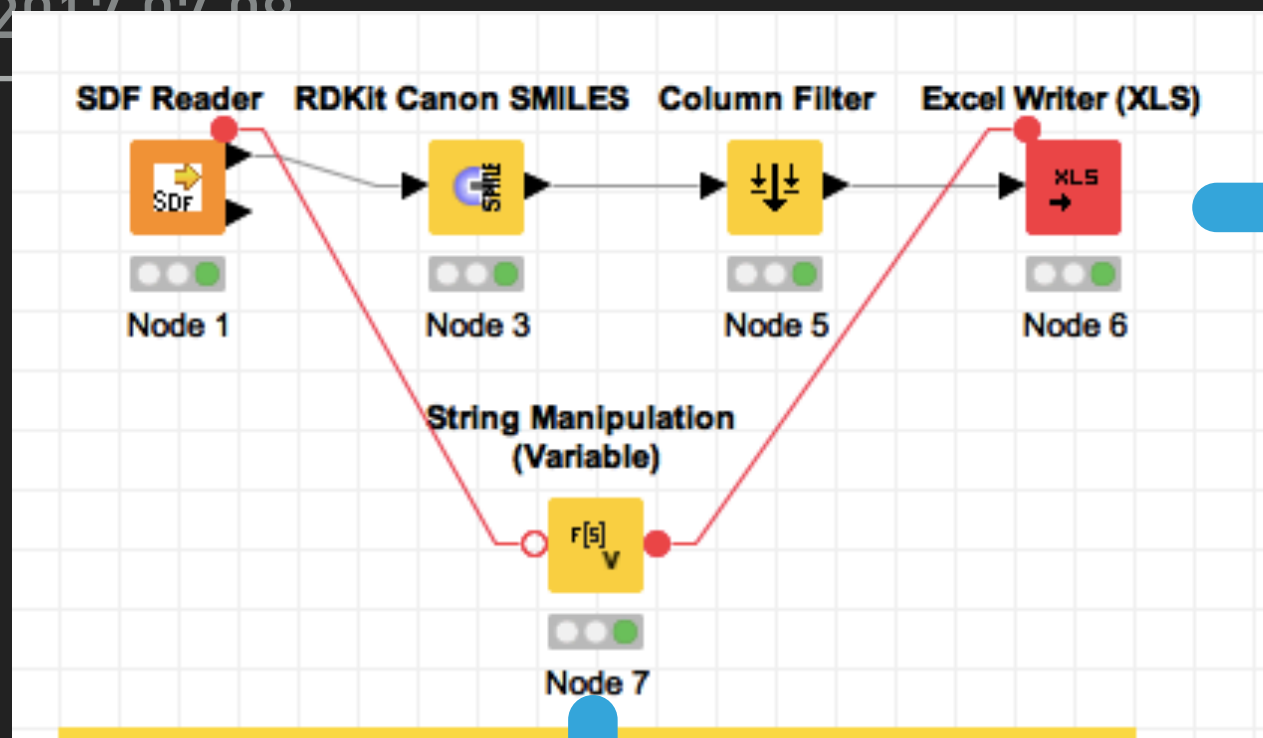


2017.07.08



入力するSDファイルのパスを
FLOW VARIABLEから読み込むように設定する

2017-07-08



Dialog - C:\G - Excel Writer (XLS)

Settings | Flow Variables | Job Manager Selection | Me

xCellOffset		
yCellOffset		
writeColHdr		
writeRowHdr		
filename	OUTPUT_FILE	
sheetname		
missingPattern		
overwrite_ok		
open_file		
file_must_exist		
do_not_overw...		
landscape		
autosize		
paper_size		
write_missing...		
evaluate_for...		
xlsxwriter?		

OK



String Manipulation | Flow Variables | Job Manager Selection | Memory Policy

Flow Variable List

- INPUT_SDFILE
- ltime.workspace

Category: All

Function

- capitalize(str)
- capitalize(str, chars)
- compare(str1, str2)
- count(str, toCount)
- count(str, toCount, modifiers)
- countChars(str, chars)
- countChars(str, chars, modifiers)
- indexOf(str, toSearch)
- indexOf(str, toSearch, modifiers)

