

Next Generation Business Analytics with RapidMiner and RapidAnalytics

Open Source Data Mining Software, Solutions, and Services by Rapid-I and its Hungarian Partner United Consult

Ralf Klinkenberg

Rapid-I GmbH, Dortmund, Germany www.rapid-i.com





Overview

- About the Speaker and the Company Rapid-I
- The Data Mining Software RapidMiner
- Data Mining and Predictive Analytics
- RapidMiner Data Mining Applications
- The Data Mining Solutions RapidMiner, RapidAnalytics, RapidSentilyzer, and RapidNet
- Rapid-I Market Place
- Future Directions
- Data Mining Services Provided by Rapid-I and its Hungarian Partner United Consult





About Ralf Klinkenberg

- Initiator and co-developer of the open source data mining project RapidMiner
- One of the two founders of Rapid-I
- Chief Business Development Officer (CBDO) of Rapid-I responsible for partner programs, sales, training courses, consulting, and projects
- 17 years of expertise in data mining, text mining, web mining, predictive analytics, and sentiment analysis
- Author of numerous data mining publications





About Rapid-I

- Leading open-source data mining provider
- Software development, consulting, and services
- Rapid-I provides innovative data mining and business intelligence solutions increasing the efficiency and profitability of its customers.
- The discovery and leverage of unused business intelligence from existing data enables better informed decisions and allows for process optimizations.
- Winner of the Open Source Business Award 2008



About RapidMiner

- Flexible solution for data mining
- Rapid Prototyping: Rapid development of data mining prototypes as well as of deployable and dependable solutions and applications
- Usable as stand-alone application and as library for the integration into your own products: Stand-alone (GUI), server (command line), Java library
- Large number of users:
 - Market research, marketing, sales, CRM
 - Manufacturing (in particular in the automotive sector)
 - Telecommunication companies
 - Financial services
 - Pharma industries

– **...**

Start of development: 2001

> Monthly downloads: 20,000

More than 600 operators

100% Java (OS-independent)



About the Open Source Business Model

From a customer view point:

- Low entrance barrier: no license fees
- Freedom to choose the amount of service and support needed
- Pay only for customized service, not pay for naked licenses
- No vendor lock-in: freely choose your service provider
- Low total cost of ownership (TCO), fast return on invest (ROI)

From a software and service provider point of view:

- Rapidly gain market share without marketing and sales cost
- Rapidly grow your business based on growing user base

500,000+ downloads

users in 60+ countries

Rapid-I team doubled every year since start

Rapid-I sales volume and profit doubled every year





Can you look into the future?





- Can you look into the future?
- Example: A glas falls to the ground and will break.
- Even before the glass touches the ground, we know:
 In a moment the glass will break into pieces.





- Can you look into the future?
- Example: A glas falls to the ground and will break.
- Even before the glass touches the ground, we know:
 In a moment the glass will break into pieces.
- How do we know that?





- Can you look into the future?
- Example: A glas falls to the ground and will break.
- Even before the glass touches the ground, we know:
 In a moment the glass will break into pieces.
- How do we know that?
- From observations and experience from the past we generate rules and use these for prediction.





- Data Mining (DM) discovers prediction rules in the "experience" of companies: in their data
 - Predictions for new situations
 - Explicit descriptions of processes allow better informed decisions and planing
- Difference to Business Intelligence (BI):

"The glas is broken into 20 parts." (BI report)

versus

"The glas will fall. Pay attention!" (DM report)







Optimization by Predictions

- Insurance: How many accidents will this customer cause? What insurance rate to suggest?
- Manufacturing: How to predict machine failures –
 before they occur? (preventive maintenace)
- Marketing: How will the new product be adopted by the market? How to advertise it best?
- CRM: What will be the reasons for buying the product? Or why will potential customers dislike the product?

