

Welcome to the third KNIME Newsletter!

Earlier this month we released KNIME 2.4. In addition to lots of additional infrastructure functionality, we are happy to announce more powerful PMML support, now including support for preprocessing and PMML editing, a series of nodes for XML processing, modular support for the most popular ensemble learning methods and parameterizable metanodes. And the KNIME product setup has grown, too: the KNIME Enterprise Suite now includes an entry level solution, the KNIME TeamSpace for small team collaboration. An interesting additional highlight is the ability to share metanodes and link them dynamically inside your own workflows. Please let us know if you are interested in a demo!

Enjoy! The KNIME Team

Customer Intelligence: competitive edge with KNIME

The value of creating new insight from data reaches far beyond Life Sciences R&D – and so does KNIME. One major cross-industry focus area for modern business analytics are customer-centered processes. “Having new fact-based insight about your customers’ needs, behavior and potential value gives you the only true edge that cannot be duplicated by your competitors,” says Phil Winters of CI Agenda, commonly known

throughout the industry as the “Father of Customer Intelligence”. And with over 300 customer

“KNIME is the modern business analytics platform other legacy and wannabe business analytics vendors dream to have.”

insight projects under his belt – gained through 30 years’ experience in senior positions at SAS® and, most recently, as strategic adviser for renowned customer strategy consultancy Peppers & Rogers Group – he should know!

“The business analytics process is a series of linked and iterative activities,” explains Phil. “Whether you use CRISP or SEMMA as your approach, each step requires a joined-up set of techniques.” Data access, data transformation, initial investigation, powerful predictive analytics, visualization and the ability to automate the re-application of that new insight back into production systems are the keys to success. “A modern business analytics platform is absolutely indispensable, and I have very strong thoughts on what is required!”

KNIME’s ability to satisfy many types of users has already been discussed (see Newsletter 2). But what about the development platform itself? According to Phil, “When today’s so-called legacy platforms were developed, new approaches were being devised simultaneously with the capabilities to be delivered: data access using scripted parameterization, data transformation with a programming language, analytics and visualization calling line-oriented object code, process automation via pseudo line-code generators, packaging using a wide variety of then-emergent (and therefore different!) coding standards, and so forth.” That was – and still is – the reality of legacy platforms. “When it finally gets up and running, it can work well – but in hindsight, all of those platform developers now long for the chance to start over.”

While the KNIME platform has the experience of years, it has been developed using a modern approach: visually building node and connector workflows used for all required techniques including application development and interactive and batch execution. “To become an expert in one of the legacy platforms, you literally need years to learn all the different coding forms. With KNIME, you can be up and developing sensible processes after watching one short video. What I particularly like is that What you Develop = What you Execute = What you Present. This is not a GUI editor built after-the-fact on top of some other code. Gone are the days when I had to somehow translate 6000 lines of code packaged in a macro language into a mocked-up PowerPoint visual to make it understandable to my business users – or

worse, to have a front end that permanently hid what’s going on under the covers!” enthuses Phil. “What’s particularly impressive is the packaging of repeatable code blocks. Whenever you make a workflow – no matter how simple or complex – you can package it into its own meta-node, making it shareable and reusable by others. Now, you can even keep a centrally maintained repository of such metanodes for use or access by different types of users.” (See 2.4 features article.)



Phil Winters

But what about feature functionality? “I’ve been truly impressed. KNIME’s open-source approach, combined with the easiest way I’ve ever found of incorporating R, is commendable. Or any other legacy application, for that matter: It even reads SAS datasets without requiring SAS! 90% of what you will need to do is already foreseen in KNIME, and for that specialist requiring the other 10%? SOMEWHERE in the community, it’s out there and just a new node definition away,” says Phil. “I always watch the Life Sciences industry for analytic trends. They have massive data, they have insight challenges on a very large



KNIME coming to San Francisco!

After a great event in Boston last year, we will be meeting you and our Life Science partners this year in the SF bay area on **July 28, 2011**.

