Dockerizing KNIME

Recipes for a KNIME Cocktail

Corpse Reviver

**Ingredients**

- 2 parts Brandy
- 1 part Apple Brandy
- 1 part Sweet Vermouth
- A twist of lime

**Instructions**

1. Combine brandy, apple brandy, and sweet vermouth in a cocktail shaker.
2. Shake with ice.
3. Strain into a chilled cocktail glass.
4. Garnish with an apple slice.

**Serves 1**
PRECISESADS gathers a wide range of data from thousands of patients suffering from auto-immune disorders that share common pathophysiological mechanisms in order to molecularly reclassify the diseases and eventually develop targeted therapies.

systemic lupus erythematosus
systemic sclerosis
Sjögren’s syndrome
rheumatoid arthritis
primary antiphospholipid syndrome
mixed connective tissue disease
Advanced cytometric analysis using KNIME leveraged with Docker and R

The prelude

CymeR Docker image

CymeR is open-source software that ties several separate components into a single program. The heart of the program is a Docker image that can be found at https://hub.docker.com/u/bmuchmore/cyme/

CymeR workspace

To make use of this Docker image, however, we provide an eclipse workspace along with a number of scripts for automated installation, start-up, updating and removal on GitHub at https://github.com/bmuchmore/CymeR

CymeR on different platforms
The Goal of CymeR

To implement graphical user interfaces (GUIs) for state of the art cytometry algorithms (e.g. clustering, dimension reduction, feature selection etc) in an open source environment.

To use these algorithms in ways that let you see your data in ways previously impossible.
Brief Introduction to Flow Cytometry

A diagram illustrating the components of a flow cytometry system, including a cell sample, nozzle, laser, dichroic mirror, filters, photomultiplier tubes (PMTs), and an analysis workstation.

- Cell sample
- Nozzle
- Dichroic mirror
- Filters
- Photomultiplier tubes (PMTs)
- Analysis workstation

Flow sheath
Obscuration bar

Components connected to the analysis workstation include:
- SSC
- FL-2
- FL-1
- FL-3
- FSC
- ADC
Introduction to Mass Cytometry
Using KNIME as a simple GUI front-end for complex functions

What the biologist sees:
Using KNIME as a simple GUI front-end for complex functions

Underlying KNIME workflow:
Using KNIME as a simple GUI front-end for complex functions

Underlying R code:
R can make KNIME (more) interactive

Sometimes, R needs interactive inputs that can not be pre-configured in KNIME. For these scenarios, I code in an interactive table editor that goes back into R as a data frame. All the biologist has to do is fill out the table when it pops up:
Essential workflow

1) Read in a FCS file

2) Apply function to FCS file

3) Write out new FCS file (and other files as needed)
KNIME makes (scientific) data fun

PhenoGraph: A cellular social network that returns clustering designation based on Louvain communities
KNIME makes (scientific) data fun

BHt-SNE: Non-linear dimension reduction using Barnes-Hut t-Distributed Stochastic Neighbor Embedding
KNIME makes (scientific) data fun

Destiny: Non-linear dimension reduction using diffusion maps:
The abbreviated list of programs needed to make everything work:

- KNIME
- KNIME extensions
- JAVA 8
- R
- Python 2.7
- Python 3.4
- X11 Server
- Firefox
- 60+ R or R/Bioconductor packages
- Many other specific packages and...

**All the dependencies required to make the above run**
A brief introduction to...

The following Docker slides have been adapted from:
http://www.slideshare.net/dotCloud/docker-intro-november?from_action=save
A standard container that is loaded with virtually any goods, and stays sealed until it reaches final delivery.

...in between, can be loaded and unloaded, stacked, transported efficiently over long distances, and transferred from one mode of transport to another.
Docker is a shipping container system for code

An engine that enables any payload to be encapsulated as a lightweight, portable, self-sufficient container…

...that can be manipulated using standard operations and run consistently on virtually any hardware platform
Containers vs. VMs

Containers are isolated, but share OS and, where appropriate, bins/libraries.
FROM ubuntu:14.04
MAINTAINER Brian Muchmore "brian.muchmore@genyo.es"

##Install Java8
RUN add-apt-repository ppa:webupd8team/java && \
    apt-get update -y && \
    echo oracle-java8-installer shared/accepted-oracle-license-v1-1 select true | /usr/bin/debconf-set-selections && \
    xvfb-run -a apt-get install oracle-java8-installer libxext-dev libxrender-dev libxtst-dev -y && \
    xvfb-run -a apt-get install liblzma-dev -y && \
    xvfb-run -a apt-get install libglu1-mesa-dev -y

##Download KNIME
RUN echo 'Installing knime' && \
    wget http://download.knime.org/analytics-platform/linux/knime_3.1.0.linux.gtk.x86_64.tar.gz -O /tmp/knime.tar.gz -q \
    tar -xf /tmp/knime.tar.gz -C /home/developer && \
    rm /tmp/knime.tar.gz && \
    apt-get install libwebkitgtk-1.0-0 -y

USER developer
ENV HOME /home/developer
WORKDIR /home/developer
CMD /home/developer/knime_3.1.1/knime -data /home/developer/knime-workspace
What are the basics of the Docker system?
Changes and Updates

Docker
Container Image
Registry
Docker Engine

Push

Host is now running A''

Container Mod A'

Container Mod A''

App A''
Bins''/Libs''
Docker Engine

App A'
Bins'/Libs'

Update

App A'
Bins'/Libs'

App A
Bins/Libs

Host running A wants to upgrade to A''. Requests update. Gets only differences.
CymeR in Linux

docker run -ti --rm
-e DISPLAY=$DISPLAY
CymeR-Data:/home/developer/Data:rw
CymeR-Workspace:/home/developer/knime-workspace:rw
/tmp/.X11-unix:/tmp/.X11-unix:rw
bmuchmore/cymer
CymeR in Windows Using Less than 3 GB of RAM

Windows 7, 8, 10

X11 Server
OpenBox

Workspace
Data

DOCKER TOOLBOX:
VirtualBox (headless)

Babun (Cygwin)

Tiny Core Linux

Docker

CymeR
CymeR in Mac

COMING SOON
(It should be much simpler to implement than CymeR in Windows)
CymeR

Advanced cytometric analysis using KNIME leveraged with Docker and R

The prelude

CymeR Docker image

CymeR is open-source software that ties several separate components into a single program. The heart of the program is a Docker image that can be found at https://hub.docker.com/u/bmuchmore/cymeR

CymeR workspace

To make use of this Docker image, however, we provide an eclipse workspace along with a number of scripts for automated installation, start-up, updating and removal on GitHub at https://github.com/bmuchmore/CymeR

CymeR on different platforms
Thanks to...

Dr. Marta Alarcon-Riquelme

IMI and PRECISESADS

KNIME

DOCKER

All of you.
For much more info, go to:

http://bmuchmore.github.io/Dock-o-KNIME/

Or

http://bmuchmore.github.io/CymeR/

Questions?

Ask now or contact me at bmuchmore@gmail.com
(I will do my best to respond in a timely manner)