•





Data Mining Applications

- Credit risk prediction, credit default prediction:
 Credit scoring, rating, risk management
- Rating customers or companies
- Analysis and prediction of financial time series:

Sales, stock market prices, exchange rates, commodity prices, indices, volatilities, ...

by uni- and multi-variate analysis and also using unstructered data like text documents

Customers: aiinvesting, NeuralMarketTrends, ...

Automated sentiment analysis from texts:

Sentiments about companies, products, brands, stocks, persons, parties extracted from news, discussion groups, and web blogs

Customers: Online Market Research: Schober Information Group, AFO Marketing, GfK, maanto, moxie interactive, T-Mobile, PayPal, ... and Financial Sector





Data Mining Applications

Automated news filtering:

Finding or monitoring all relevant articles on a given topic and topic tracking, e.g. for brand monitoring

Automated document classification:

Automated assignment of documents into pre-defined categories automatically learned from given examples

Automated e-mail routing:

Routing incoming e-mails automatically based on their content to the department or person in charge

Customer: mobilkom austria, Austria's leading mobile phone service provider





Data Mining Applications

- Market research and competitve intelligence:
 Online market research, competitive intelligence
 Customers: GfK, Schober Information Group, AFO Marketing, ...
- Automated product recommendation:
 Individually customized/personalized product recommendations, fitting additional products (cross-selling), and higher-valued products (up-selling)
- Marketing, Sales, Customer Relationship Management (CRM):

Optimizing and controlling direct marketing activities, increasing response rates, estimating response probabilities for each customer, maximizing the return on investment (To whom advertize what when how?)





Example: Web Discussion Survey, Blog Mining, Automated Sentiment Analysis

- Market Research: How will the product be adopted by the market and why?
- CRM: What are the reasons why customers buy the product? Or why do potential customers dislike it?
- Online Market Research: Which sentiments are expressed in internet discussion groups and web blogs regarding a company or product?
- Goal: Automatically extract the expressed sentiments.





Selected Customers of Rapid-I and Users of RapidMiner in Finance

Banking and Investment:

- BNP Paribas (France)
- Bank of America (USA)
- comdirect bank (Germany)
- Landesbank Berlin (LBB) (Germany)
- Merrill Lynch (USA)
- PayPal (USA)
- aiinvesting.com (USA)
- NeuralMarketTrends.com (USA)

Insurance Companies:

- Ineas (France)
- PentaSecurity (Chile)
- Allianz (Germany)
- Wüstenrot (Austria)





Selected Customers of Rapid-I and Users of RapidMiner in Industry

Manufacturing:

- ThyssenKrupp Nirosta (Germany)
- ThyssenKrupp Steel Europe (Germany / Europe)
- Mannesmann Salzgitter (Germany)
- Schott (Germany)

Retail:

- DiTech (Austria)
- Libri (Germany / Europe)
- Rossmann (Germany)
- Tchibo (Germany)

Utilities:

- E.ON Ruhrgas (Germany)
- RWE (Germany)





Selected Customers of Rapid-I and Users of RapidMiner in Industry

Aviation:

- Lufthansa (Germany)
- EADS (France / Europe)

Pharma and Medical:

- Sanofi-Aventis (Germany / France)
- Siemens (Germany / USA)

Telecom:

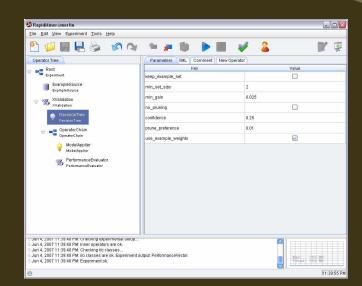
- Telekom Deutschland (Germany)
- T-Mobile International (Germany / Europe)
- mobilkom austria (Austria)

Market Research:

- GfK (Germany / Europe)
- PepsiCo (USA)
- Schober Information Group (Germany)
- moxie interactive (USA)

