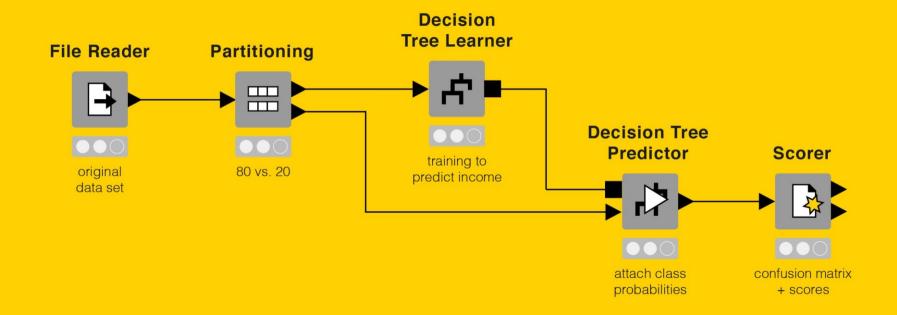
KNIME® BEGINNER'S LUCK

EXTRACT

"How to Download and Install KNIME Analytics Platform (Chapt. 1.3)



Copyright[©] 2021 by KNIME Press

All Rights Reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording or likewise.

This book has been updated for KNIME 4.3.

For information regarding permissions and sales, write to:

KNIME Press Technoparkstr. 1 8005 Zurich Switzerland

knimepress@knime.com

ISBN: 978-3-033-02850-0

Foreword

This is the first book I wrote in 2010 for the KNIME Press on how to use KNIME Analytics Platform. Since we are getting close to the 10-year anniversary of this book, we (the KNIME Press Team and I) thought that it might be in need of a new Foreword text. It does not actually need an overall update, as ever since its birth it has been updated twice a year every year, following each new release of KNIME Analytics Platform; not immediately after - but close enough.

That is right! KNIME Beginner's Luck, like all other e-books from KNIME Press, is a live e-book, constantly changing to fit the newest version of the software. This liveness of the e-book is also the reason why it has only rarely been printed. Updating printed pages is undoubtedly harder than updating a pdf file!

As this is the first book, it is inevitably about the basics: the basics of KNIME Analytics Platform of course and also the basics of a data science project. This book guides you through the most important access functions, data transformation operations, and of course machine learning nodes available in KNIME Analytics Platform. Supplemented with many example workflows, exercises, and screenshots, it will quickly familiarize you with the basic functions of the software. If you are looking for more advanced topics, you won't find them here, instead....

If you want to learn more about advanced machine learning algorithms, flow variables, or loops, check the sequel to this book: "KNIME Advanced Luck". If you want to learn more about text processing, have a look at the book, "From Words To Wisdom". If you come from that school of thoughts where reading manuals or instructions is overrated, you can start directly with reading about solutions to case studies in various application fields in our collection "Practicing Data Science". If your job is more about integrating and blending different data sources and data types, then the book for you is the "Will they blend?" collection. More useful booklets are available on the KNIME Press page, if you are transitioning from SAS, Excel, or Alteryx.

All this is to say that the KNIME Press team and I have been working hard to provide you with the learning material, books, and tutorials, to become progressively more and more productive with KNIME Software and data science concepts.

Rosaria Silipo (Author of a number of KNIME Press Books, PhD)

Chapter 1. Introduction

1.1. Purpose and structure of this book

We live in the age of data! Every purchase we make is dutifully recorded; every money transaction is carefully registered; every web click ends up in a web click archive. Nowadays everything carries an RFID chip and can record data. We have data available like never before. What can we do with all these data? Can we make some sense out of it? Can we use it to learn something useful and profitable? We need a tool, a surgical knife that can empower us to cut deeper and deeper into our data, to look at it from many different perspectives, to represent its underlying structure.

Let's suppose then that we have this huge amount of data already available, waiting to be dissected. What are the options for a professional to enter the world of Business Intelligence (BI) and Data Science (DS)? The options available are of course multiple and growing rapidly. If our professional does not control an excessive budget, he or she could turn to the world of open-source software. Open-source software, however, is more than a money driven choice. In many cases it represents a software philosophy for resource sharing and control that many professionals support.