For details see:

www.knime.org/LifeScienceDay2011

Upcoming User Training

If you want to learn more about how to use KNIME and KNIME Reporting, you can now enroll for one of our very popular monthly training courses.

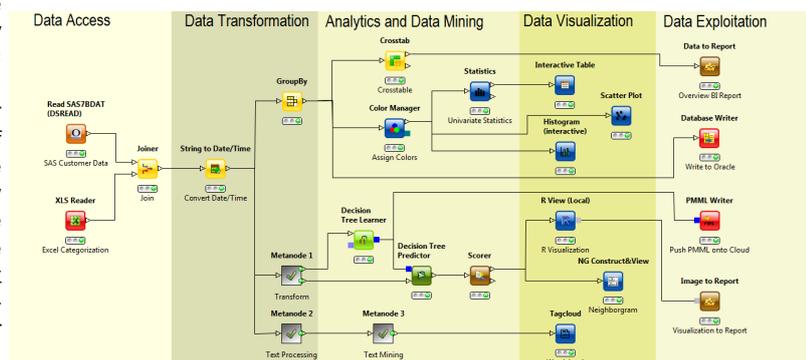
August 3 - 5, 2011

Technopark
Zurich, Switzerland
Visit:

www.knime.org/training
to register and for more information.

scale and they have always demanded professional and stable systems. Their take up of KNIME says it all.”

“Let’s face it: the proof is in the doing. I have now used the open-source (and free!) version of KNIME on successful customer intelligence projects ranging from banks to casinos. Recently at The Economist, we used KNIME to redefine the way they look at their extremely large base of worldwide loyal subscribers – both print and online. KNIME is the modern business analytics platform other legacy and wannabe business analytics vendors dream to have.”

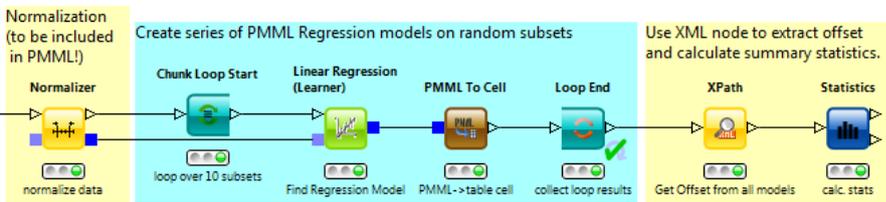


What's New in KNIME 2.4?

Selected New Nodes

A selection of new nodes to accommodate various preprocessing needs has been added in 2.4, among them a node to perform the popular CrossTab operation and the revised Webservice node, which was moved over from KNIME Labs.

The range of new KNIME nodes includes nodes for XML Processing, enhanced PMML support and also a series of nodes to model ensemble learning techniques, plus many other nodes for data preprocessing. The workflow on the top right demonstrates one of these new XML nodes in conjunction with our PMML extensions. First a normalization is applied and then a number of regression models are trained on the normalized data inside the chunking data loop. The resulting models are collected in a table containing a series of PMML models. From this table, the XML XPath node extracts a certain property from each model (the offset of the regression line in this particular case) and the final node calculates some statistics on this model parameter.



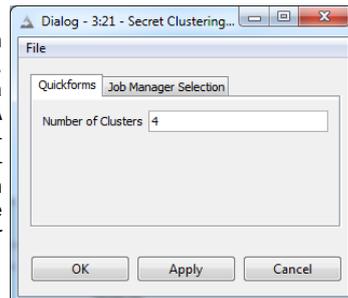
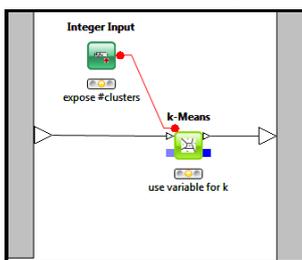
Featured Usability Enhancements

Collapse/Expand Metanode: With one click of your mouse it is now possible to collapse a selected set of nodes into a metanode. The newly created metanode automatically has the right number of ports. The corresponding expand action does exactly the opposite, moving all of the nodes inside a metanode into the parent workflow.