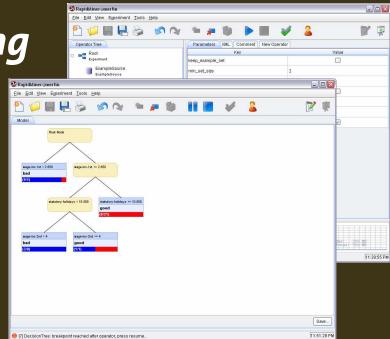


More than 600 data mining operators



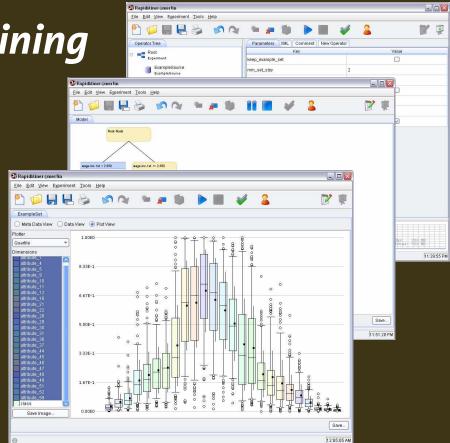


- More than 600 data mining operators
- Usable as stand-alone application or library



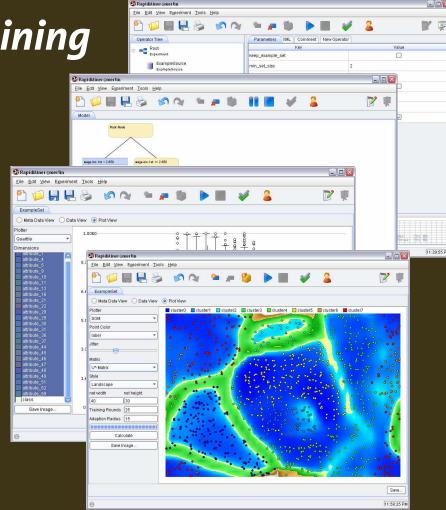


- More than 600 data mining operators
- Usable as stand-alone application or library
- Standardized process descriptions and outputs



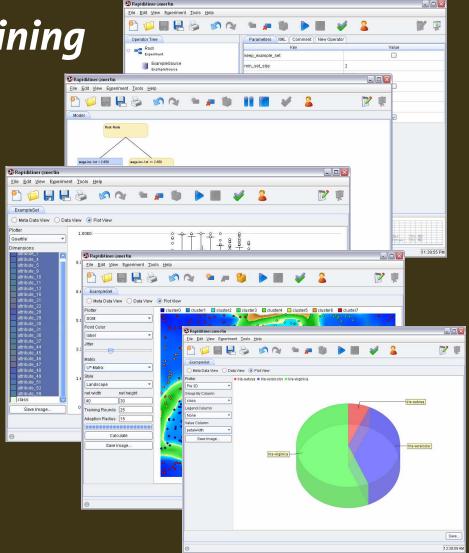


- More than 600 data mining operators
- Usable as stand-alone application or library
- Standardized process descriptions and outputs
- Efficient multi-layered data view