Inside the open-source software world, we can find a few Data Science and BI tools. <u>KNIME Analytics Platform</u> represents an easy choice for the non-initiated professional. It does not require learning a specific script and it offers a Graphical User Interface (GUI) to implement and document analysis procedures. In addition - and this is not a secondary advantage - KNIME Analytics Platform can work as an integration platform into which many other BI and Data Science tools can be plugged. It is then not only possible but even easy to analyze data with KNIME Analytics Platform and then to build dashboards on the same processed data with a different BI tool.

Even though KNIME Analytics Platform is very simple and intuitive to use, any beginner would profit from an accelerated orientation through all of the nodes, categories, and settings. This book represents the beginner's luck, because it is aimed to help any beginner to gear up his/her learning process. This book is not meant to be an exhaustive guide to the whole KNIME software. It does not cover implementations under the KNIME Server, which is not open-source, or topics which are considered advanced. Flow Variables, for example, and implementations of database SQL queries are discussed in the sequel book "KNIME Advanced Luck".

This book is divided into six chapters. The first chapter covers the basic concepts of KNIME Analytics Platform, while chapter two takes the reader by the hand into the implementation of the very first KNIME application. From the third chapter, we start the exploration of data science concepts in a more in-depth manner. The third chapter indeed explains how to perform some basic data exploration and visualization, in terms of nodes and processing flow. Chapter four is dedicated to data modeling. It covers a few demonstrative approaches to machine learning, Naïve Bayes, decision trees, and artificial neural networks. Finally, chapters five and six are dedicated to reporting. Usually, the results of an investigation based on data visualization

or, in a later phase, on data modeling must be shown at some point to colleagues, management, directors, customers, or external workers. Thus, reporting is a very important phase at the end of the data analysis process. Chapter five shows how to prepare the data to export into a report, while chapter six shows how to build the report itself.

Each chapter guides the reader through an <u>ETL</u> or a machine learning (ML) process step by step. Each step is explained in detail and offers some explanations about alternative employments of the current nodes. At the end of each chapter several exercises are proposed to the reader to test and perfect what he/she has learned so far.

Examples and exercises in this book have been implemented using KNIME 4.3. They should also work under subsequent KNIME versions, although there might be slight differences in their appearance.

1.2. KNIME community

Being an open-source software, KNIME Analytics Platform benefits of a number of forums and groups of KNIME users all around the world. This is a good safety net for advises, hints, and learning material. We report below the most popular sites and groups.

Useful Web Pages	
http://www.knime.com	The root page in the KNIME web site.
https://www.knime.com/software-overview	The first place to look for an overview of all KNIME products . The open source KNIME Analytics Platform can be downloaded here.
https://www.knime.com/knime-self-paced-	This is the root page of the KNIME self-paced courses (e-learning) where you can learn more about
courses	the specific KNIME functionalities. It covers the whole data science creation cycle from data access
	and data exploration to machine learning and control structures.
	The KNIME Hub is a public repository for KNIME material: workflows, examples, extensions,
https://hub.knime.com/	component templates, and nodes. If you want to start with practical examples, here you can find
	many according to the search term you type in.
	What I find particularly useful is the KNIME Forum. Here you can ask questions about how to use
https://forum.knime.com/	KNIME Analytics Platform or about how to extend it with new nodes. Someone from the KNIME
	community answers always and quickly.

Courses for KNIME Analytics Platform	KNIME periodically offers onsite and online courses for KNIME software. This includes basic and advanced elements. To check for the next available date and to register, just go to the KNIME Events web page https://www.knime.com/learning/events and select the tab Online Courses.		
KNIME Webinars	A number of webinars are also frequently available on specific topics, like chemistry nodes, text mining, integration with other analytics tools, automated machine learning, deep learning, time series analysis, best practices, and so on. To know about the next scheduled webinars, check the KNIME Events web page at https://www.knime.com/learning/events and select the tab Webinars.		
KNIME Meetups and KNIME Summits	KNIME Meetups and KNIME Data Talks are held periodically all over the world. These are always good chances to learn more about KNIME software, to get inspired about new data science projects and to get to know other people from the KNIME Community (https://www.knime.com/learning/events)		
KNIME TV Channel on YouTube	KNIME has its own video channel on YouTube, named KNIMETV. There, a number of videos are available to learn more about many different topics and specially to get updated about the new features in the new KNIME releases (http://www.youtube.com/user/KNIMETV)		

