What’s New

Bernd Wiswedel
KNIME
Outline

• “What’s new“ presented in two use cases, presented by the team

• Questions/Discussions/Concerns: Find us!
• Demo booths in the registration area
What’s new - Agenda

- Too much to show it all.
- Some highlights:
  - Audio Processing
  - Text Processing
  - Analytics & Scripting
  - KNIME Server
  - Cloud Offerings
  - Guided Analytics

Use Case 1: IMDB Reviews

Use Case 2: Census Data
KNIME Software Pieces
What’s new - Agenda

- Too much to show it all.
- Some highlights:
  - Audio Processing
  - Text Processing
  - Analytics & Scripting
  - KNIME Server
  - Cloud Offerings
  - Guided Analytics

Use Case 1: IMDB Reviews
Use Case 2: Census Data
KNIME as an Open (Source) Platform
KNIME Analytics Platform – Source Code on GitHub

- Source Code available on GitHub & BitBucket
- Ongoing effort, enough to get started

https://github.com/knime/
SVN To Git – Fun Facts

• Switched from SVN to Git earlier the year
• ~3 weeks conversion time (parallelized over 10 physical machines)
• Split into 70 repositories (60 open source)
• Code Base:
  – ~20GB (includes text models, R, etc)
  – ~1.7 Mio Lines of Code (open source only)
  – 1500 integration tests per version (excl. unit tests)
KNIME Analytics Platform – Nightly Builds

• Nightly Builds publicly available:

Really, really, really important disclaimer

This is most definitely not production quality code. These nightly builds are recent developments, so they are not tested as thoroughly as stable versions. Functionality may change substantially (or disappear entirely) with each nightly build.
KNIME Personal Productivity – free & open-source

- Part of the KNIME Analytics Platform 3.4

- What is it?
  - Workflow Diff Viewer
  - Metanode Templates
  - Workflow Linking
    (“Call Local Workflow”)
  - “Local” Workflow Coach
KNIME Personal Productivity – Workflow Diff
KNIME Personal Productivity – Workflow Linking

- Workflow Orchestration by calling out to other workflows
KNIME Personal Productivity – free & open-source

• Part of the KNIME Analytics Platform 3.4

• Big Data Extensions in December (3.5)
Speech Recognition
– Tobias Koetter –
KNIME Audio Processing

- New audio cell type
- Listen to audio files within KNIME
- Extract acoustic features
- Speech recognition


https://www.youtube.com/watch?v=1L6zwLMvA-U
Example Use Case

• Sentiment classification of movie reviews
  – PDF parsing
  – Feature vector creation
  – Predictive modeling
Example Use Case

Goal

• Build predictive models to predict sentiment labels of reviews, “positive” or “negative”.

• “Ah, Moonwalker, I'm a huge Michael Jackson fan ...”

• “This film has a very simple but somehow very bad plot. ...”

Positive or negative?
Data Set

• The Large Movie Review Dataset v1.0
  – English movie reviews
  – Associated sentiment labels “positive” and “negative”
  – http://ai.stanford.edu/~amaas/data/sentiment/

• 2000 PDFs
  – 1000 positive reviews
  – 1000 negative reviews
Textprocessing

• Tika integration
• StanfordNLP NER Learner & Tagger nodes
• Document Vector Hashing
• Word2Vec integration
Analytics & Scripting
– Bernd Wiswedel –
Analytics

• Logistic Regression - new solver

New Integrations – more on this tomorrow
- Deep Learning via Keras
- H2O AI Platform