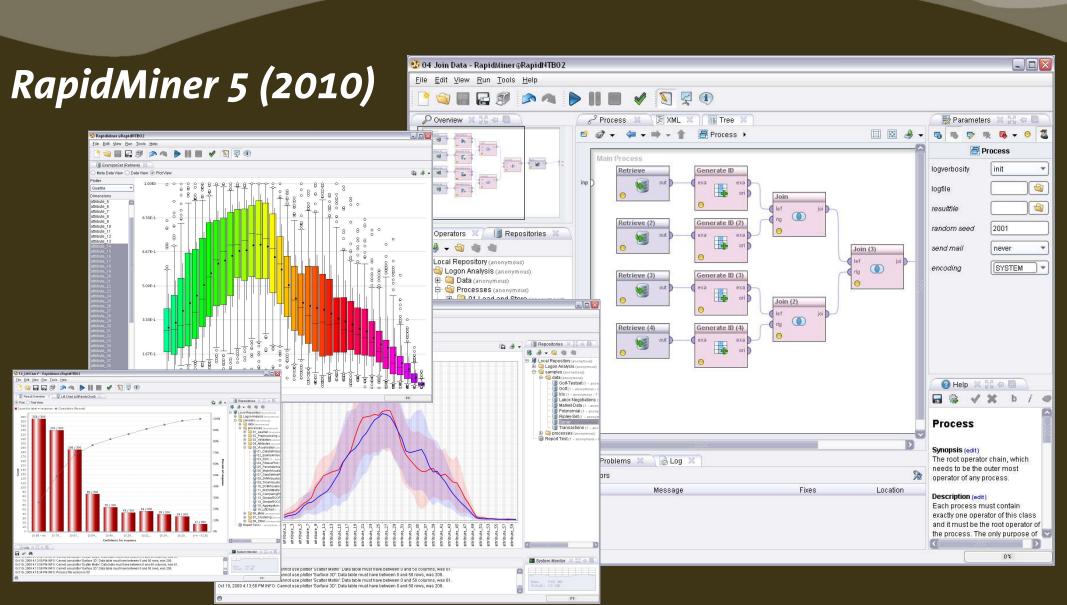




- More than 600 data mining operators
- Usable as stand-alone application or library
- Standardized process descriptions and outputs
- Efficient multi-layered data view
- Large-scale DM via scripting language, interactive via explorative data analysis





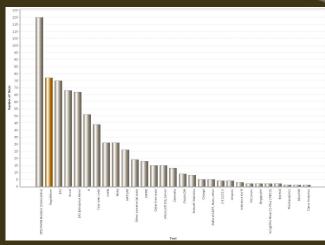




Adoption of RapidMiner

- KDnuggets Polls 2007, 2008, and 2009:

 RapidMiner most widely used open source data mining tool, second overall, and 21% of all data mining experts work with RapidMiner
- SourceForge.net:
 More than 500,000 Downloads,
 users in more than 60 countries
- Data Mining Cup 2007, Study by Christian Schieder "Open Source Data Mining Tools", and BARC 2009: RapidMiner leads in the leaders' quadrant







Study: Open Source Data Mining

"Open Source Data Mining" by Christian Schieder (Technical University of Chemnitz, Germany) presented at the international Data Mining Cup (DMC) 2007 and BARC presented at CeBIT 2009:

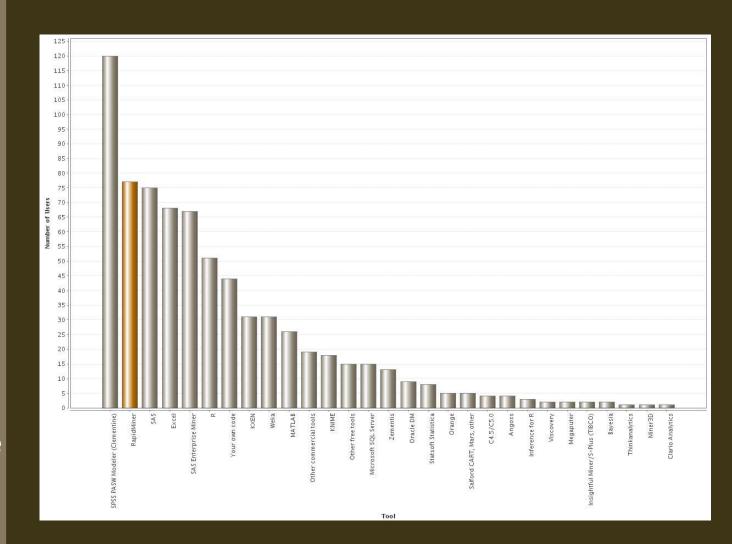
RapidMiner is the world-wide leading Open Source Data Mining Software