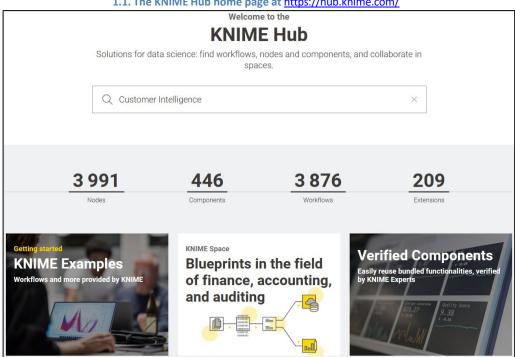
Advanced Features in KNIME Analytics Platform	For the advanced use: Rosaria Silipo, Jeanette Prinz, "KNIME Advanced Luck" (https://www.knime.com/knimepress/advanced-luck)
Reporting Suite	The KNIME Reporting Suite is based on <u>BIRT</u> , another open source tool for reporting. Here is a basic guide on how to use BIRT: D. Peh, N. Hague, J. Tatchell, "BIRT. A field Guide to Reporting.", Addison-Wesley, 2008
Data Science and KNIME	For an overview of data science, data mining, and data analytics, please check: Berthold M.R., Borgelt C., Höppner F., Klawonn F., Silipo R., "Guide to intelligent data science", Springer 2020.

KNIME Hub

However, there is a privileged place where to find information about KNIME nodes and example workflows for your next projects: the KNIME Hub (https://hub.knime.com/).

The KNIME Hub is a repository of applications, components, and nodes to recycle, reuse, and assemble on KNIME Analytics Platform. Or as it says on the home page: The KNIME Hub is "the place to find and collaborate on KNIME workflows and nodes. Here you can find solutions for your data science auestions."

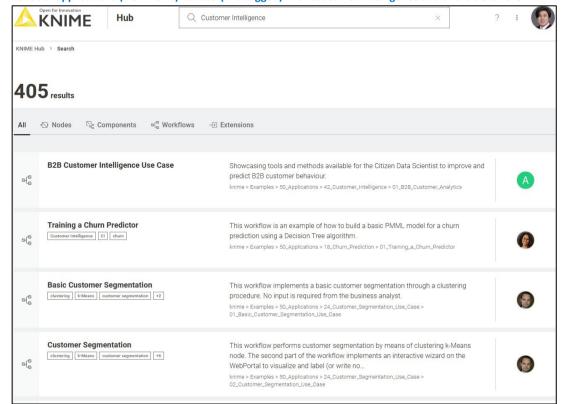
When you access the KNIME Hub the first time, you end up with the page in figure 1.1. This page offers a few links to the starting guide documentation, the KNIME Forum, and the KNIME blog. Most importantly at the top it offers a search box to find applications, nodes, and components uploaded by KNIME users in this shared place of the KNIME community.



1.1. The KNIME Hub home page at https://hub.knime.com/

If we type "Customer Intelligence" in the search box, we end up with a list of nodes and workflows related with Customer Intelligence. Let's select just "Workflows" in the top menu. Then below in figure 1.2 you can see the list of applications (workflows) implementing some aspects of Customer

Intelligence - and appropriately tagged - as uploaded by users of the KNIME community. Indeed, you can upload your own developed applications on the KNIME Hub. All you need is an account with the KNIME Forum.