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:30 am</td>
<td><strong>KNIME Deep Dive: Deep Learning (and some Images)</strong></td>
</tr>
<tr>
<td></td>
<td>Christian Dietz (KNIME)</td>
</tr>
<tr>
<td>12:00 pm</td>
<td><strong>Leveraging KNIME for Fire Analysis and Public Safety</strong></td>
</tr>
<tr>
<td></td>
<td>Mohammed Ayub (National Fire Protection Association)</td>
</tr>
<tr>
<td>12:30 pm</td>
<td><strong>Guided Analytics in Action: Census and Sensibilities</strong></td>
</tr>
<tr>
<td></td>
<td>Greg Landrum (KNIME)</td>
</tr>
<tr>
<td>1:00 pm</td>
<td><strong>Lunch Break - Picnic Ziler Buffet: Tejas Conference Dining</strong></td>
</tr>
<tr>
<td>2:00 pm</td>
<td><strong>Integrating High Performance Machine Learning: H2O and KNIME</strong></td>
</tr>
<tr>
<td></td>
<td>Mark Landry (H2O) &amp; Christian Dietz (KNIME)</td>
</tr>
<tr>
<td>2:30 pm</td>
<td><strong>Citizen Data Science Initiative at Seagate</strong></td>
</tr>
<tr>
<td></td>
<td>Allan Luk (Seagate Technology)</td>
</tr>
</tbody>
</table>
Scripting – Python (Labs)

• Supports Python 3
• Major Speedup
• Used for new Deep Learning Integration
Type Extensions in Java Snippet

- Java Snippet: Swiss Army knife for data manipulation
- Read/write support for 3rd party types: SVG, XML, JSON, Images, Molecules, ...

```java
// expression start
long numAtoms = c_smilesRDKitMol.getNumAtoms();
RDKFuncs.assignStereochemistry(c_smilesRDKitMol, false, true, true);
long nUnspecifiedChiralCenters=0, nSpecifiedChiralCenters=0;
for (int i = 0; i < c_smilesRDKitMol.getNumAtoms(); ++i) {
    if (c_smilesRDKitMol.getAtomWithIdx(i).getChiralTag() != Atom.ChiralType.CHI_UNSPECIFIED) {
        ++nUnspecifiedChiralCenters;
    } else if (c_smilesRDKitMol.getAtomWithIdx(i).hasProp("_ChiralityPossible")) {
        ++nSpecifiedChiralCenters;
    }
    numAtoms = numAtoms;
    nUnspecifiedChiralCenters = nUnspecifiedChiralCenters;
    nSpecifiedChiralCenters = nSpecifiedChiralCenters;
}
```
Date & Time Handling Nodes revised

- Richer data types
- Timezone support
- Duration “math”
Three new helpers!!

- Generate a file path based on folder, file name and extension type
- Simple changes to string based variables
- ... and for number variables as well.
... and now for something completely different

– Rosaria Silipo –
E-Learning
Free E-Learning Course: Web Page

• Hands-on e-learning course
• Data Access, ETL, Analytics, Control Structures, Visualization
• Around 50 small units
• ... with exercises
• ... and with solutions on the EXAMPLES server
• Final exercises to test your knowledge!

https://www.knime.org/knime-introductory-course
KNIME Text Mining Training

This one day course is an intensive training focused on the processing and mining of textual data with KNIME using the Textprocessing extension. Learn how to read textual data in KNIME, enrich it semantically, preprocess, and transform it into numerical data, and finally cluster it, visualize it, or build predictive models. Text mining experience is not necessarily required for this training.

<table>
<thead>
<tr>
<th>Duration</th>
<th>1 day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs</td>
<td>€ 750 per user</td>
</tr>
</tbody>
</table>

Topics

1. Introduction to KNIME
2. Reading and Importing Textual Data
3. Text Preprocessing, Semantic Enrichment, and Transformation
4. Text Classification
5. Visualization
6. Text Clustering

PDF files for KNIME Text Mining Training material can be downloaded here.
E-Books

At the KNIME Press:

https://www.knime.com/knimepress
On-line Learning
On-line Courses

• On-Line Courses vs. On-Site Courses with Teacher
• First Experiment with Joe Porter’s group in November/December
• Basic Course on KNIME Analytics Platform
• With Homework and Homework Correction

Send us an email: education@knime.com
Webinars

In 2018 Series of Webinars:
• Data Science Cycle
• Introduction to KNIME Analytics Platform
• Data Access
• Database
• Logistic Regression
• Ensemble Models
• Text Processing
• ...