KDnuggets Poll 2009

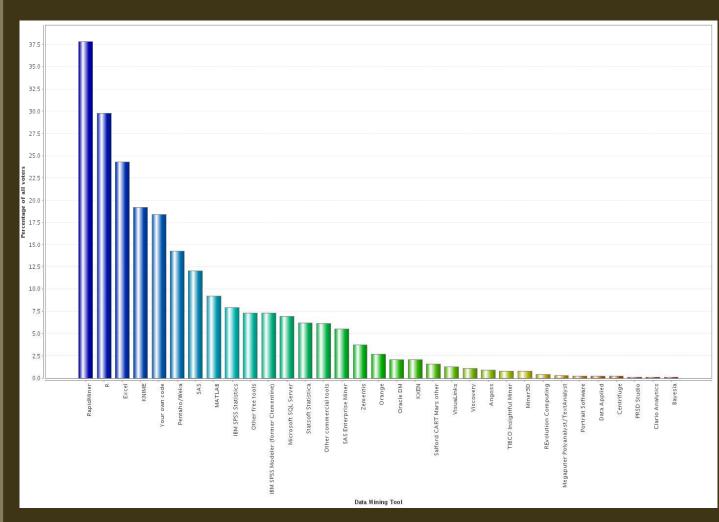
- Poll of the leading data mining portal KDnuggets among 364 data mining experts
- 21% work with RapidMiner
- RapidMiner is the second most widely used data mining tool overall and the most widely used open source data mining software and library