Expression

```
join(, ${INPUT_SDFILE}$, ".xlsx" )
```

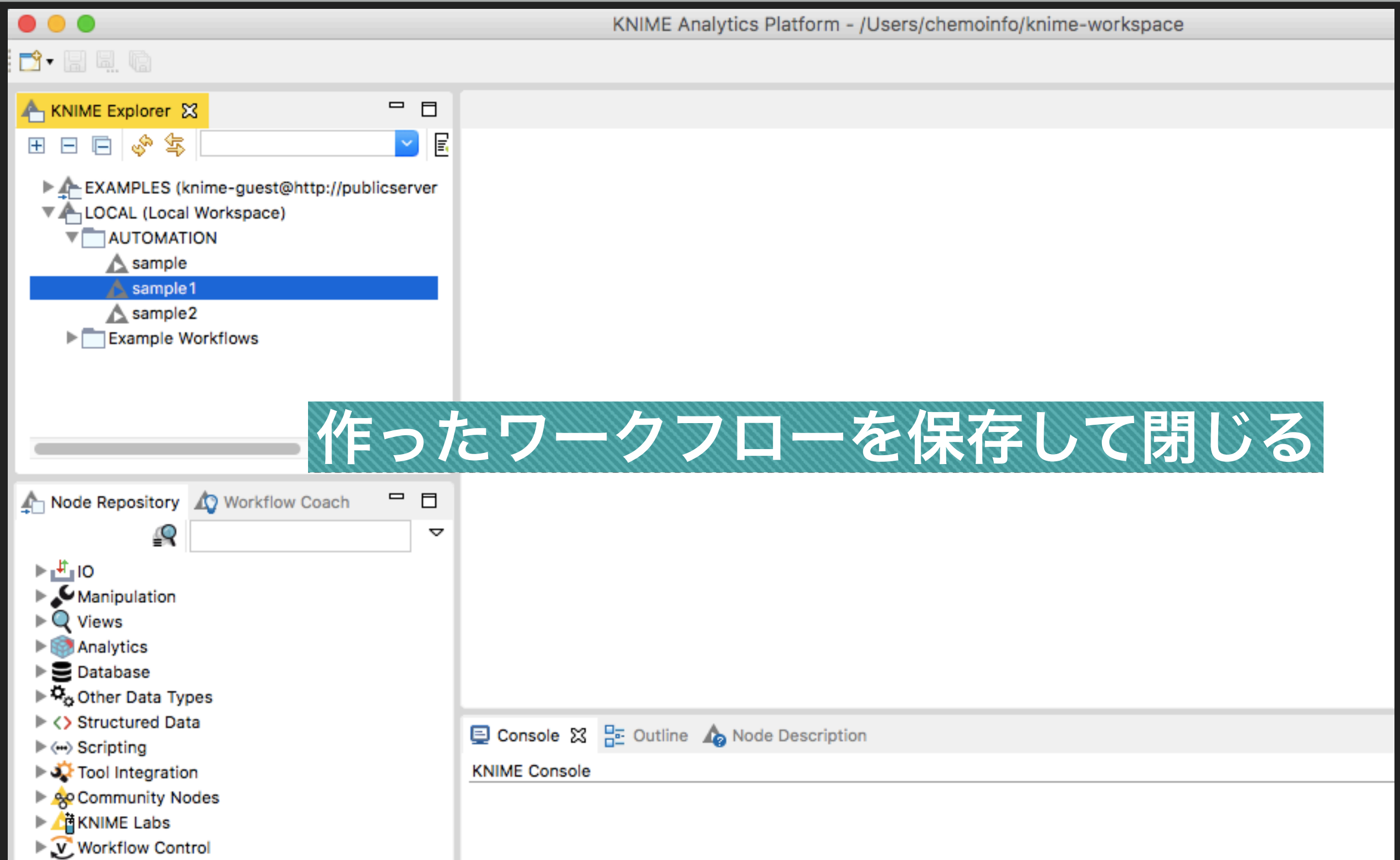
Append Variable: OUTPUT_FILE

Replace Variable: INPUT_SDFILE

Syntax check on close

出力するファイル名の変数OUTPUT_FILEを作る

2017.07.08



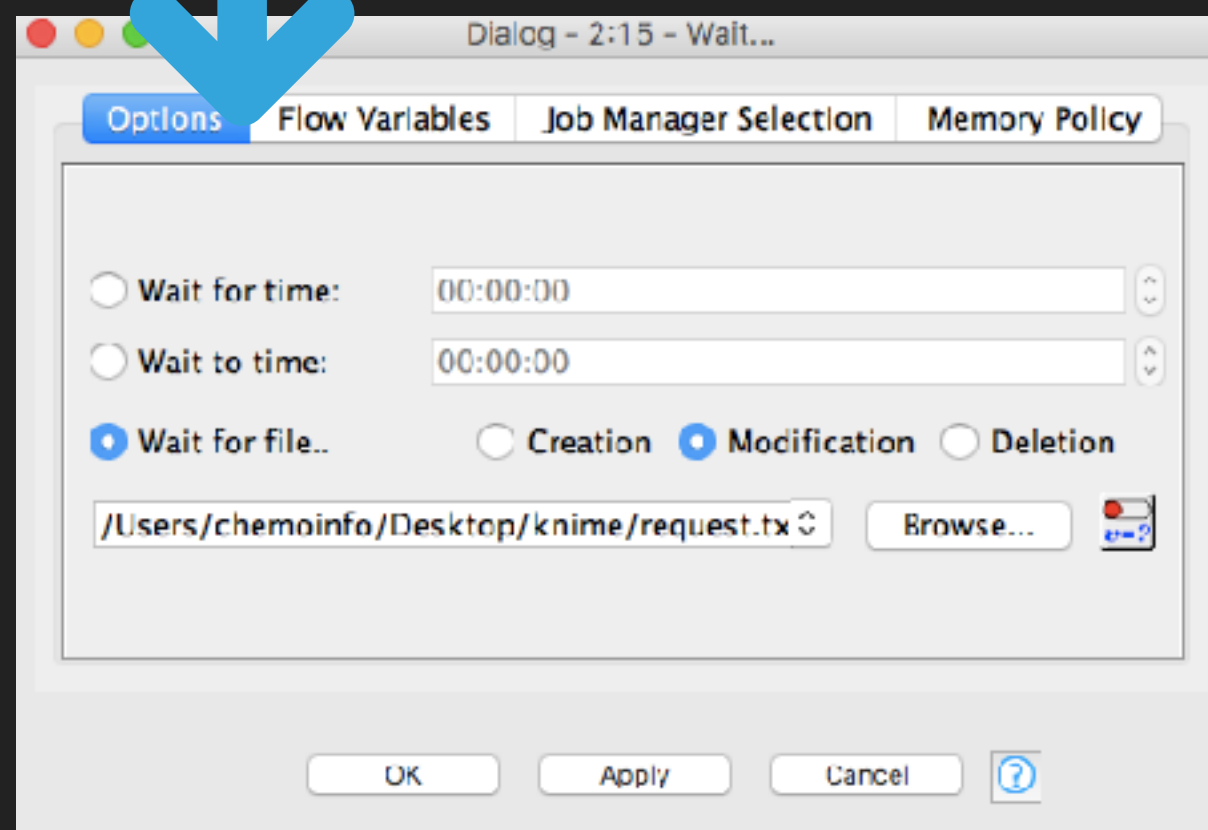
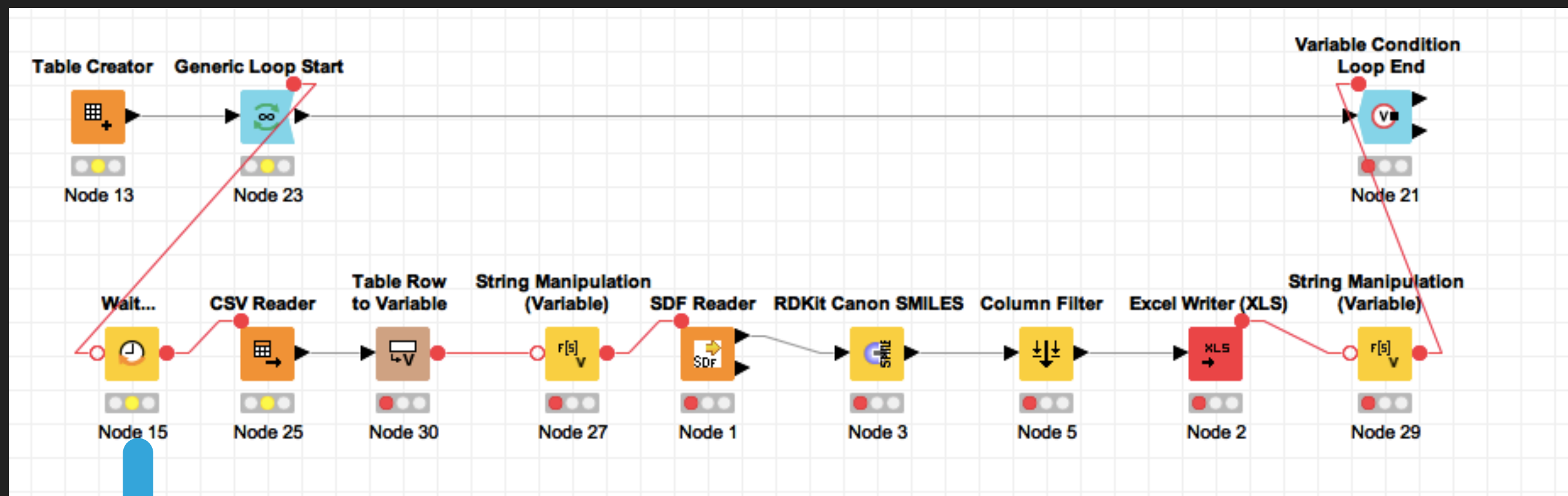
BATCH MODEを使ってみた

```
/Applications/KNIME\ 3.3.1.app/Contents/  
MacOS/Knime -consoleLog -nosplash -reset -  
noexit -application  
org.knime.product.KNIME_BATCH_APPLICATION -  
workflowDir="/Users/chemoinfo/knime-  
workspace/AUTOMATION/sample1"  
-workflow.variable="INPUT_SDFILE,/Users/chemoinfo/Desktop/knime/  
sample.sdf,String"
```


BATCH MODEの使い道

- ▶ GUIよりも起動時間は半分くらい。でも早くはない。
 - ▶ メモリも半分くらい。
- ▶ 別途WEBサーバーを立てて、batch modeからMOEとか SchrodingerのExtensionを呼び出すこともできる。が、あまり早くはない。でも。KNIME Serverも結構高い。

2017.07.08



おまけ：起動時間を省略する
方法
無限LOOPを作って、
WAITでファイルを監視する。
ファイル内に処理したいファ
イル名を書いておく