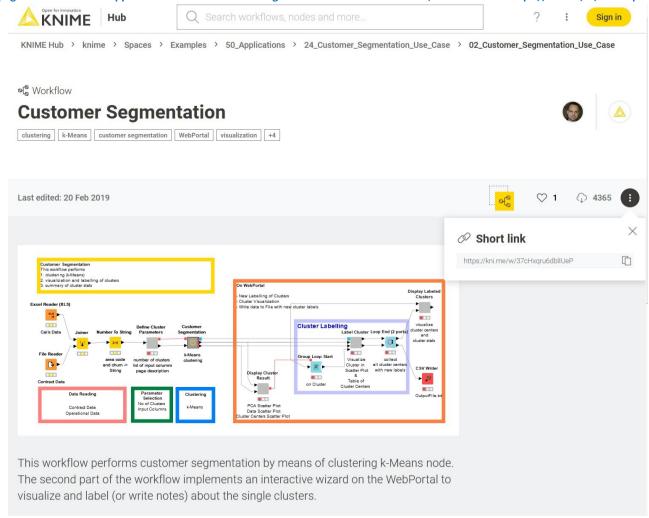
1.2. The list of applications (workflows) related (and tagged) with Customer Intelligence and available on the KNIME Hub

Clicking one of the applications in the list opens the corresponding web page (Fig. 1.3), with a nice explanatory picture of the implemented workflow.

On the top right corner, you can see the button to log in with the KNIME Forum account. Being logged in allows you to upload, download, comment, like, and update the spaces for which you have permissions. Below that, you can find the author picture and below that a number of utility buttons: to download the workflow, to like it, to drag & drop it onto your KNIME Analytics Platform installation, and the short permanent link to this workflow to share.

If you hover over the author image, and if you have editing permissions for this Hub space, a pen will appear. Clicking on it will allow you to allow other KNIME users to upload and change this space.

Notice that the KNIME Hub is a repository for workflows, but also for nodes, components, and extensions.



1.3. The page dedicated to the application named "Customer Segmentation" on the KNIME Hub, with short link https://kni.me/w/37cHxqru6dblIUeP

1.3. Download and install KNIME Analytics Platform

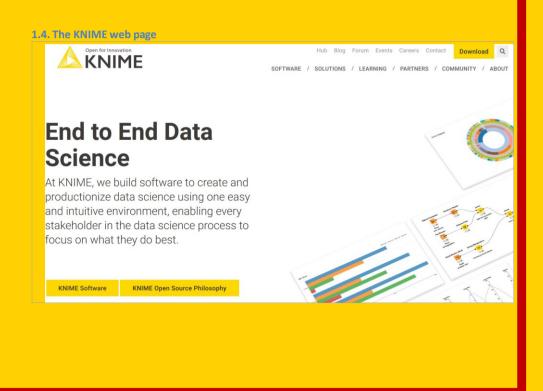
There are two available KNIME products:

- the open source KNIME Analytics Platform, which can be downloaded free of charge at https://www.knime.com/software-overview under the GPL version 3 license
- the KNIME server, which is described at https://www.knime.com/knime-server

Analytically speaking, the functionalities of the two products are the same. The KNIME Server in addition includes a number of useful IT features for team collaboration, enterprise workflow deployment and management, data warehousing, integration, and scalability for the data science lab. In this book, however, we will work with KNIME Analytics Platform (open-source). To start playing with KNIME Analytics Platform, first, you need to download it to your machine.

Download KNIME Analytics Platform

- Go to www.knime.com
- In the upper right corner of the main page, click
 "Download"
- Provide a little information about yourself (that is appreciated), then proceed to step 2 "Download KNIMF"
- Choose the version that suits your environment (Windows/Mac/Linux, 32 bit/64 bit, with or without Installer for Windows) optionally including all free extensions
- Accept the terms and conditions
- Start downloading. You will end up with a zipped (*.zip), a self-extracting archive file (*.exe), or an Installer application
- For .zip and .exe files, just unpack it in the destination folder. If you selected the installer version, just run it and follow the installer instructions.

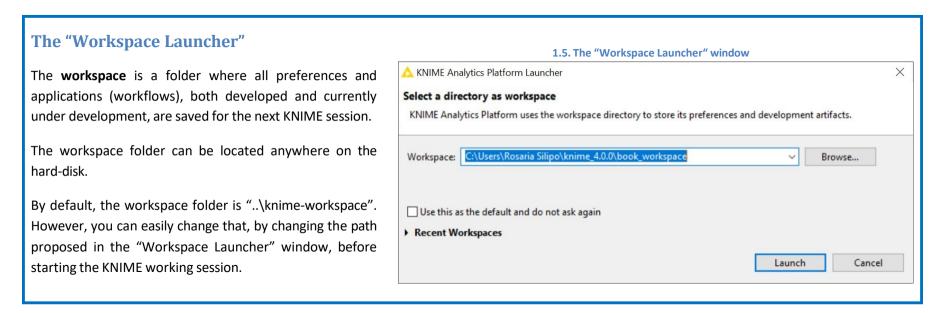


1.4. Workspace

To start KNIME Analytics Platform, open the folder where KNIME has been installed and run knime.exe (or knime on a Linux/Mac machine). If you have installed KNIME using the Installer, then you can just click the icon on your desktop or on your Windows main menu.