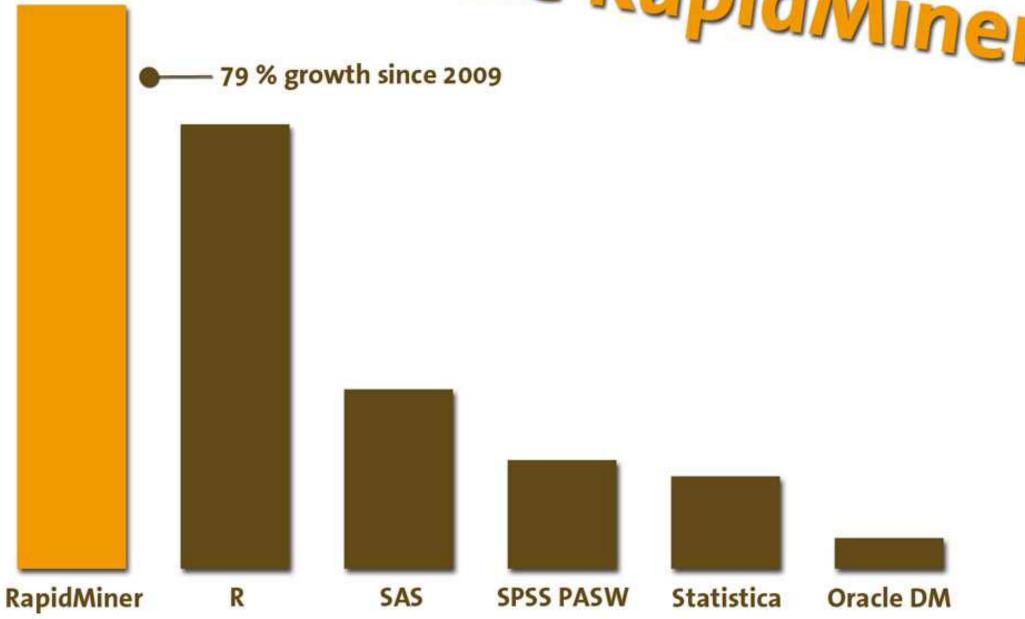


KDnuggets Poll 2010

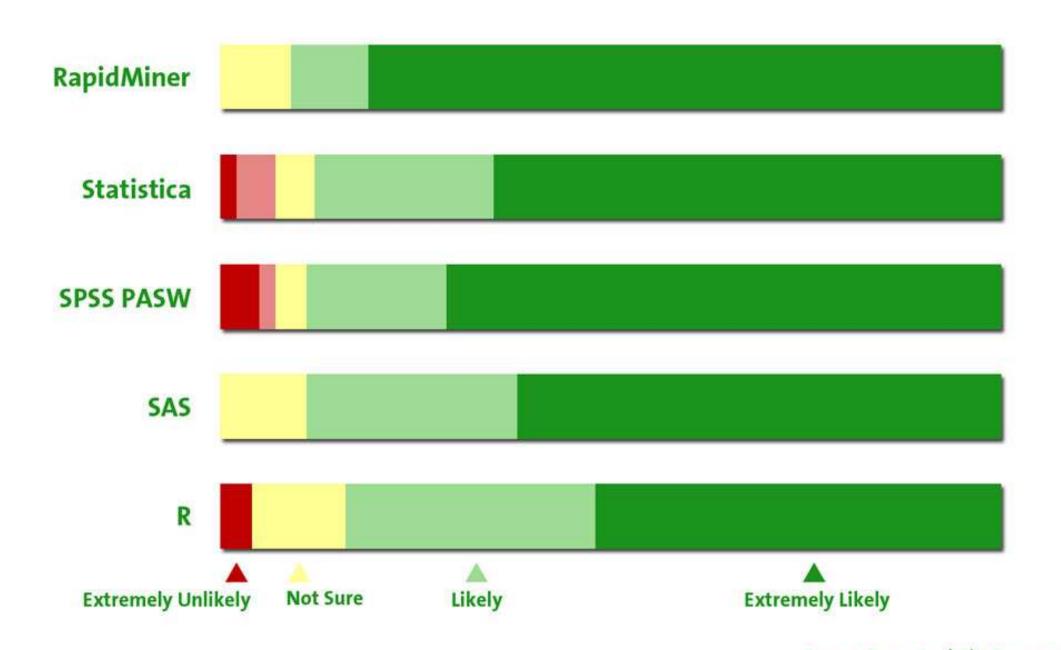
- Poll of the leading data mining portal KDnuggets among 912 data mining experts
- 37.8% work with RapidMiner
- RapidMiner is the most widely used data mining tool world-wide
- Open source data mining solutions gain significantly



38% use RapidMiner



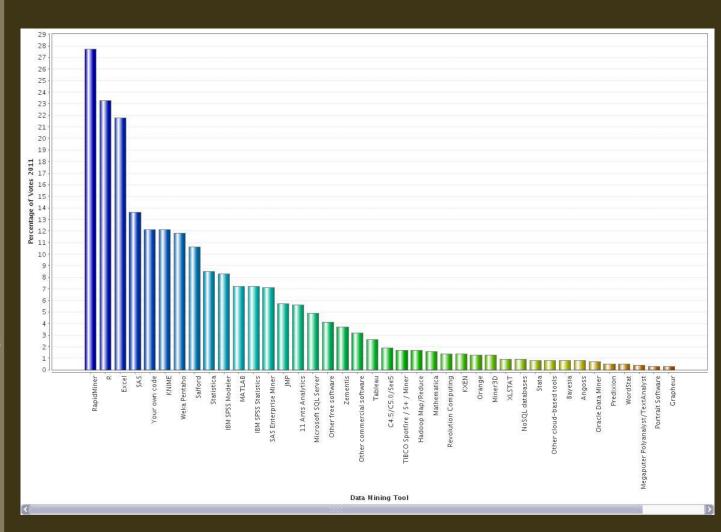
91% are likely to continue using RapidMiner for the next 3 years





KDnuggets Poll 2011

- Poll of the leading data mining portal KDnuggets among more than 1100 data mining experts
- 27.7% work with RapidMiner
- RapidMiner is the most widely used data mining tool world-wide
- Open source data mining solutions gain significantly





RapidMiner Family



Rapid-Sentilyzer







RapidDoc

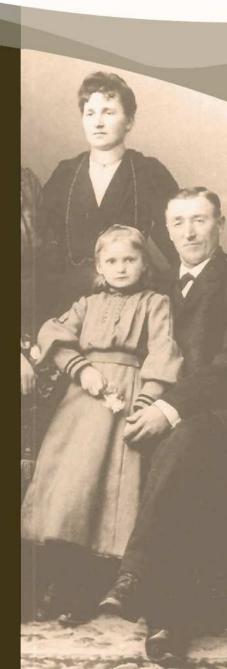
RapidMiner (GUI)

