

2年前の紹介時点と比べて、変わったところと、
自動化する際のポイントを紹介する

@kochi0603

~~2017.07.08~~

2019.09.21

~~KNIMEのBATCH MODEを使ってみた~~
使ってる

DB更新のデータ処理はKNIME化

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

社内に強力な人たちが増えた

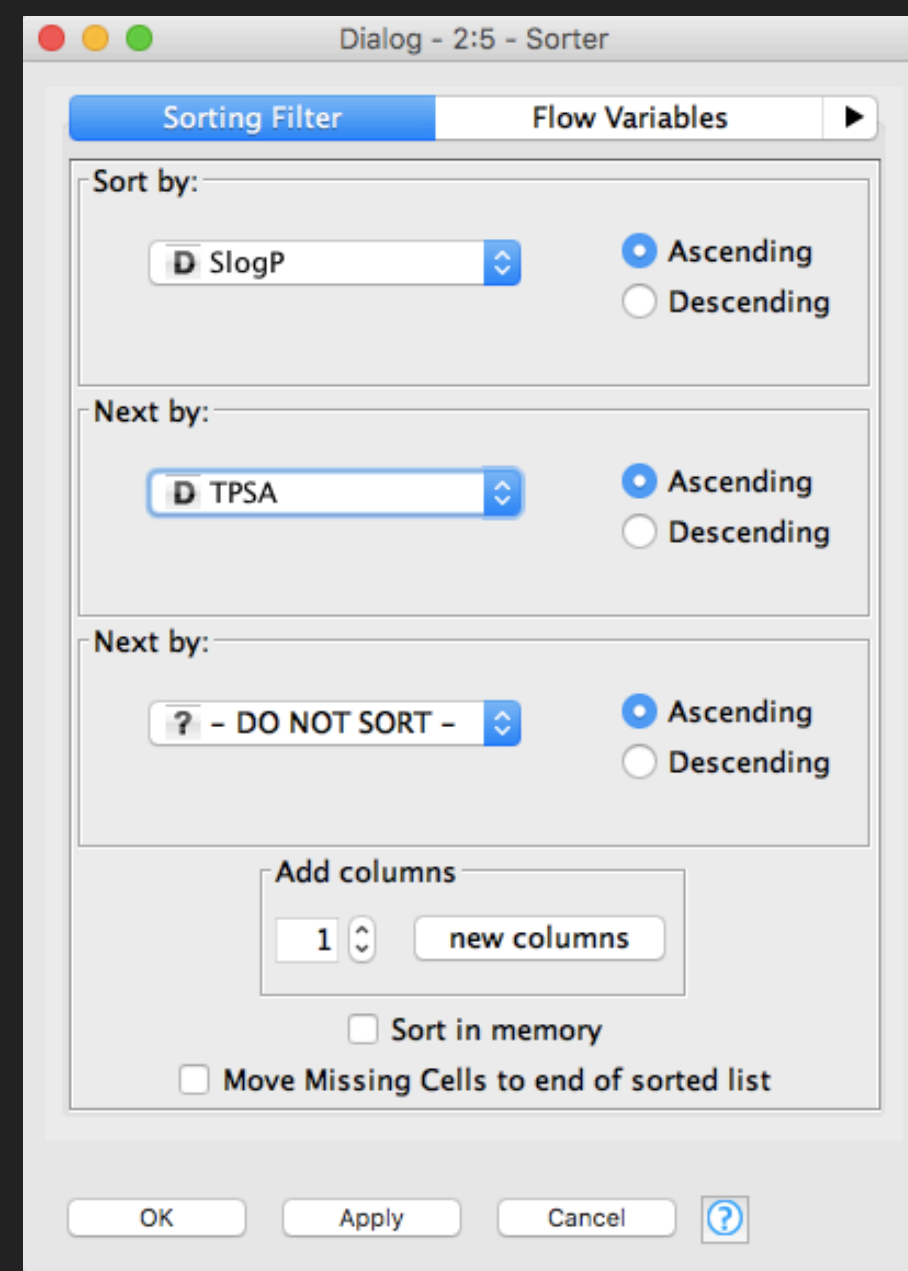
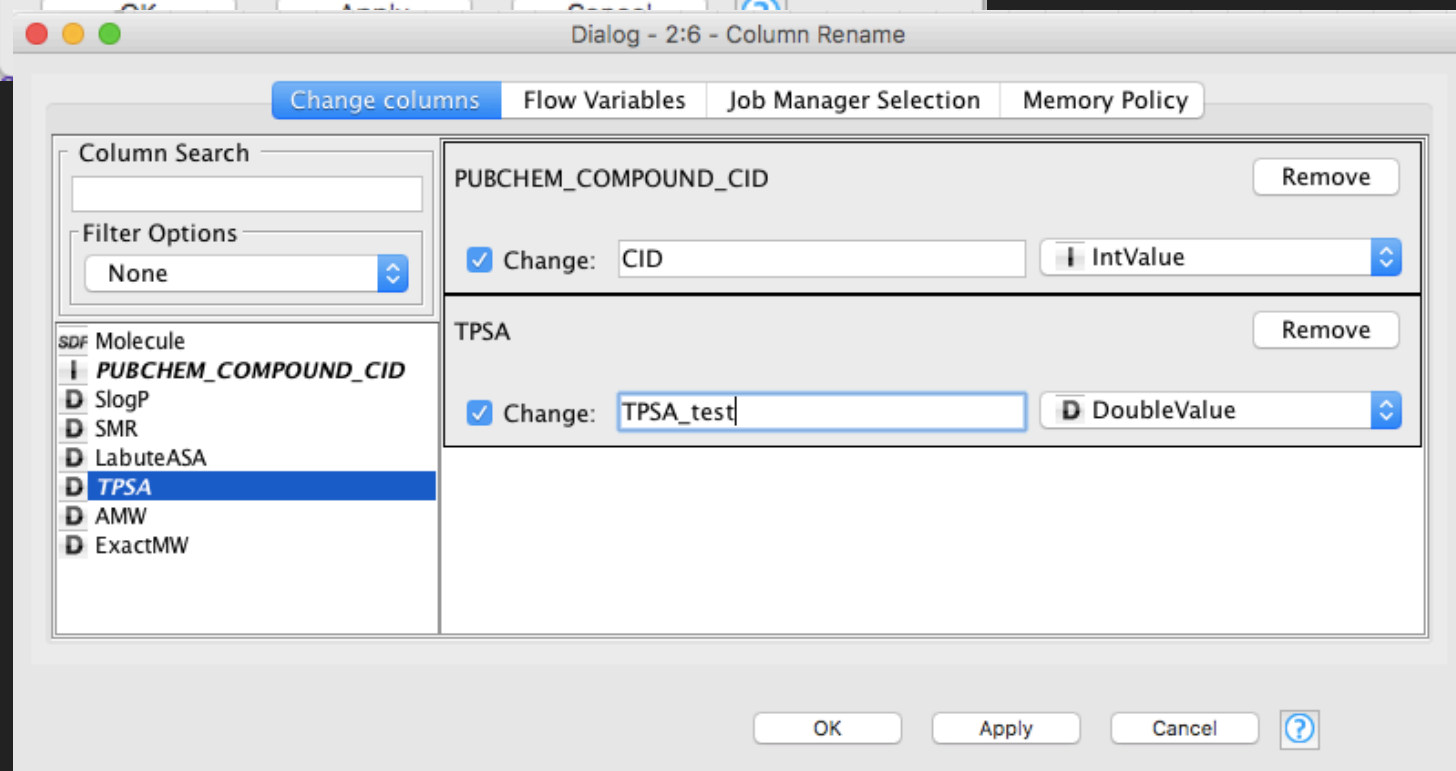
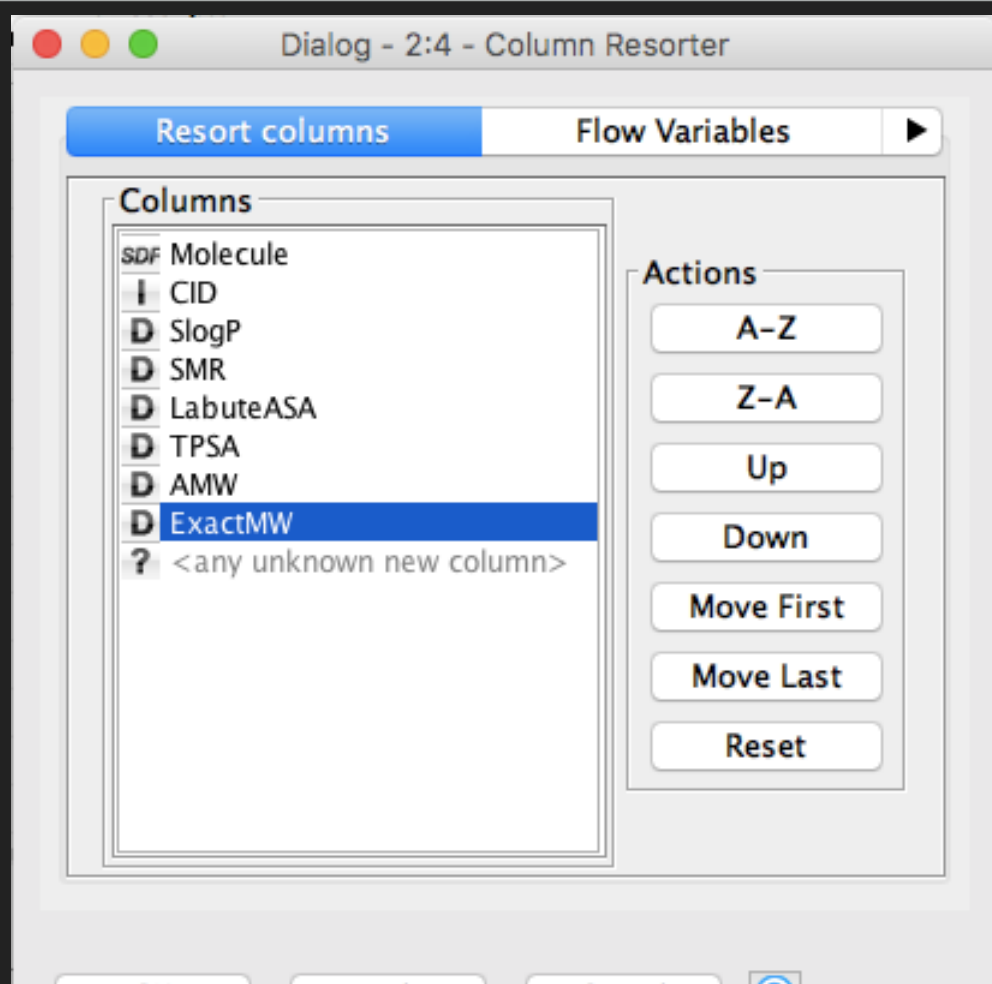
2017.07.08