Keep an eye on our web site www.knime.com
Learning @Meetups
Learnathons

Data Science Cycle: ETL, Model Training, Deployment

Locations: London, Berlin, Zurich, Milan, New York, Austin, Sao Paulo, Somewhere on WC
KNIME Server
– Jon Fuller–
KNIME Software Pieces

Open Source Extensions

KNIME® Analytics Platform

Community & Partner Extensions

KNIME® Server
KNIME Server

Shared Repositories
- Data
- Partitioning
- My Predictor
- Nodes & Metanodes
- Workflows

Access Management
- Nodes, Files, Applications
- Java® Powered

Web Enablement
- Content, Services
- REST API

Flexible Execution
- Remote, Scheduled
Deploy to Server
Open in WebPortal

WebPortal Explorer

- knime-server (Emil@https://knimeserver:8443/toomee/sjb)
  - Deployed Workflows
    - 1. Train model (on schedule)
  - Examples
    - Orchestration
    - Reporting
    - REST
    - Shared Metanodes
    - Supplementary Data (really only the data)
    - Test Workflows (add your own for databases)
  - WebPortal
    - 01 - WebPortal Data Mining (Basic)
    - 02 - WebPortal Customer Segmentation
    - 03 - WebPortal IF statements

Examples (knime-guest@http://publicserver)

- LOCAL (Local Workspace)
  - Example Workflows
  - Predict Income
    - data
      - 1. Train model (on schedule)
      - 2. Predict Results on a Schedule
      - 3. Predict Results Using REST API

New Workflow Group...
Deploy to Server...
Execute...
Open in WebPortal
Authentication

- Authentication with LDAP/Active Directory.
- Single Sign On (SSO) with Kerberos
KNIME Server – Admin made easy

Status Info

Host: FordPrefect
Version: 4.6.0
Uptime: 0 days, 2 hours, 34 minutes, 54 seconds,
Executors Info: KNIME Executors associated with user: (none): 0: RMI Executor 0 (user: (none), port: 50100, uptime: 02:32:55, status: AcceptingJobs, #jobs: 0)
Config Info ➤

Download Server Log Files ➤

License and Users Info

License Type: KNIME Server
Expiration Date: 2018-11-19
Company: KNIME AG
Customer: Jon Fuller
Comment:

Users ➤
KNIME in the Cloud
– Jon Fuller –
Why cloud?

•  Bring analytics to the data
•  Work fast, work agile
•  Pay for what you need
•  New ways to work
Cloud Platforms

- AWS and Azure
- The two leaders in Cloud (see Gartner 2016)
KNIME Cloud Products (Marketplace)

- KNIME Cloud Analytics Platform
- KNIME Server (5-User)
- KNIME Server + Big Data Extensions (5-User)
- KNIME Server (+ Big Data Extensions) BYOL
KNIME Server in the Cloud

About KNIME
At KNIME, we build software for fast, easy and intuitive access to advanced data science, helping organizations drive innovation. Our KNIME Analytics Platform is the leading open solution for data-driven innovation, designed for discovering the potential hidden in data, mining for fresh insights, or predicting new futures. Organizations can take their collaboration, productivity and performance to the next level with a robust range of commercial extensions to our open source platform. For over a decade, a thriving community of data scientists in over 60 countries has been working with our platform on every kind of data: from numbers to images, molecules to humans, signals to complex networks, and simple statistics to big data analytics. KNIME’s headquarters are based in Zurich, with additional offices in Konstanz, Berlin, and Silicon Valley. We are open for innovation, so visit us at KNIME.com.