If you are starting KNIME Analytics Platform for the first time, you will be asked if you want to share your usage statistics with KNIME. These numbers will be used to fuel the best practice recommendation engine provided within KNIME Analytics Platform workbench: the Workflow Coach. No personal information will ever reach KNIME and your anonymous statistics will never be shared with anybody.

After the splash screen, the "Workspace Launcher" window requires you to enter the path of the workspace.



Once KNIME Analytics Platform has been opened, from within the KNIME workbench you can switch to another workspace folder, by selecting "File" in the top menu and then "Switch Workspace". After selecting the new workspace, KNIME Analytics Platform restarts, showing the workflow list from the newly selected workspace. Notice that if the workspace folder does not exist, it will be automatically created.

If I have a large number of customers for example, I can use a different workspace for each one of them. This keeps my work space clean and tidy and protects me from mixing up information by mistake. For this project I used the workspace "KNIME_4.3.1\book_workspace".

1.5. KNIME workflow

KNIME Analytics Platform does not work with scripts, it works with graphical workflows.

Small little icons, called nodes, are dedicated each to implement and execute a given task. A sequence of nodes makes a workflow to process the data to reach the desired result.

What is a workflow

A workflow is an **analysis flow**, i.e. the **sequence of analysis steps** necessary to reach a given result. It is the pipeline of the analysis process, something like:

- Step 1. Read data
- Step 2. Clean data
- Step 3. Filter data
- Step 4. Train a model

KNIME Analytics Platform implements its workflows graphically. Each step of the data analysis is implemented and executed through a little box, called **node**. A sequence of nodes makes a workflow.

In the KNIME whitepaper [1] a workflow is defined as follows: "Workflows in KNIME are graphs connecting nodes, or more formally, direct acyclic graphs (DAG)." (http://www.kdd2006.com/docs/KDD06_Demo_13_Knime.pdf)

Below is an example of a KNIME workflow, with:

- a node to read data from a file
- a node to exclude some data columns
- a node to filter out some data rows
- a node to write the processed data into a file

1.6. Example of a KNIME workflow



Note. A workflow is a data analysis sequence, which in a traditional programming language would be implemented by a series of instructions and calls to functions. KNIME Analytics Platform implements it graphically. This graphical representation is more intuitive to use, lets you keep an overview of the analysis process, and makes for the documentation as well.

What is a node

A node is the **single processing unit** of a workflow.

A node takes a data set as input, processes it, and makes it available at its output port. The "processing" action of a node ranges from modeling - like an Artificial Neural Network Learner node - to data manipulation - like transposing the input data matrix - from graphical tools - like a scatter plot, to reading/writing operations.

Every node in KNIME has 4 states:

- Inactive and not yet configured
- Configured but not yet executed
- Executed successfully
- Executed with errors

- → red light
- → vellow light
- → green light
- → red with cross light

Nodes containing other nodes are called **metanodes** or **components**.

Below are four examples of the same node (a File Reader node) in each one of the four states.

1.7. File Reader node with different states



1.6. .knwf and .knar file extensions

KNIME workflows can be packaged and exported in .knwf or .knar files. A .knwf file contains only one workflow, while a .knar file contains a group of workflows. Such extensions are associated with KNIME Analytics Platform. A double-click opens the workflow inside KNIME Analytics Platform.