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 shows the KNIME Analytics Platform interface. On the left, the 'KNIME Explorer' pane displays the project structure, including 'EXAMPLES', 'LOCAL (Local Workspace)', 'AUTOMATION', and 'Example Workflows'. Below it, the 'Node Repository' pane lists various nodes, including 'RDKit' nodes like 'RDKit Descriptor Calculation', 'RDKit Calculate Charges', 'RDKit Generate Coords', 'RDKit Optimize Geometry', 'RDKit Add Conformers', 'RDKit Open 3D Alignment', 'RDKit RMSD Filter', and 'Fingerprints'. 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, showing the 'Options' tab. The 'RDKit Mol column' is set to 'SDF Molecule'. The 'Available descriptors' section lists several descriptors, with 'NumLipinskiHBA', 'NumLipinskiHBD', 'NumRotatableBond', 'NumHBD', 'NumHBA', and 'NumAmideBonds' listed in the 'Exclude' section. The 'Include' section lists 'SlogP', 'SMR', 'LabuteASA', 'TPSA', 'AMW', and 'ExactMW'. The 'Flow Variables' tab is also visible, showing a list of variables with 'WARN' and 'ERROR' status indicators. The 'Job Manager Selection' and 'Memory Policy' tabs are also present. The 'OK', 'Apply', and 'Cancel' buttons are at the bottom of the dialog.