RapidNet

RapidLab

Rapid-Analytics

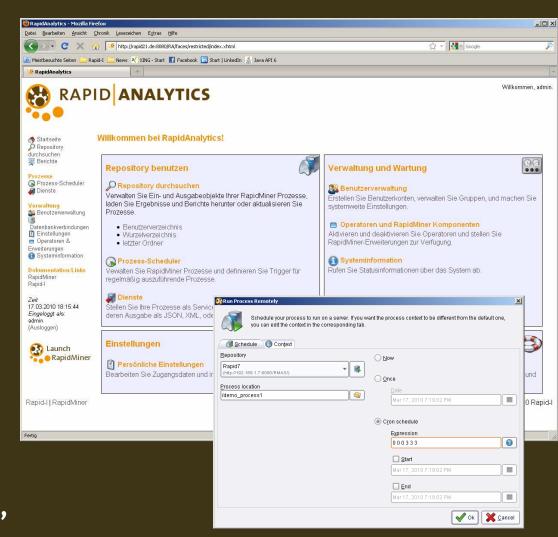
RapidMiner (Engine)





RapidAnalytics

- Enterprise analytics server with the flexibility, functional breadth, and predictive power of RapidMiner
- Remote process execution and scheduling of analysis processes
- User and rights management
- Option to provide RapidMiner processes as web services
- Collaborative work on shared data, meta-data, analysis processes, results, and reports via shared repositories





RapidAnalytics

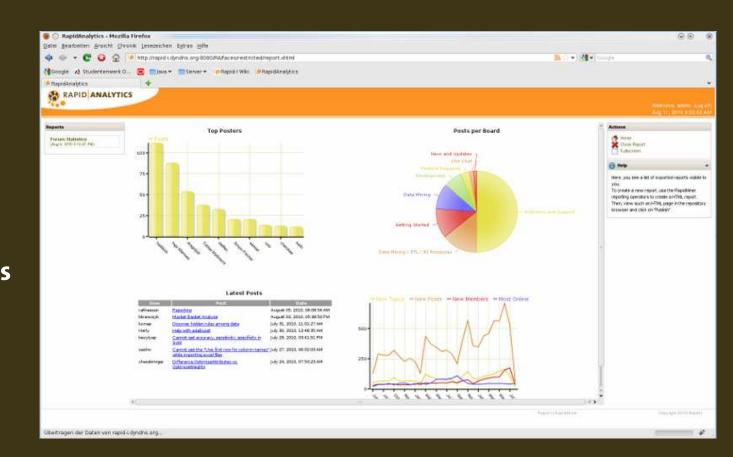
- Based on industry standards for application servers
- Web-based access to processes, results, and reports
- Access via RapidMiner client software
- Automated calendar-driven process scheduling
- Seamless integration in existing environments via export of results as web services





RapidAnalytics

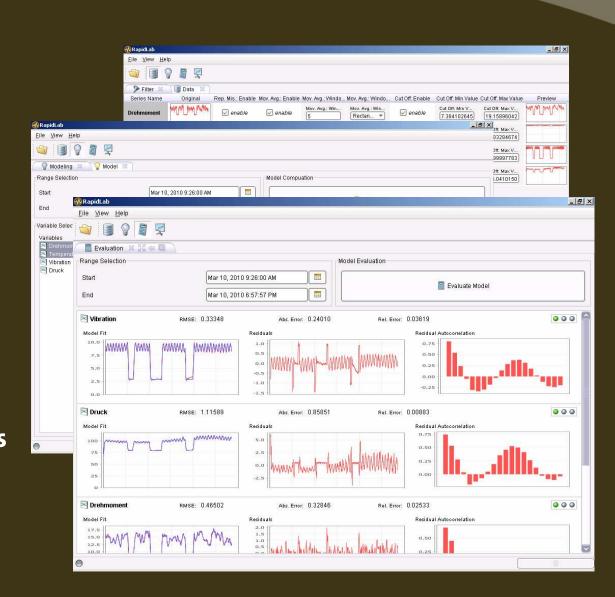
- Static reports in HTML, PDF, Excel, RTF format
- Dynamic reports via the web
- Dynamic report elements (charts, tables, etc.) via web service embedable into other systems





