

Integrating Data, Tools, and Science

BioMax Symposium Munich 3. May 2012



Who is KNIME?



History

- 2004: KNIME Development started
- 2006: KNIME v1 released
- 2006: Spin Off in Konstanz, Germany
- 2006-2007: First Commercial Partners
- 2008: KNIME moves to Zurich
- 2010: Enterprise Product Suite released
- 2012: 5th KNIME UGM atmost triples attendance



Status Quo

- KNIME 2.5 released (Feb 2012)
- ~10,000 active users
 ~50% Life Science
 ~50% Business Intelligence, Analytics
- 15+ commercial partners
- Local distributors in Japan, Italy, China, Brazil, Portugal
- Global Support and Distribution Agreement with PerkinElmer



Status Quo

- KNIME 2.5 released (Feb 2012)
- ~10,000 active users
 ~50% Life Science
 ~50% Business Intelligence, Analytics
- 15+ commercial partners



- Local distributors in Japan, Italy, China, Brazil, Portugal
- Global Support and Distribution Agreement with PerkinElmer



What is KNIME?



The KNIME Platform





KNIME loads and integrates data from diverse data sources:

- Different data bases
- Various file formats (CSV, XML, SDF, etc.)







KNIME provides huge repository of modules for easy-to-use, modular

- Data preprocessing
- Data fusion
- Data transformation







Histogram

Node 0:1:18

Pie chart

Node 0:1:19





and insights into the learned models.

Interactive linking&brushing techniques allow for powerful exploration of models and data.





R Snippet (Local)

Node 0:1:21

Logistic

Node 0:1:22

Weka Predictor

Due to its open API and "node-in-a-sandbox"-approach additional (also external) tools are easily integrated,

e.g.

- Access to the statistics tool R
- Complete integration of the machine learning library WEKA
- Application area specific integration, e.g. CDK (Chemical Development Kit)

KNIME is Eclipse-based: Integrating other Eclipse projects such as BIRT, DTP, etc. provides even more functionality





Life Science Partners

- CambridgeSoft
- Schrödinger
- Symyx Technologies
- Tripos
- ChemAxon (via Infocom)
- CDK
- Molecular Discovery
- Treweren
- Tibco / Spotfire
- Chemical Computing Group
- BioSolve IT
- Cresset
- And others...



Additional Technology Partners

PERVASIVE

Unleash the Power of Your Data



- Pervasive
- Dotmatics
- Revolution Analytics
- Dymatrix



ATRI

LTING GROUP

• •••

• and others













d d

~~<u>~</u>

. f. a. ...

×00 🕬

O

SDF Reader

SDF

File Reader









ale			
Table "default"	-Rows: 9379 Spe	c - Columns: 2 Properties	Flow Variables
Row ID	aar Molecule	+ PUBOL.	







File			
Table "default"	-Rovs: 9379	Spec - Columns: 2 Prop	erties Flow Variables
Row ID	sor Holecul	le PUBOL	
	e. h.		

🔺 Table View - 2	:45 - Interactive Table (76 x 7)		. 0 X		Table View - 2:	44 - Interactive Table (9379 x 8)				×
Ele LEPae Nu 1	unting View Output				File Hilite Navig	ation View Output	1		1=	_
File Hilite Navig	jation View Output				Row ID	OS Molecule	PUBCH	S SID	S Source	4.1
Row ID	Of Fragment	Atom c	Bond c							â
Row9	N V N	11	12	, <mark>-</mark>	3246032_1		3246032	4252558	NCGC	
Row 10		7	7	ert 1	3234518_2		3234518	4239411	NCGC	
Row11		10	10	ctui	476416_3		476416	11112724	NCGC	Ŧ
3 X734* X704* 3 1113** 860* 3 1113** 860* 3 1113** 860* 3 1113** 860* 3 1113** 860* 3 1113** 860* 3 243** 110* 3 2112** 214* 3 212** 214* 3 212** 214* 3 212** 214** 3 212** 214** 3 212** 214** 3 212** 214** 3 214** 214** 3 214** 214** 3 214** 214** 3 214** 214** 3 214** 214** 3 214** 214** 3 214** 214** 3 214** 214** 3 214** 214** 3	SOC IP Ann B.B. SOC IP Ann B.B.			_						

Cell Assay Image Mining





Cell Assay Image Mining

- Various modules for segmentation and feature extraction
- (Interactive) Machine