2017.07.08

列の順番を変えるとか、
列名変更とか、
データのソートとか。

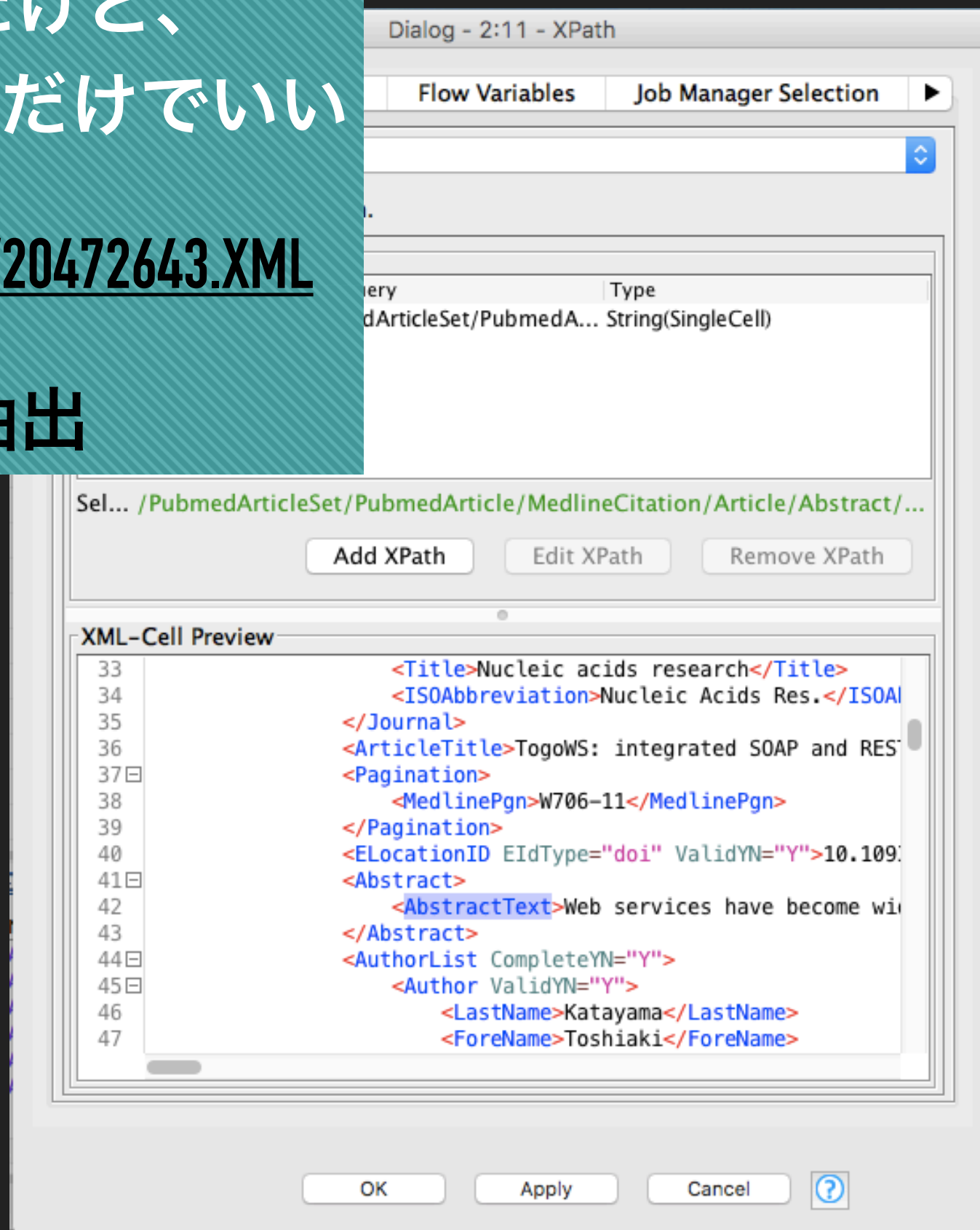
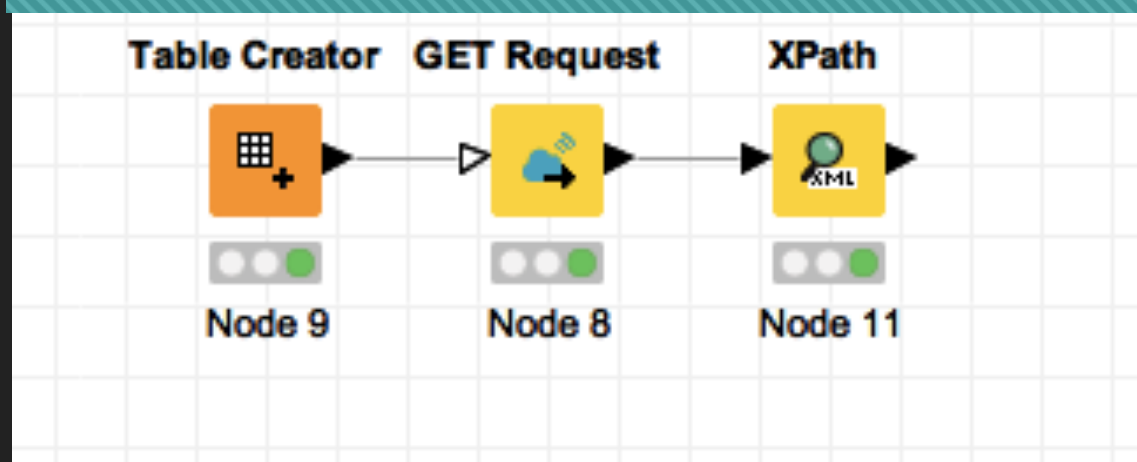


2017.07.08

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

Install

Available Software

Check the items that you wish to install.

Work with: --All Available Sites-- Add...

Find more software by working with the ["Available Software Sites"](#) preferences.

type filter text

Name	Version
<input type="checkbox"/> KNIME Community Contributions - Cheminformatics	
<input type="checkbox"/> 3D-e-Chem KNIME nodes	1.0.12.v201703071314
<input type="checkbox"/> CheS-Mapper-Node extension for KNIME Workbench	2.2.200.v201510090835
<input type="checkbox"/> EMBL-EBI Nodes for KNIME	1.0.400.v201510090835
<input type="checkbox"/> Enalos Nodes for KNIME	1.0.100.v201510090835
<input type="checkbox"/> Erlwood Knime Open Source Cheminformatics	3.3.0.v201701271222
<input type="checkbox"/> Erlwood Knime Open Source Core	3.3.0.v201701271222
<input type="checkbox"/> EU-OPENSOURCE	1.0.100.v201510090852
<input type="checkbox"/> Indigo KNIME integration	2.0.0.v201603210922
<input type="checkbox"/> Indigo KNIME integration (deprecated)	1.1.501.v201604281629
<input type="checkbox"/> KNIME-CDK	1.5.600.v201607162102
<input type="checkbox"/> Lhasa Metabolism Feature	1.1.2.v201607181415
<input type="checkbox"/> Lhasa public plugin	1.1.500.v201705211831
<input type="checkbox"/> OCHEM Integration Nodes	2.0.800.v201510091117
<input type="checkbox"/> RDKit KNIME integration	3.2.2.v201705150604
<input type="checkbox"/> RDKit KNIME Wizards	3.2.2.v201612061535
<input type="checkbox"/> Vernalis KNIME Nodes	1.10.0.v201706231506
<input type="checkbox"/> KNIME Community Contributions - Image Processing and Analysis	

Select All Deselect All

Details

RDKit, CDK, 色々使える。

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
@kzfmix KNIMErary
- ▶ 国内にコミュニティがない。教えてくれる人がいない
社内mattermost

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??