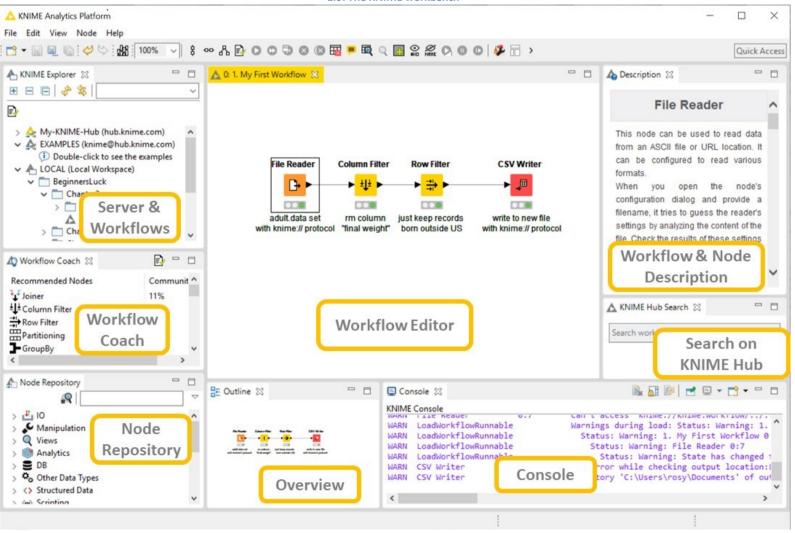
1.8. .knwf and .kngr files are associated with KNIME Analytics Platform. A double-click opens the workflow(s) directly inside the platform

△ 01_From_Strings_to_Documents.knwf	10/4/2017 9:45 AM	KNIME Workflow	18,619 KB
△ 04_Interaction_Graph.knwf	9/29/2017 8:20 AM	KNIME Workflow	9,465 KB
△ 06_REST_Examples_Google_Geocode.knwf	7/29/2017 7:09 PM	KNIME Workflow	62 KB
△ 06_Semantic_Web_updated.knar	11/3/2016 2:24 PM	KNIME Archive File	178 KB
🛕 AzureDemoWorkflowArchive.knar	5/5/2017 11:24 AM	KNIME Archive File	24,104 KB
🛕 Building a Simple Classifierknwf	2/18/2017 5:46 PM	KNIME Workflow	43 KB
△ Cookbook_Ch5.knar	11/24/2017 10:03	KNIME Archive File	477 KB
🛕 Cookbook_Ch6.knar	11/24/2017 10:26	KNIME Archive File	155 KB
△ Corsair.knwf	7/10/2017 4:20 PM	KNIME Workflow	106 KB

1.7. KNIME workbench

After accepting the workspace path, the KNIME workbench opens on a "Welcome to KNIME" page. This page provides a few links to get started, such as for example to the KNIME Hub, to some basic documentation, to the current courses and events, to available updates, and so on. The "KNIME Workbench" consists of a top menu, a tool bar, and a few panels. Panels can be closed, re-opened, and moved around.

1.9. The KNIME workbench



The KNIME Workbench

Top Menu: File, Edit, View, Node, Help

Tool Bar: New, Save (Save As, Save All), Undo/Redo, Open Report (if reporting was installed), zoom (in %), Align selected nodes vertically/horizontally. Auto layout, Configure, Execute options, Cancel execution options, Reset, Edit node name and description. Open node's first out port table. Open node's first view, Open the "Add Meta node" Wizard, Append IDs to node names, Hide all node names, Loop execution options, Change Workflow Editor Settings, Edit Lavout in Components, configure job manager.

KNIME Explorer

This panel shows the list of workflow projects available in the selected workspace (LOCAL), on the EXAMPLES server, on the Mv-KNIME-Hub (vour own space on the KNIME Hub), or on other connected KNIME servers.

Workflow Coach

This is a node recommendation engine. It will provide the list of the top most likely nodes to follow the currently selected node.

Workflow Editor

The central area consists of the "Workflow Editor" itself.

A node can be selected from the "Node Repository" panel and dragged and dropped here, in the "Workflow Editor" panel.

Nodes can be connected by clicking the output port of one node and releasing the mouse either at the input port of the next node or at the next node itself.

Node Description

If a node or a workflow is selected, this panel displays a summary description of the node's functionalities or the workflow's meta information

Search box for KNIMF Hub

To search for material on the KNIMF Hub

Node Repository

are available in your KNIME installation. It is something similar to a palette of tools when working in a report or with a web designer software. There we use graphical tools, while in KNIME we use data analytics tools.

