KNIME News

July 2012 - Vol. 2, No. 2

KNIME News

In this issue, we highlight not only the newest version of KNIME - Version 2.6 - along with more great tips and tricks but also our KNIME Community, without which there would be no KNIME! The KNIME Community has some of the sharpest, most creative minds out there and we appreciate and value their commitment and support. This new feature in KNIME News gives us the opportunity to highlight the dedicated people who have made tremendous contributions by sharing their expert knowledge and creative solutions. It is also a fun way for our global community to get to know each other, promote further collaboration and share experiences. Speaking of community: come by and share your KNIME story or best practice with us in Boston or across Europe - including our first KNIME Users Day in Italy - throughout the fall!

Meet the KNIME Community

Our very first "Meet the KNIME Community" article features top community contributor Simon Richards. If you've spent any time on our KNIMEtech forum, you'll understand why we chose "richards99"!

Simon Richards
"richards99"
Medicinal Chemist
Eli Lilly

How long have you been using KNIME? For about 3 years now. When I started using it I didn't know anything about pipeline tools - KNIME was my first experience. Now I'm completely addicted to KNIME and what it can do for me.

What is it you find so addictive about KNIME? In my work as a medicinal chemist, you get so much data on the molecules you make that it can be overwhelming how to use it all. KNIME has made it so much easier to pull it together in an automated way, saving a lot of time. Because of this, it allows hypotheses for future target designs to be developed in a much shorter time period, and to be very data driven in the approach. You can pull out trends in data so easily and that makes it very addictive. All this data allows creative solutions to be developed to solve problems. With all the nodes available to manipulate and manage data in so many ways, you can always get to the answer you need.

What have been your greatest successes with KNIME? There have been quite a few! But if I had to identify something I would point to my presentation at the KNIME User Group Meeting on Matched Pair Analysis Tools. It's one of the nodes Eli Lilly built and offered to the community for free. This opens a readily accessible avenue for medicinal chemists in gathering statistical analysis on single point changes in molecules and how this affects the data. Doing this previously was incredibly time consuming and often impractical to do. Another big success in KNIME is pulling data together from different locations. In Pharma, data can often be generated at multiple sites and stored separately to measure the same property. KNIME easily combines multiple data sets together, allowing data trends in molecules to be identified more easily. Finally, a third success of KNIME is generating correlation models on data. This is really useful for predicting one property of a molecule by measuring a different property. KNIME has so many types of predictive models available, it is usually possible to generate some highly predictive tools for chemists to use. But the most important success of KNIME is in bringing all these tools to the medicinal chemist to use for themselves, and this therefore helps make us all much more statistically aware of how to use and analyse data.

Since joining the forum, Simon has become our top contributor with 568 entries, contributing 7.4% of all comments and 9% of all posts - not to mention hundreds of tips and tricks as well as a new KNIME sample workflow showcasing the use of the community nodes from Erlwood, Indigo and RDKit (available on the KNIME public download site).

What do you find most helpful in your day-to-day usage of KNIME? Bernd (Wiswedel)! But after Bernd, I would say the entire Forum Community, which is extremely active. There is a good mix of different types of users; from other chemists, to statisticians, to biologists, to IT-orientated users. Everyone is always handy in troubleshooting and provides varied responses and solutions. Being a varied community group means we get to see many different problems. The forum shows there is more than one way to solve a problem with KNIME.

What motivates you to contribute and share so much with our KNIME community? I enjoy solving other people's problems. This helps me learn new things, which may eventually give me the answers I need further down the road. I am grateful that KNIME is offered as a free download and feel a duty to help out where I can. We are all in this together... to make it better and help each other. If it's commercial you are less inclined to solve problems as this is part of the paid-for support; being free, it's good to provide something back to the community and the KNIME team.

What's your biggest KNIME desire? When are you converting KNIME to the iPad? That way, I could KNIME on the move. Also, are there any plans to open a KNIME-oholics clinic to help those who have become addicted to KNIME? I need some help in this regard!

What do you do when you are not KNIMEing? I'm happy chilling out at the weekends actually. I spend time gardening, and also enjoy cooking. I live in England, so when we get a brief (usually very brief) spell of nice weather, I enjoy walks in the countryside. Getting away from busy towns and seeing nature and its wildlife is always relaxing. I'm also very fond of cats. I have six that keep me busy in the evenings.

KNIME news

KNIME Rated No. 1 in Customer Satisfaction

THANK YOU to all our KNIME users! KNIME scored number one in overall satisfaction for an open source data mining platform. KNIME took top marks of ALL packages for:

- *Best in Ease of Use*
- *Best User Interface*
- *Best Data Manipulation Capabilities*

Visit: KNIME Satisfaction

KNIME Life Science User & Community Meeting

September 28, 2012
Boston
Microsoft New England Research & Development (NERD) Center

Join KNIME developers, users, contributors and partners for a complimentary get-together in Boston's NERD Center

Visit: KNIME LSD 2012

Come and Meet KNIME !

KNIME Italian User Day
October 9, 2012
Milan
Visit: KNIME Italia

Predictive Analytics World
Sept. 30 –Oct 4, 2012
Boston
Nov. 6 & 7, 2012
Düsseldorf
Nov. 27 & 28, 2012
London

DMEXCO
September 12 –13, 2012
Köln

Visit www.knime.com for more information on these and other upcoming events.