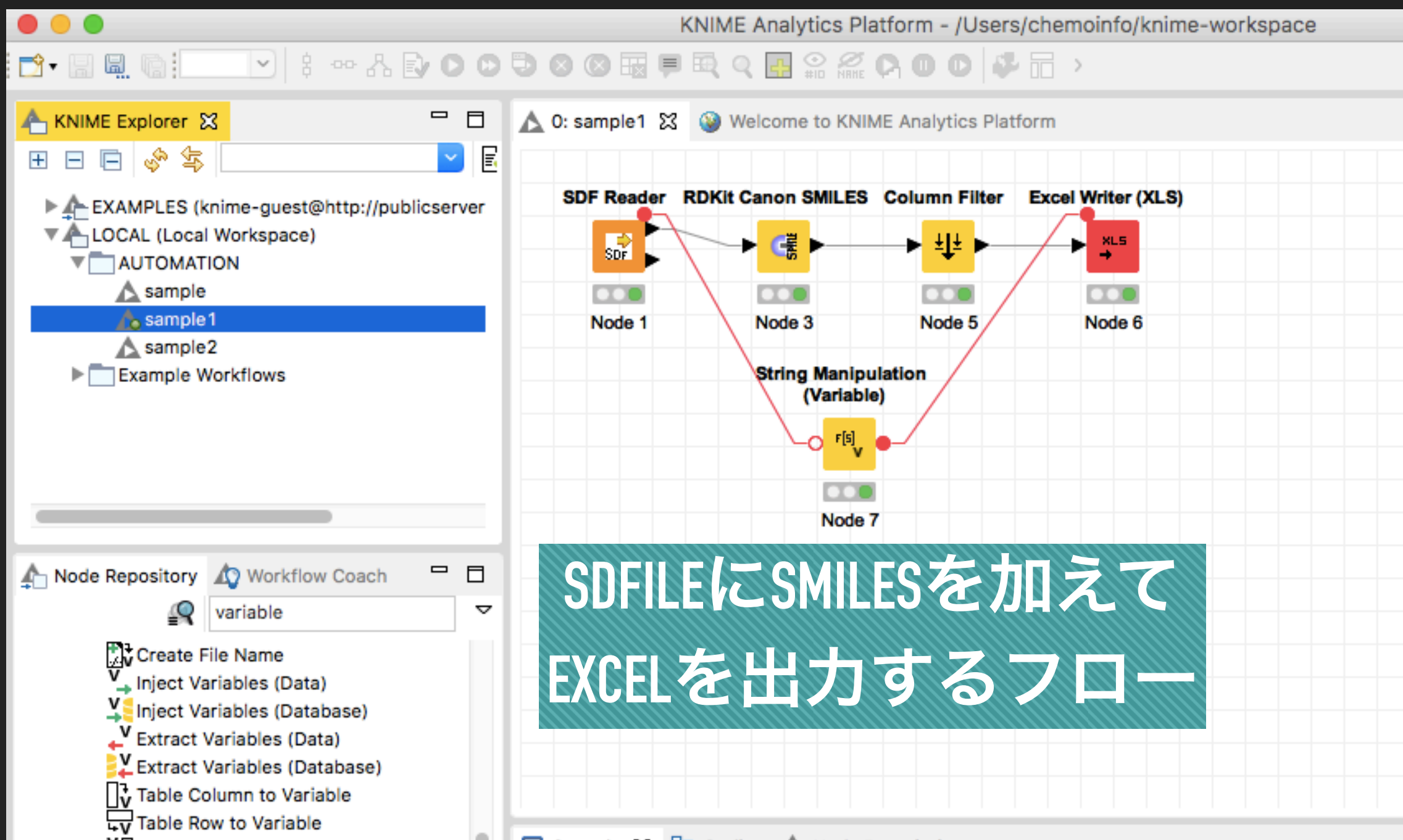
BATCH MODEを
使ってみる

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