Text Mining





Internet in control HLA Thr2 circul
 Irad cyclokii IL-18 kareini seperioti Th-1 arkum
 Vaccine escap affin brain, cyclokii cyclokii cyclokii cyclokii cyclokii cyclokii cyclokii Carlokii Carl

igida immangen CP-10 frongrocest ELISPOT II-15 international ES7-6 humoert radial MEE mack CD6+ A-induced perform OM-124 cellular: ONHM these intake profiler origin corropreserv T-SPOTTB cycpyrin alphaGSUFTCR8b mutat apoptosi neutrophil studi Rancial-12 Li -27 PmacL12 angles -specific vireniis QTF toler BES-1-deficient modifi lesion radiotherapi LKR-13 regul cultur miL-18 vervet InK retrovir polymorphneoforman mi EGT Lunorigenesi AIDS-CMVstromal MHV MMA CACLI o chang receitor EUF14 CMR mRNA mous promiscu loa DCSS8L-19 plut NILAK ag induc plaqu IGRA X-trradiation adult NYVAC-Cperson inhibit model FGF1 IRAK m BISM ROI case I hybrid VCP142 The local cetesoarcoma DNA-Creacing MDM murin specif VSV polit CD4panceat A11C0-40L treat CD40L method TDLN SGPSynergid GVHD donor gene breast legand predict SHU-9890 TABaha IF19 plut NILAK ag induc plaqu IGRA X-trradiation adult NYVAC-Cperson inhibit model FGF1 IRAK panceat A11C0-40L treat CD40L method TDLN SGPSynergid GVHD donor gene breast legand predict SHU-9890 the SHU bactilostenci gene-gen HUHCV Forschrift therapi signal acquare Li -18 CDNN to stediet, maind Adel L Durin tandrict science Non-responders MMW control CD98 viru A105 inicie metatata effus astrocyt popule ndog gammadelta week ut alert Median assai T-cell immun DNA develop administr subject growth CD72 sarco calcapt pml PBMC gain der epitop CD4 infect responders mice tuberculosi lung NKDC latent lipozom cancer interferon hepat caspase-1 HUV tumoor cancer interferon hepat caspase-1 HUV tumoor out alpha-640Cer monkie metanoma bLF express KKL DC combin HEE speci Lp18 B7-1 HDC B16

IL-18R count HIV-1-specific Salmonella test IL-12 vaccin activ pituitari HCV protect children rate fungal N-

11-Ibeta dai GM-CSFTh1 skin Candida treatment CTL patient individu dose pleural correl declin QFT-2G vitamin ASC-hypersensit anli-CD40 stimul abi IL-10 progress product anli-tumor IL-2 CMV-specific migrat toxin advanc GAL PDGF Listra glioma IM-gamma metatati BCG viral HIV-infected function IFI type IL-21 FBI antimetariat laboratori efficaci tuberculosit-specific intestin SPC detect cytokin resist inhibitori subtyp usund conjug IPN-ajphabeta interrupt Cg6 PHPC pro-IL-18 mortal rL-19 neopterin transpitar EL-4 Colon PEC IR-F1 Tablap1 IL-180p-Fc agreem complex AdV PCR BMT resolv acquisit matur GBV-C interleukin-18 infant fibroblast CKLB Fgh2b pylori AdS colon pEgr-IL-18-B7 Gag-specific microglia baboro C215Fab-SE radiosomsti diagnost CD11 macagu tumour tgAAC09 associ bacteria huma gene-modified blood HIV-specific restor respond AGM pregnanc OK-PSAresult lymphoma annual dual amino mechan CCL2LI hypoplasia boost IRD ednkt IRD editor.

Network Analysis

Network Reader K-Core Clustering







Life Science Case Study: Novartis

"We are very impressed with our early work with KNIME. Being able to satisfy the needs of our different types of users from within one modern platform is very important and is the ultimate measure of success." Andy Palmer, Global Head of Software and Database Engineering of Novartis





KNIME: Tool Integration





KNIME: Tool Integration





KNIME: Open Integration

- KNIME: Integration
 - Data: form one view for all activities (transformation, modeling, analytics, reporting)
 - Tools: open source and commercial (own and others)
 - Science/Expertise: archive and reuse workflows
- KNIME.com: Enterprise Scale Server tools
 - Workflow and template repository
 - Web based access to workflows and reports
 - Grid/Cluster Execution



Bioinformatics: The Future



Bioinformatics: The Future

- Research in bioinformatics:
 - New Algorithms from Bioinformaticians
 - New Problems from <u>Bio</u>informaticians



Bioinformatics: The Future

- Research in bioinformatics:
 - New Algorithms from Bioinformaticians
 - New Problems from <u>Bio</u>informaticians
- Future of Bioinformatics: Sustainable and modular Integration of new Tools and Methods:
 - \Rightarrow Continuous access to state of the art
 - \Rightarrow Simple transfer to new applications
 - \Rightarrow Synergy



Questions?

New to KNIME?

Get "KNIME Beginner's Luck" from KNIME Press



www.knime.com/knimepress