KNIME 2.6 - Extensions and Enhancements for all Types of Users!

KNIME 2.6 is now available! This newest release has a large number of extensions and new nodes, many based on requests from our community. For developing KNIME workflows, we've added the ability to add/remove ports on metanodes. There are also enhancements to the GUI editor such as a grid layout and snapping, node movement with arrow keys, and storing the zoom level that make layout and structuring of workflows even easier. For the heavy coder, extensions to both the String Manipulation and a powerful new Java Snippet node, along with a GroupLoop node for iterating over groups will be appreciated. We've also made the automating of workflows even stronger by improving dynamic column handling in many nodes.

New nodes include those for hypothesis testing such as *T-test* and *Anova*, new graphics nodes such as JFreeChart's Heatmap, and new data processing nodes like Column Aggregator and Empty Table Switch. The newest WEKA release 3.6 is also fully supported.

One major focus has been to further extend KNIME to business and non-specialists users. For creating packaged workflows on the desktop and via the web portal, there are now many more QuickForm nodes such as List Box selection, Check Box and Column Filter. For those surfacing KNIME to the occasional KNIME workflow builder via Teamspace and Server, there is now the ability to create a customizable node repository surfacing only the nodes a user requires. On the Server, you can also use the new QuickForms in the WebPortal. In addition, we've added an execution wizard to support in testing workflows via the KNIME WebPortal.

These are just a few of the highlights. For more information, check out the entire list of features at www.knime.com/Version260. Or better yet, just download and use the new version!

KNIME Open Source (KOS) Days

September 3-7, 2012
University of Konstanz
Konstanz, Germany

Join us at the University to develop KNIME projects in open programming sessions while enjoying beautiful Lake Constance.

Visit: KNIME KOS Days

Upcoming Training Dates

KNIME Developer Training
August 20-21, 2012
Technopark

KNIME Basic & Advanced User Training
September 17-19, 2012
Technopark
Zurich, Switzerland

Enroll now in our Developer Training for a comprehensive overview of the KNIME architecture. Our Basic & Advanced courses are an ideal introduction to KNIME and its effective use.

Introducing Peter Ohl: KNIME's Man Behind the Server


Peter Ohl

Born and raised in Southern Germany, Peter has worked for KNIME since its founding. But the relationship goes back much further. Peter has known Michael Berthold since 1982, when the two met in high school and happened to study together at the same university in Karlsruhe. Their paths crossed again a few years later in California although they were working for different companies. Once they were back in Germany, Michael approached Peter with his ideas and vision for KNIME and, joining forces with Bernd Wiswedel and Thomas Gabriel, they launched KNIME.
Since then Peter has not only been developing the KNIME server, which is his main role, but also handling the diverse and rapidly growing responsibilities in human resources, finance, legal and procurement. His intricate knowledge of the organization keeps KNIME's daily operations fine-tuned and running smoothly. Procurement becomes particularly important in summer when the Technopark offices become unbearably warm and the freezer needs to be kept stocked with Vanilla Macadamia Magnum bars!

Aside from events when the KNIME team gets to enjoy much needed social time together at sushi parties and BBQs, being on board from the beginning and working in a start-up environment with a small, cool, knowledgeable group is what is most satisfying for him.

Catching up with Peter in the evenings or weekends finds him keeping life simple. He is happy not to drive as he commutes from Konstanz in Southern Germany to Zurich. A typical weekend for him is hanging out, relaxing and enjoying time with his family. His love of cooking developed from a joy of eating, which he tries to counter by playing weekly on his handball team. What would Peter be doing if he weren't KNIMEing? He would move to Italy, close to the sea, build a wood shop and create beautiful things out of nothing. But then he'd probably get bored after three months and return to start programming again. And of course, he would miss Michael. You can't just break up a relationship like that after so many years!

Tips and Tricks—Extracting Cluster Information from a k-Means Model


Rosaria Silipo, Data Mining Consultant, Zurich, Switzerland

Being able to use the k-Means cluster information for further processing or alternate visualization can be helpful. The information is output by a k-Means node within the PMML model. To use this cluster information, we first convert the PMML model into an XML table cell with the **"PMML To Cell"** node. Next, begin with XML parsing of the cell content. If you were to write the XML content of the cell into a file with the "XML Writer" node, you would see that the root is <PMML>, followed by <ClusteringModel>, and then by <Cluster>. A <Cluster> contains all information for a particular cluster of the k-Means model. This is the information we will extract.
The **"XPath"** node implements this XPath query "/dns:PMML/dns:ClusteringModel/dns:Cluster"and extracts all <Cluster> into one cell. The **"Ungroup"** node assigns each <Cluster> to one data row. And finally a number of "XPath" nodes extracts further information about each cluster. For example, an "XPath" node with XPath query "/dns:Cluster/@name" extracts the cluster name while an "XPath" node with XPath query "/dns:Cluster/dns:Array/text()" extracts the text content of <Cluster>, in this case the prototype values of the used attributes. The attribute values are all concatenated together in a long string. Now use a "Cell Splitter" node to get the individual values.

To the right is the sub-workflow used to extract the information about the k-Means prototype from a PMML model produced by a k-Means node. You can find this example (013003_XML_Processing_K-means_centers) and others by connecting KNIME to the public server.