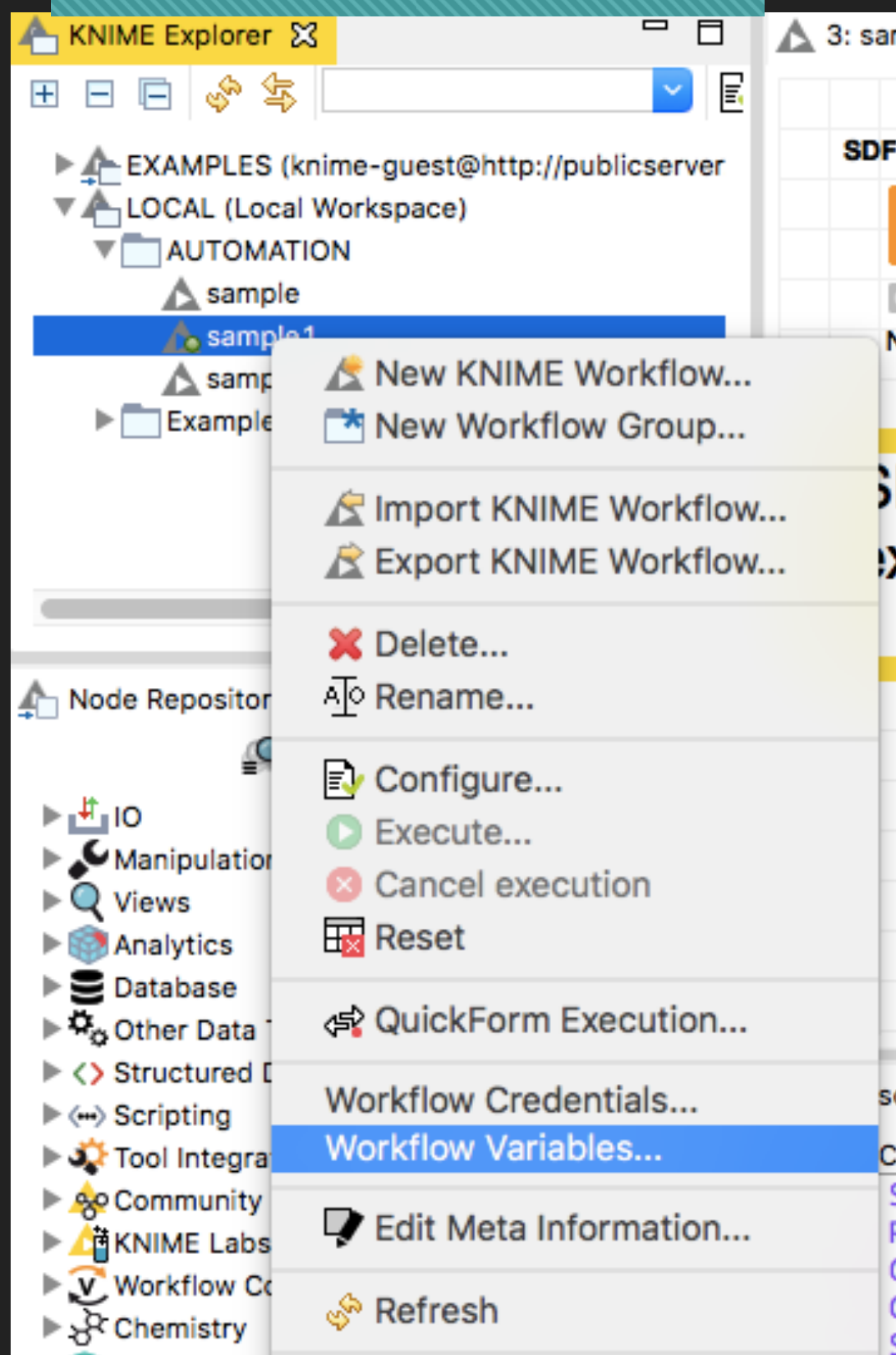
2017.07.08

こんな感じで作ります。

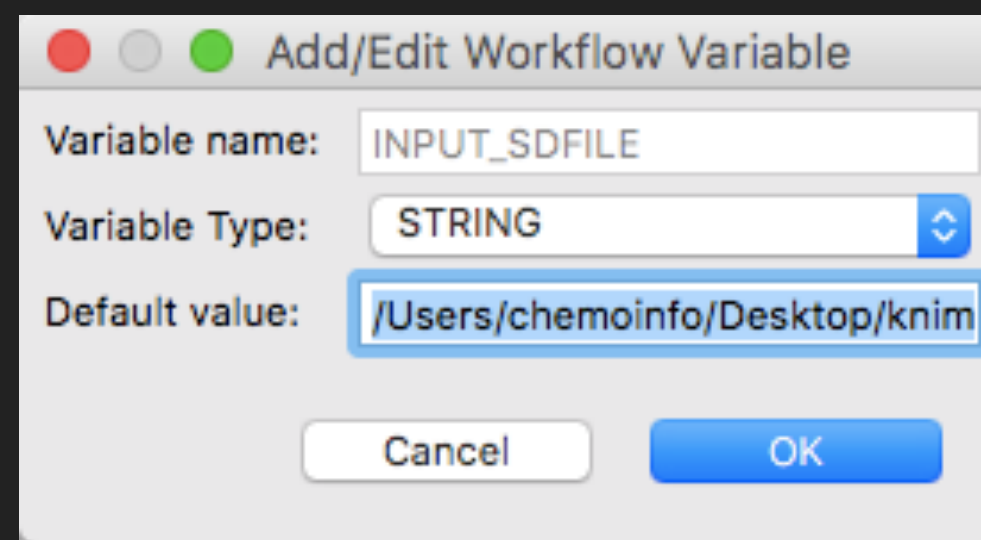
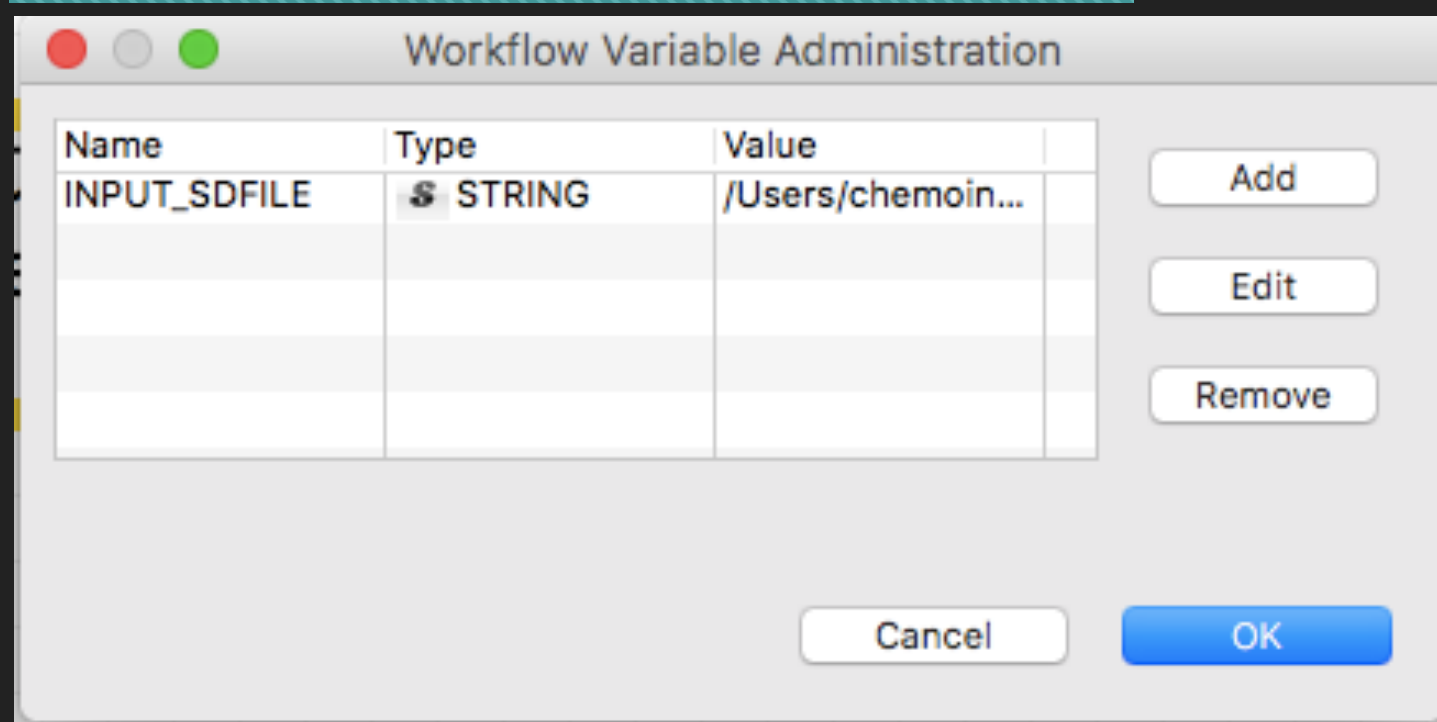