Outline

This panel contains all the nodes that The "Outline" panel contains a small overview of the contents of the "Workflow Editor". The "Outline" panel might not be of so much interest for small workflows. However, as soon as the workflows reach a considerable size, all the workflow's nodes may no longer be visible in the "Workflow Editor" without scrolling. The "Outline" panel, for example, can help you locate newly created nodes.

Console

The "Console" panel displays error and warning messages to the

This panel also shows the location of the log file, which might be of interest when the console does not show all messages.

There is a button in the tool bar as well to show the log file associated with this KNIME instance.

Top menu File Fdit View Mew... Ctrl+N Alt+Shift+O. C Undo Ctrl+Z Console Save Ctrl+S Redo Ctrl+Y Description Save As... Error Loa Alt+Shift+Q, L Cut Ctrl+X Save All Ctrl+Shift+S KNIME Explorer Close All Ctrl+Shift+W Copy Ctrl+C KNIME Hub Search Paste Ctrl+V Recent Workflows Node Monitor Print... Ctrl+P Delete Delete Node Repository Import KNIME Workflow... Select All Ctrl+A Export KNIME Workflow... Outline Alt+Shift+O. O Export to SVG... Workflow Coach Switch Workspace Other... Alt+Shift+Q, Q Preferences Reset Perspective... Export Preferences... Ouick Node Insertion... Ctrl+Space Import Preferences... Install KNIME Extensions... Open KNIME log Update KNIME... Restart Exit File includes the traditional File commands, like View contains the list of all panels that can be opened in the **Edit** contains edit commands. "New" and "Save", in addition to some KNIME KNIME workbench. Undo and Redo refer to the last specific commands, like: performed actions. A closed panel can be re-opened here. Import/Export KNIME workflow... Cut, Copy, Paste, and Delete refer Also, when the panel disposition is messed up, the option "Reset **Export to SVG** Perspective" re-creates the original panel layout when the to selected nodes in the workflow. **Switch Workspace** workbench was started for the first time. Preferences with Export/Import Preferences Select All selects all the nodes of Install KNIME Extensions the workflow in the workflow Option "Other" opens additional views useful to customize the **Update KNIME** editor. workbench.

Help Node Configure... F6 Help Contents Execute F7 Search Execute All Shift+F7 Execute and Open Views Welcome Page Cancel F9 About KNIME Analytics Platform (Cancel All Shift+F9 Install New Software... Edit Node Name and Description... Alt+F2 Open First Out-Port View Shift+F6 Show Active Keybindings... O Open First View F10 Cheat Sheets... Update Metanode Links Ctrl+Alt+U Open Metanode Wizard... Show Node IDs Ctrl+Alt+W Hide Node Names Ctrl+Alt+O Link selected nodes Ctrl+L Unlink selected nodes Ctrl+Shift+L Ch Step Loop Execution Ctrl+Alt+F6 Pause Execution Ctrl+Alt+F7 Resume Loop Execution Ctrl+Alt+F8 Ctrl+Alt+F10 {} Select Loop Workflow Editor Settings... Component Usage And Layout > Configure JobMgr... Help Contents provides general Help about the Workbench, BIRT, and **Node** refers to all possible operations that can be performed on a node. A node can be: KNIMF Configured **Search** opens a panel on the right of the "Node Description" panel to Executed search for specific Help topics or nodes. Cancelled (stopped during execution) Reset (resets the results of the last "Execute" operation) Welcome Page (re-)opens the Welcome Page Given a name and description Set to show its View (if any) **Install New Software** is the door to install KNIME Extensions from the KNIME Update sites. Options are only active if they are possible. For example, an already successfully executed node cannot be re-executed unless it is first reset or its configuration has **Show Active Keybindings** summarizes all keyboard commands for the been changed. The "Cancel" and "Execute" options are then inactive. workflow editor. Option "Open Meta Node Wizard" starts the wizard to create a new meta node in the Cheat Sheets offer tutorials on specific topics: the reporting tool, csv, workflow editor. Plug-ins.