KNIME Products (3)

KNIME Server + Big Data Extensions

KNIME Server (BYOL)

KNIME Server
Cloud Connectors

- Enable working directly with data on the AWS or Azure clouds

  - S3
    - Amazon S3
  
  - Blob Storage
    - Microsoft Azure Blob storage
  
  - SQL Server
    - Azure SQL DB
KNIME Cloud Connectors

- Amazon S3/Azure Blob Storage: Save (large) unstructured data in the cloud
- Connect through KNIME’s file handling nodes
  - List
  - Upload/Download
  - Delete
  - URL access
Cloud Connectors Workflow Example

Amazon S3 Connection

Enter your own credentials here

Amazon S3 File Picker

Using the path we create a URL from which the CSV reader can read directly

File Reader

Read the CSV file

Process as a stream
Azure SQL Server Connector

Subscriptions: All 2 selected

Filter by name...

1 items

<table>
<thead>
<tr>
<th>NAME</th>
<th>STATUS</th>
<th>REPLICAION ROLE</th>
<th>SERVER</th>
<th>PRICING TIER</th>
<th>LOCATION</th>
<th>SUBSCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>demodb</td>
<td>Online</td>
<td>None</td>
<td>knime</td>
<td>Basic</td>
<td>West Europe</td>
<td>Microsoft Azure Spons...</td>
</tr>
</tbody>
</table>

SQL Server Connector
- Connect to Azure SQL DB

Database Table Selector
- Select Customer Table

Database Row Filter
- Filter only customers with suffixes in their names

Database Connection Table Reader
- Read data from database to KNIME Analytics Platform
More connectors: Amazon Redshift

- Fully managed data warehouse
- SQL interface
- Distributed and parallelized queries

- Amazon Redshift connector integrates seamlessly with existing KNIME Database connectors
Amazon Athena

• Serverless, interactive query service
• SQL interface

• Amazon Athena integrates seamlessly with existing KNIME Database connectors
Amazon Athena
S3/Blob Storage support for Spark Data Source Nodes
Cloud Resources

- **Data storage**
  - MS Azure Blob Storage, Amazon S3

- **Databases**
  - Azure SQL Server, Amazon RDS (MySQL, PostgreSQL etc)

- **Data warehouse**
  - Amazon Redshift, Amazon Athena, Azure SQL-DW

- **Hadoop/Spark**
  - Azure HDInsight, Amazon EMR
KNIME in the Cloud: Summary

- KNIME Analytics Platform available on Azure Marketplace
- KNIME Server available on Azure and AWS marketplaces
- Cloud connectors for S3, Blob Store, Azure SQL server
Guided Analytics

– Greg Landrum –
Reminder of what we’re talking about here

Someone who needs data / analytics

KNIME expert
What often happens

Someone who needs data / analytics

KNIME expert
What we’d like to provide

Someone who needs data / analytics

Guided Analytics workflows

KNIME expert
Guided Analytics workflows

• Enable people who are not KNIME experts to work with/explore/analyze/learn from data
• Typically deployed via KNIME WebPortal
• Heavy use of interactive JavaScript views for the user experience
Guided Analytics workflows

- Enable people who are not KNIME experts to work with/explore/analyze/learn from data
- Typically deployed via KNIME WebPortal
- Heavy use of interactive JavaScript views for the user experience

- Bonus for everyone: the JavaScript views that we’re building for Guided Analytics are also really useful in “normal” KNIME workflows.
So what’s new?

• JavaScript Views:
  – Parallel coordinates plot
  – Table view
  – Decision tree view
  – Network view
  – Stacked area chart
  – Sunburst chart

• Interactivity: Selection and filtering
  – Range filter widgets
  – Linked views
  – Accessing views in wrapped metanodes, composite view
  – Layout editor
But let’s start with a demo
The workflow behind the demo:
JavaScript Stacked Area Chart

Stacked Area Chart

- #analytics
- #bigdata
- #machinelearning
- @DMR_Rosaria
- @knime
- @KNIME
- @dr_greg_landrum
- #textmining
- #deplearning
- @JonathanCFuller

© 2017 KNIME AG. All Rights Reserved.
JavaScript Sunburst Chart
(Short) Coffee Break
- all of us -
The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME.com AG under license from KNIME GmbH, and are registered in the United States.