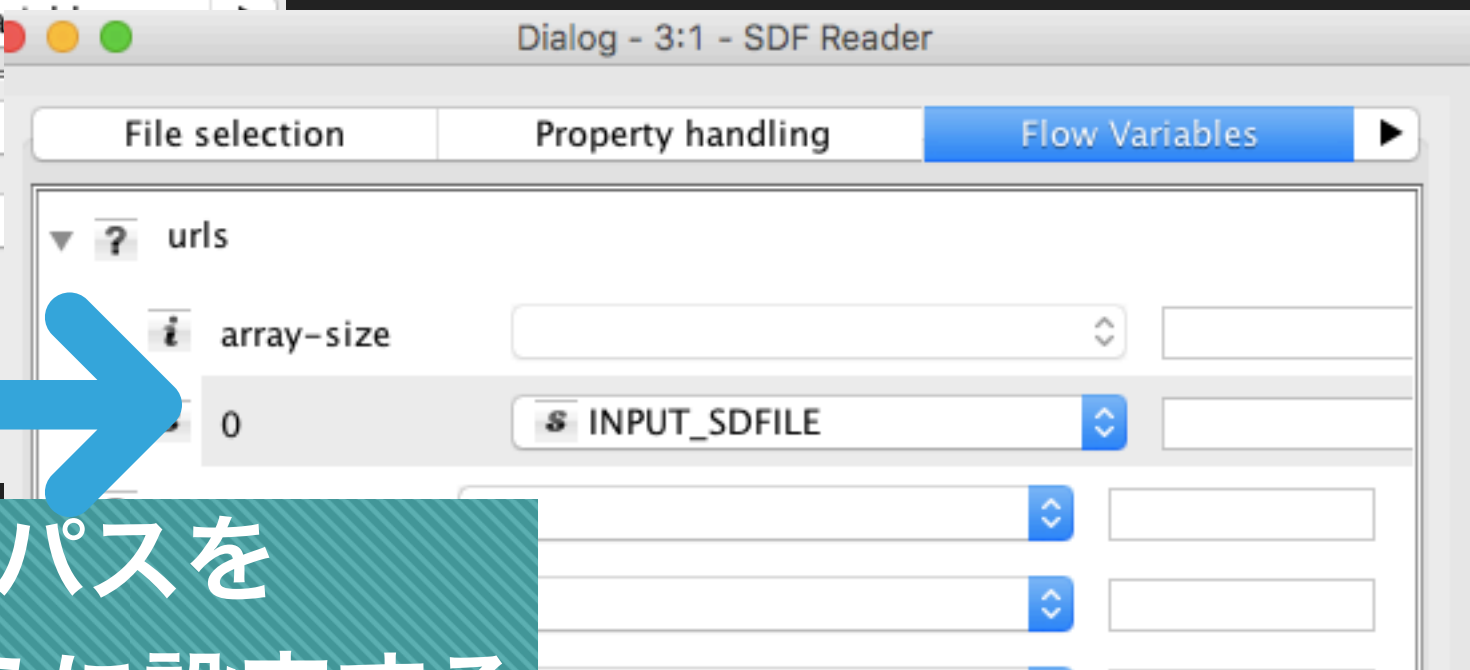
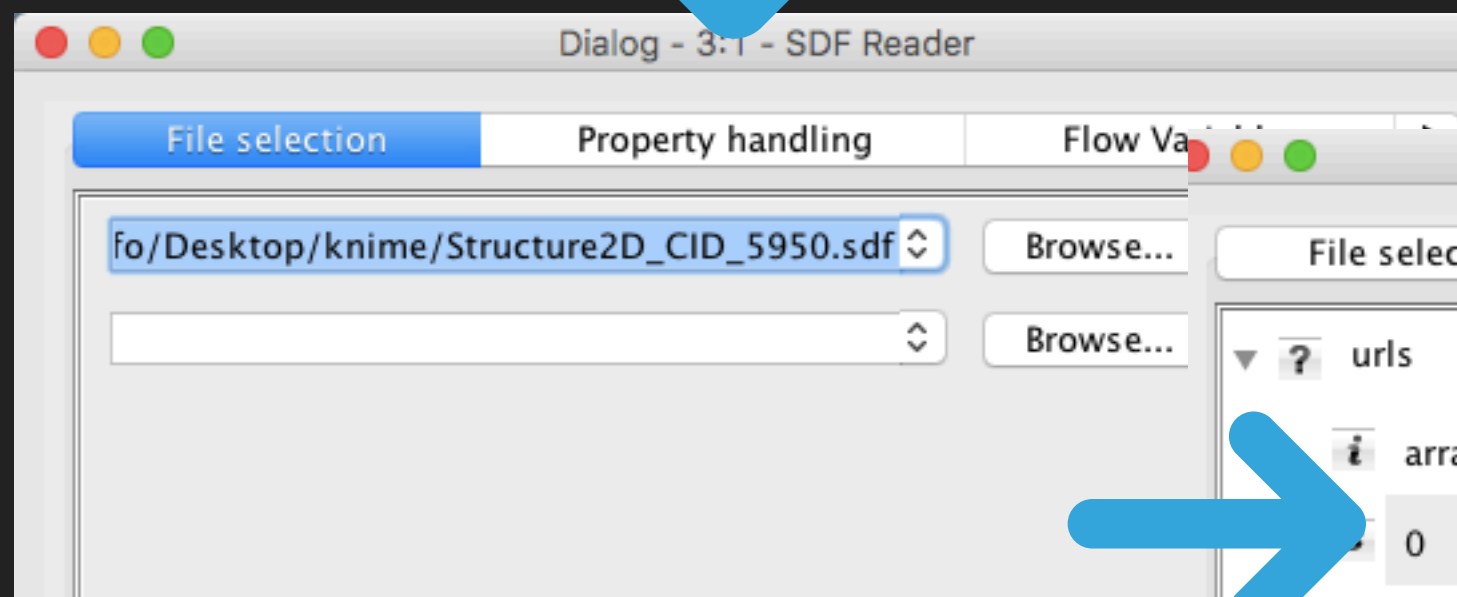
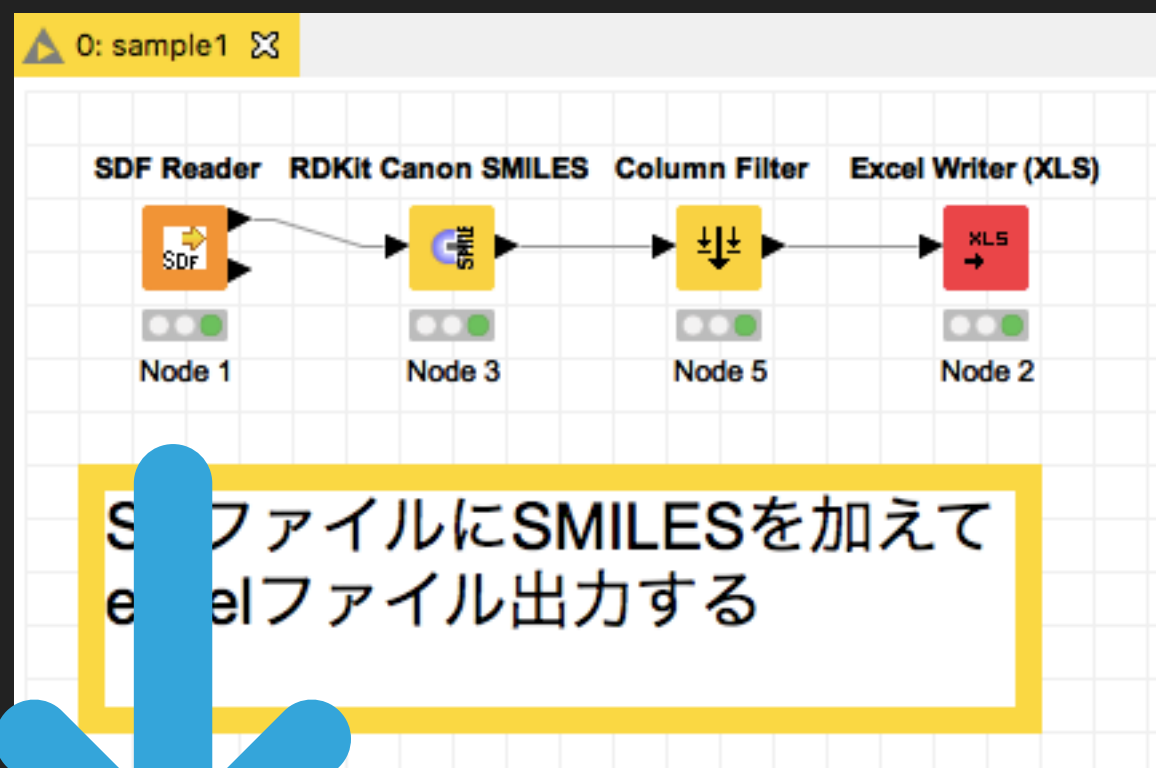
2017.07.08

作成したWORKFLOW に



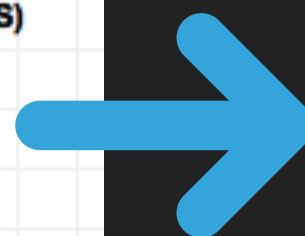
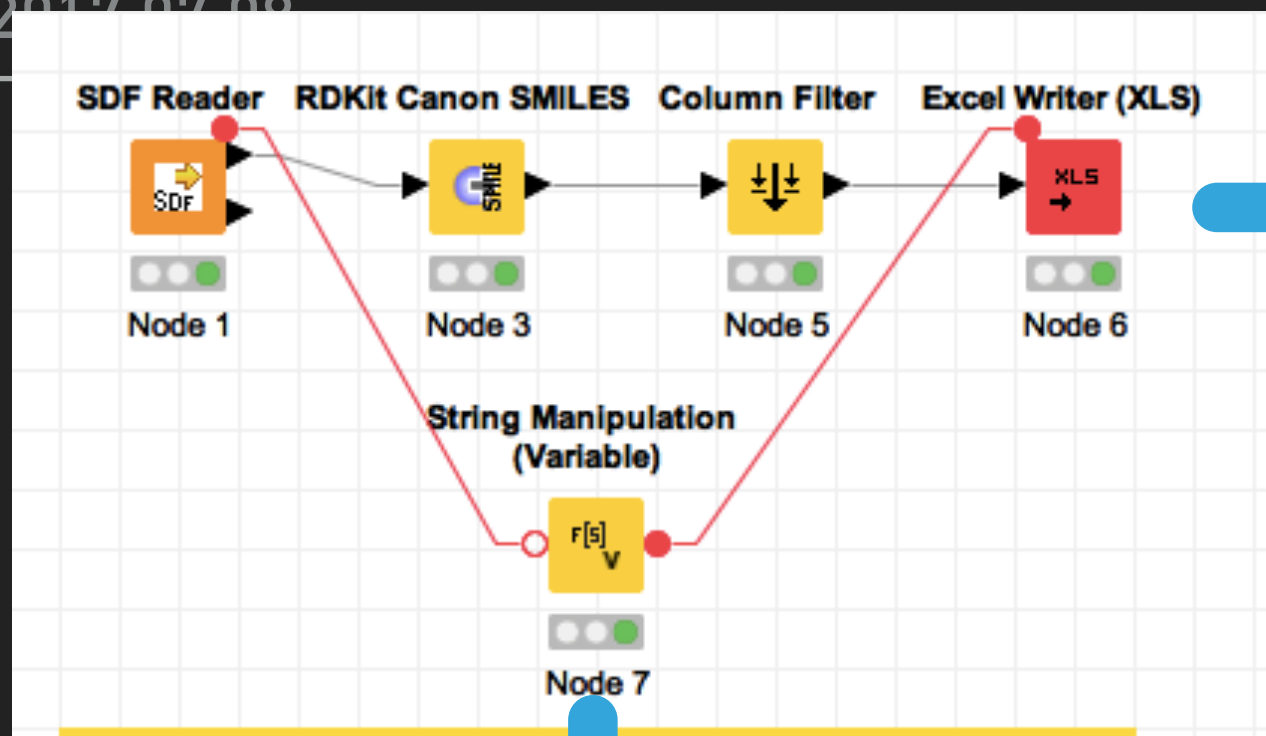
変数INPUT_SDFILEを追加する





