

Welcome to the first KNIME Newsletter of 2012!

There was a lot going on during the first quarter of 2012. The results of the most recent Rexter Analytics Survey have indicated that KNIME is growing strongly in data analytics setups and report excellent user satisfaction across all domains. However, our 4th KNIME User Group Meeting was undoubtedly the highlight of the past quarter! Not only did it have the highest attendance ever, but we also saw various presentations about the productive use of KNIME and our Enterprise tools in diverse application areas. In addition, eight KNIME partners showed their offerings around KNIME in our exhibition area and the demo of upcoming KNIME improvements raised lots of interest—look out for information in future KNIME Newsletters! You can read more about the UGM on the second page of this newsletter.

Social Media: Network Analytics Meet Text Mining

In today's world of social media with its wide variety of social media channels, there is a huge amount of data available. The challenge comes in accessing that data and transforming it into something that is usable and actionable.

"It doesn't rely on a thumbs up to rate importance, but rather identifies influence."

KNIME is the ideal platform for organizations that want to use the social media data to understand the needs and behavior of their customers or specific targeted groups of individuals with respect to the company's current or future products or services. "For a major European Telco, we used KNIME to combine text mining and network analysis to generate new fact based insight." Says Kilian Thiel, University of Konstanz. "Since those results are confidential, but the techniques used are very good, we created a white paper using publicly available data."

Generally, social media data contained in public forums is now accessible through standard APIs or tools that allow the data to be downloaded from the cloud. There are also specialized service providers that will scan and deliver all the data in a usable form. Text and network mining in social media data as analytic approaches are now widely used to reveal new insights in the data. In texts from blogs, comments or posts, the sentiments and opinions of users on certain topics, products, or persons are often mined. However, each technique follows its own very specific goal.

"In text mining, the emphasis is on translating the textual data into sentiment in a carefully controlled process that places the emphasis on words and expressions within a given context," says Thiel. "However, the information about the actual creator of the text, the sentiment expressed in it, and the counts and numbers of readers and responders cannot reveal the relevance of that person with respect to all others in that community, nor can it reveal the relative interactions between that person and others how they relate to each other."

"Network mining, on the other hand, does a very good job of identifying how individuals interact

with each other," says Tobias Kötter, University of Konstanz. "It does not rely on a categorical captured "thumbs up" or "star rating" of individuals to rate the importance of a person, but rather identifies those people of influence and those that are followers through physical nodes and connectors."



Dr Kilian Thiel, Fellow, University of Konstanz



Dr Tobias Kötter, Fellow, University of Konstanz

"Taking advantage of having both network mining and text processing available within the KNIME environment, we combined the results from the sentiment analysis with the results from the network analysis in order to better position each user inside his/her community in terms of influence—leaders vs. followers—and sentiment—positive, neutral, and negative users," says Kötter. "You can see one result of that here." (See Fig. 1.)

"KNIME is a fantastically flexible platform for combining these techniques and we can highly recommend it," says Thiel. "We'd like to thank KNIME for helping us publish these results. You can find more details along with workflow examples on the KNIME website."

www.KNIME.org/white-papers

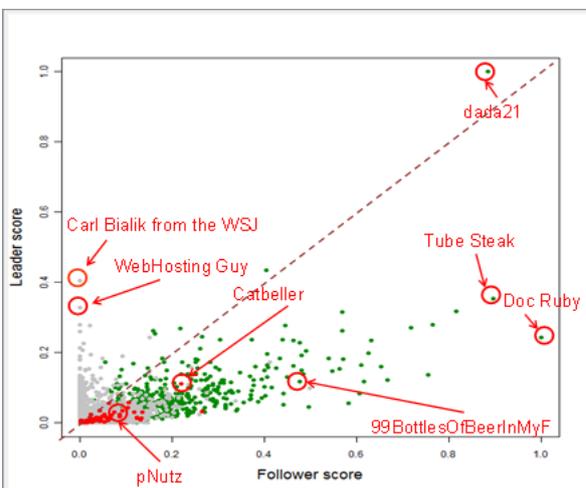


Fig. 1 Leader and follower scores



Visit KNIME at Predictive Analytics World

Boston **Sept 30–Oct 4**
Düsseldorf **Nov 6-7**
London **Nov 27-28**

See: www.predictiveanalyticsworld.com

Upcoming User Training

KNIME User Training

If you want to learn more about how to use KNIME and KNIME Reporting, you can now enroll for another one of our well known training courses.

May 7-9, 2012
Technopark
Zurich, Switzerland

July 2-4, 2012
Oxford, UK

Visit: www.knime.org/training to register and find more information.

Next Developer Training

During this course you will be given a comprehensive overview of the KNIME architecture and learn how to extend KNIME by programming your own custom nodes.

June 11-12, 2012
Technopark
Zurich, Switzerland

Visit: www.knime.org/training to register and for more

KNIME Cookbook

Recipes for the Advanced User. The long-awaited sequel to KNIME Beginner's Luck is now available. Purchase the new KNIME Cookbook: www.knime.org/knimepress and start mining your data today!