Workflow Autolayout: This is a feature that has been on our wish list for a long time. Select a subset of nodes or your entire workflow and let KNIME take care of the layout for you! Click multiple times to choose the layout you like best.

Metanode Templates & QuickForm Nodes: Package your Workflows

A completely new set of features is provided by our QuickForm nodes. Use them to visually specify which variables control execution of a workflow. This, in turn, can be used to expose a selected set of workflow parameters from inside a metanode on the outside. The screenshot on the left shows an example: A QuickForm node inside the metanode is applied to specify the number of clusters used by the k-Means clustering node. This parameter is subsequently accessible externally via the configuration dialog of the metanode itself (shown on the right). This new feature realizes powerful potential in conjunction with the KNIME TeamSpace and Server where this kind of metanode can be deployed for use by others and complex operations are hidden from view.



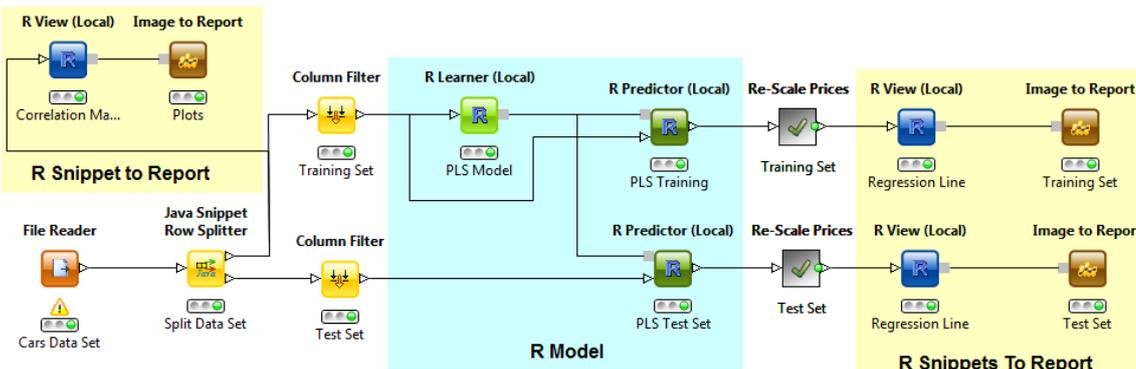
The New Enterprise Suite from KNIME.com



Professional Open-Source Analytics in a Modern Corporate Environment

KNIME.com continues to expand its enterprise offerings around the modern open-source data analytics platform KNIME. In addition to the open-source KNIME Desktop, you can now benefit from KNIME Professional which caters to providing support and maintenance for the open-source platform. KNIME TeamSpace leverages workflow sharing and linked metanodes among small teams, while the KNIME Server is an enterprise-scale solution, which hosts workflows featuring user roles and authentication, remote execution, and—now even more powerful with the new QuickForms, our KNIME WebPortal which allows workflows and reports to be accessed and parameterized from an easy-to-use web frontend. And finally, the KNIME Cluster Execution utilizes existing inhouse clusters to speed up computationally expensive workflows.

Tips & Tricks: Plots and Statistics with R



Michael Mazanetz
Computational Chemist
Evotec, UK
Certified KNIME Coach

Integrating R Models and Views in KNIME Workflows and Reports

The R integration into KNIME combines the wealth of KNIME's data mining tools with the efficient R functions for data analysis and manipulation, calculation and graphical display. R is supported by a strong development community. Coupled with R's own scripting language this gives an analyst the ability to mine data effectively and efficiently. The example above demonstrates how easy it is to use the R nodes in a KNIME workflow.

The R package comes with a variety of statistical modeling functions. Here we are interested in creating a least-squares model for some car attributes in order to predict the car price. We use a "Java Snippet" to split our data set into training and test sets. The "R Learner" node runs our PLS Model, which we use to predict our data with the "R Predictor" node and the "R View" node runs an R script to process the data into a graphic. KNIME's reporting features enable the results from the R calculations to be viewed easily. Three "R View" nodes are used to send results to "Image to Report" nodes. The final report is shown on the right and illustrates how data from a KNIME workflow can be pulled together and shared.