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 - 0:6 - 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...		
xlsxwriter2		

OK



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

Flow Variable List

- \$ INPUT_SDFILE
- \$ knime.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(0) ${INPUT_SDFILE}$$, ".xlsx" )
```

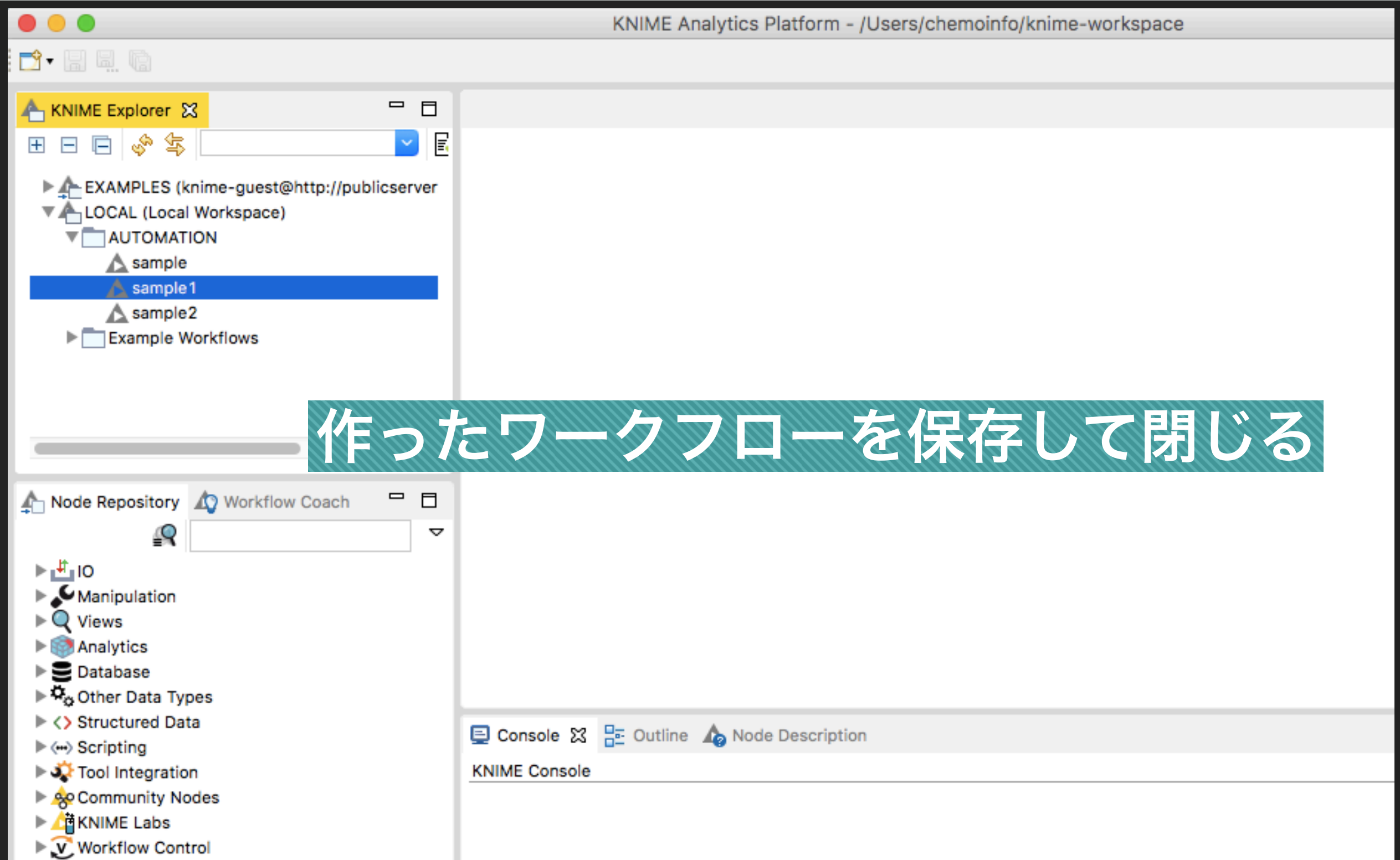
Append Variable: OUTPUT_FILE

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



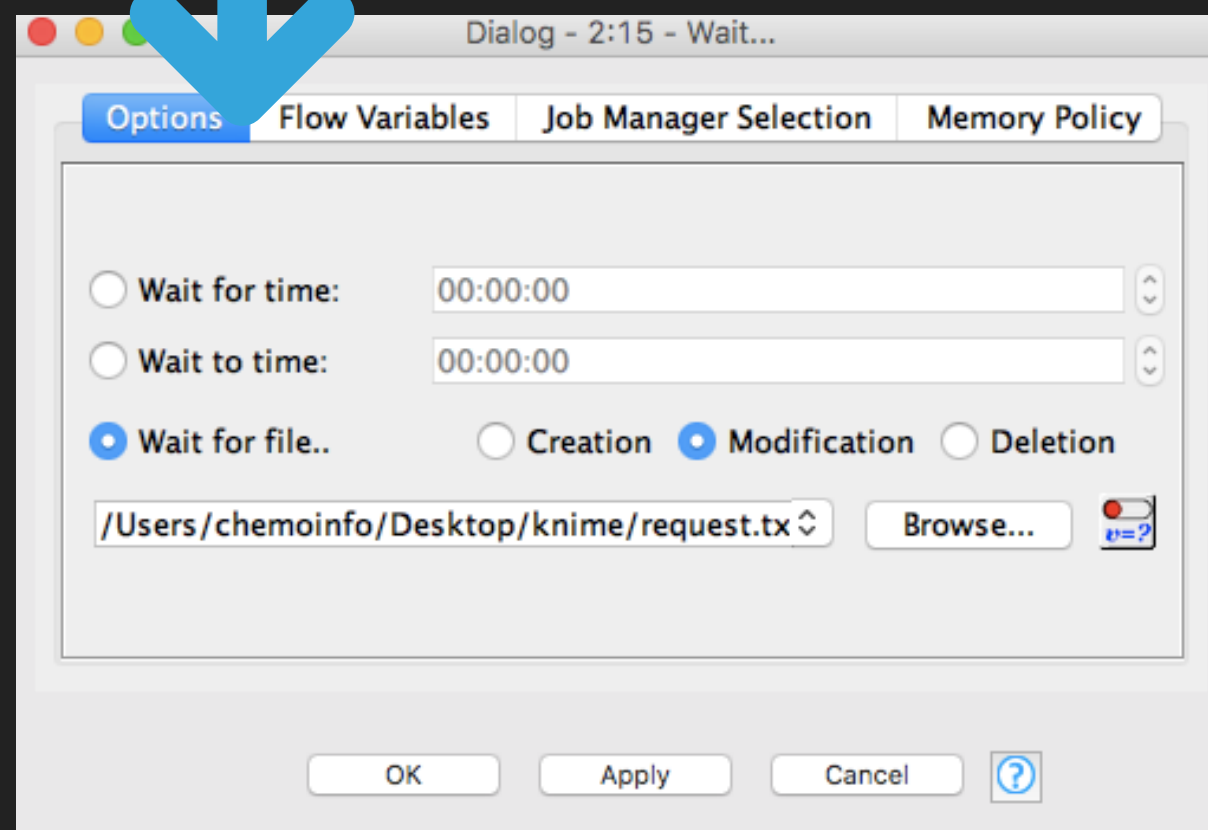
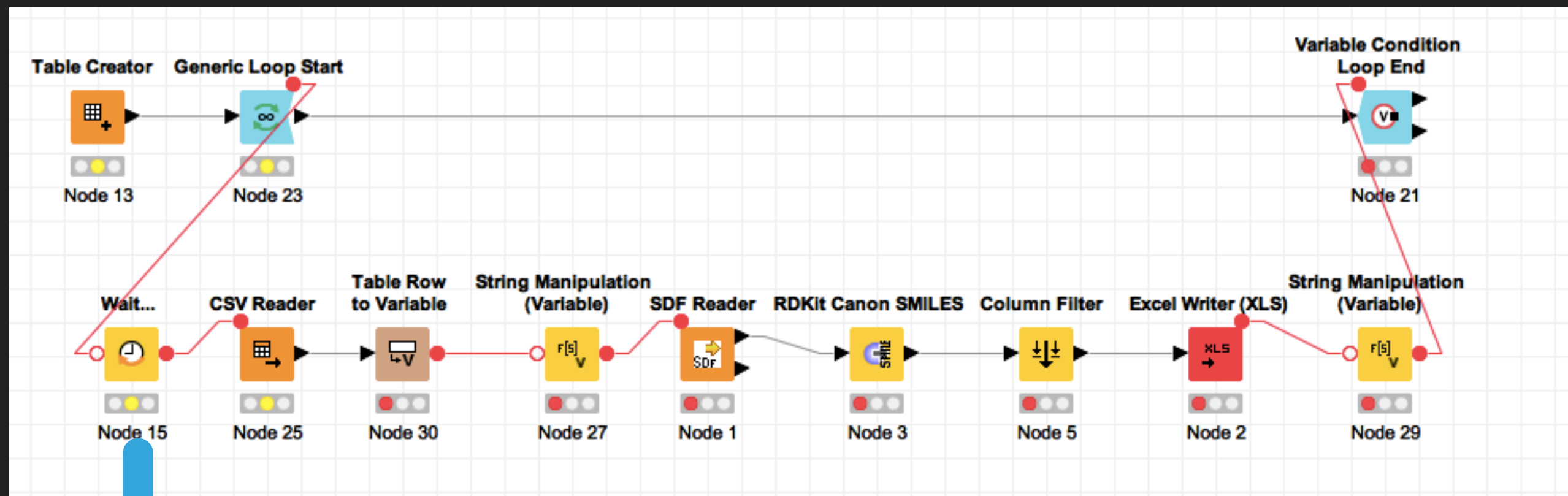
BATCH MODEを使ってみた