The KNIME Users Group Meeting in Zurich

Over 130 KNIME fans came together for the 5th KNIME user group meeting held Feb 1-2 in Zurich. Attendees represented a wide range of industries, covering everything from life science through retail, telco and banking to entertainment and government. Day one, designed for all attendees, started with an update of KNIME.com AG, which was able to report a strong increase in not only open source customers, but a significant number of new small and medium sized customers as well as three new customers doing major worldwide rollouts. "Thanks to our professional customers, we can report a growing and profitable organization in 2011 that will continue to invest in R&D for all customers - including our open source users," said Michael Berthold, CEO of KNIME.com AG.



UGM, Technopark in Zurich

The opening session included a detailed look at functionality enhancements over the last year. "Our emphasis was not only on new nodes and techniques, but enhancing the usability for the workflow builder and adding additional functionality for KNIME in the enterprise environment," said Bernd Wiswedel, CTO. KNIME development does not stop, and there was a brief look at the strategic development directions of KNIME.

A major topic for KNIME users and industry in general is the topic of **Big Data**. Mike Hoskins of Pervasive, one of the leaders in Big Data Processing and an integrated KNIME partner, was able to highlight not only the challenge that big data brings to organizations, but how the KNIME/Pervasive combination tackles the challenge. It was made even more impressive by running - live - a complex processing of 126 million rows faster than Mike could describe what was happening. The day continued with a ground-breaking session that combined both text mining and network analytics on social media data to produce new actionable insight that had never been seen before (see previous page). The second day of the meeting was dedicated to Life Science users and our partners in this industry gave us reports on how their large-scale applications are using KNIME.

An exciting User Group Meeting came to a close with special focus sessions on text mining, network analytics and image processing, just to name a few. For those that missed the user group and are interested in the presentations, send an email to info@KNIME.com. And for those that did attend the user group - those strange things in your attendee bag were KNIME logoed baggage straps, to help take care that all the goodies collected at the conference got home safely. The next KNIME User Group Meeting is scheduled for 6-7 March, 2013 in Zurich, Switzerland. Mark your calendars now!

Tips and Tricks—String Manipulation

String Manipulation in KNIME
 In KNIME, string manipulation nodes are spread across sub-categories of the Column category, like the String Replacer node in the Convert & Replace sub-category or all the Cell Splitter nodes in the Split & Combine sub-category. However, starting with KNIME 2.5, one very powerful node has been made available for string manipulation under the Transform sub-category. Together with the nodes mentioned above, the String Manipulation node can probably solve all your string manipulation needs.



Rosaria Silipo, Data Mining Consultant, Zurich, Switzerland

The String Manipulation Node

The String Manipulation node is very powerful, encapsulating more than 10 string manipulation functions. It can implement functions, for example, to calculate a string length, to compare two strings, to change a string into only uppercase or lowercase characters, to replace all the occurrences of a substring inside a string, to capitalize the string words, to find the positions of a substring occurrence, to extract a substring from a string, just to mention a few!

The power of the String Manipulation node does not reside solely in the number of functions that it can implement, but also in its flexibility to combine them to create arbitrarily complex string manipulation functions. For example, if you want to combine three strings - where one is a constant string like "_new_", one is the whitespace stripped and reversed version of the first input string, and one is the uppercase version of the second input string - all you need to do is appropriately combine the functions **join**, **reverse**, **strip**, and **uppercase** in the expression editor of the configuration window of the "String Manipulation" node.

The configuration window of the String Manipulation node has an Expression Editor in the central bottom part. Here a number of string functions can be combined to obtain the desired string transformation.

The available string functions are listed above in the Function list panel. The Description panel on the right explains the task of the selected function. Functions can also be visualized in smaller groups, by selecting a category in the Category list menu over the Function list panel.

On the left, in the Column List panel, all available input data columns are displayed. Double-clicking a column or a function automatically inserts it in the Expression Editor.

Constant string values have to be reported in quotation marks, for example "new", when introduced to the Expression Editor.

The Insert Missing As Null flag enables a null string to be produced instead of an empty data cell when there are missing values in the input data.

The configuration window finally requires the name of the new or existing column, depending on whether the resulting string has to overwrite existing data or not.

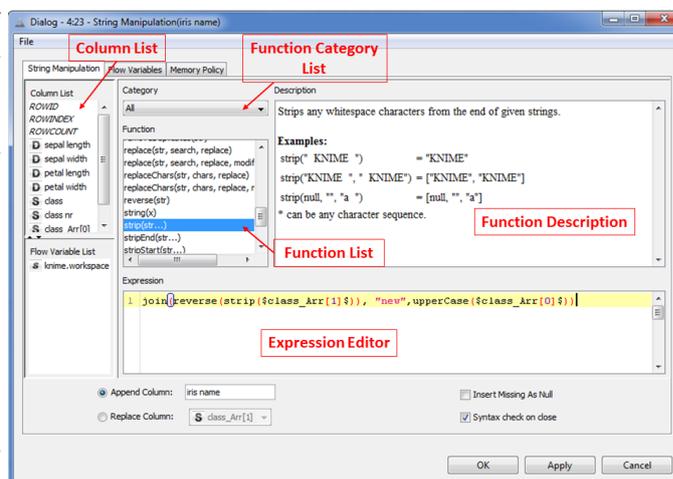


Fig. 2 Configuration window of the String Manipulation node