```
-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

DBの自動更新系だと気にならない

- ▶ 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。



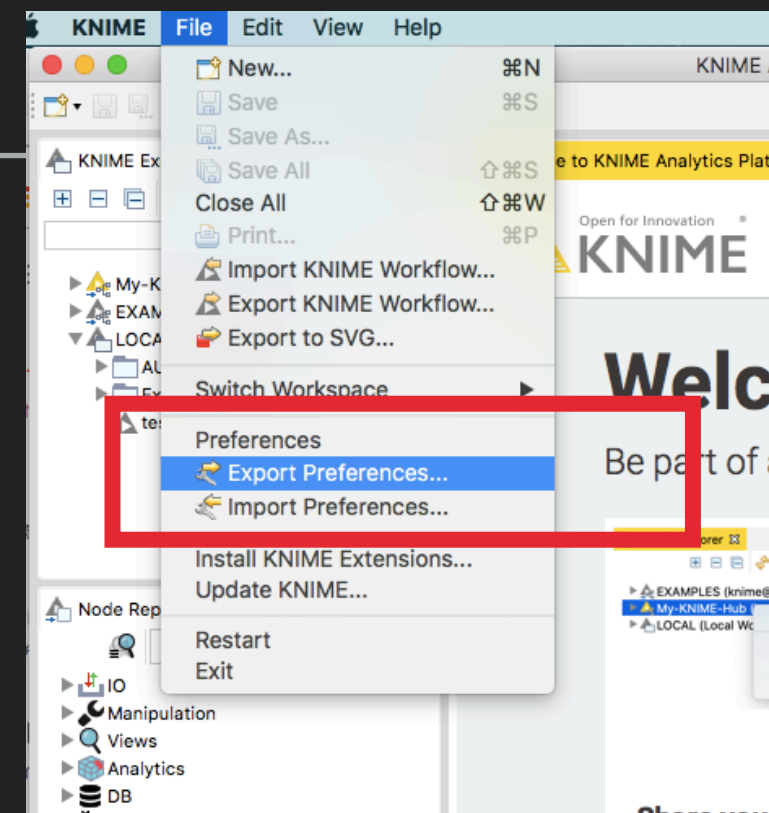
おまけ：起動時間を省略する
方法
無限LOOPを作って、
WAITでファイルを監視する。
ファイル内に処理したい
ファイル名を書いておく

自動化する時 の注意点

テキスト

(1)

- ▶ KNIMEにプラグインを追加した場合、 BATCH_MODEから起動するには、設定ファイルを読み込まないといけない。
 - ▶ あらかじめ export preference で設定ファイルを出力しておく



```
knime.exe -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" -preferences=設定ファイ  
ル
```

(2)

- ▶ External Tool(Labs) Nodeを使えば、KNIMEから直接外部プログラムを呼び出せますが、shell(winならバッチ)経由で間接的に呼び出す方が確実です。
- ▶ 外部プログラムに渡したい引数を、External Tool Nodeがうまく表現できない。