RapidLab

- Visualisation of process data
- Easy modelling of causal relations
- Model-based simulationen and scenario analysis support process understanding and optimization
- Easy-to-use user interface supports intuitive data cleaning, modelling, and evaluation





RapidSentilyzer BuzzBoard

- Real-time online market research for brand monitoring, market insight, customer insight, and competitive intelligence
- Automated web crawling of the latest news, forum posts, and blog entries about your company, your products, and/or your competitors
- Automated categorization and sentiment analysis of all texts with RapidMiner as text mining engine
- Easy-to-use web portal solution without configuration or maintenance effort for you





RapidSentilyzer BuzzBoard

- Get answers to questions like
 - What do my customers think?
 - How successful are my marketing and PR campaigns?
 - How competitive is my company?
 - How will the needs and wishes of my customers develop?
- Sentiment analysis
 - Positive / negative / neutral
 - Development over time / trends
 - Sources / opinion leaders
 - Reasons / Topics / Key Issues/ Key Words





RapidSentilyzer BuzzBoard

- Buzz Statistics: number of relevant documents found and their sources
- Sentiment Statistics: aggregated market sentiment
- Sentiment History: sentiment trends over time
- Sentiment Insight:
 detailed information about the
 reasons for the observed opinions
 and sources of particular sentiments
 and opinions (opinion makers)

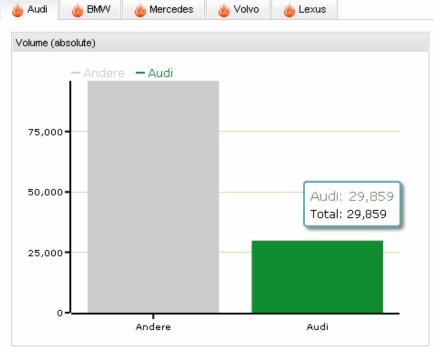


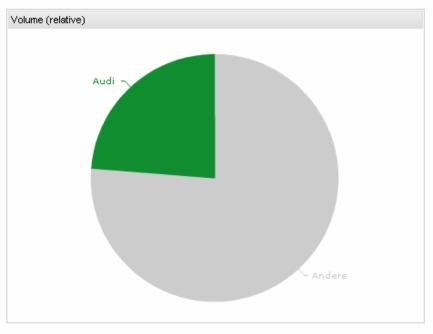


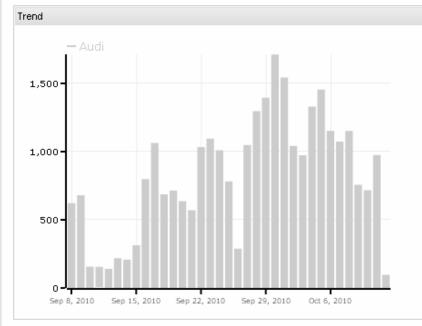
RapidSentilyzer BuzzBoard -- Real-Time Market Insight

- Examples from two domains on the following slides:
 - Automobile: comparing the sentiment about five international car brands.
 - Consumer electronics: comparing the sentiment about three brands.
- RapidSentilyzer BuzzBoard elements:
 - Selection of brand and content language (English/German/French/.../all).
 - Volume: Number and relative share of messages about the target brand.
 - Sentiment: Share of positive, negative, and neutral messages overall and over time (trend) as well as most recent most positive and most negative mesages.
 - Sources: Sources of the buzz and the sentiments as well as the most prominent positive and negative topics (word cloud).



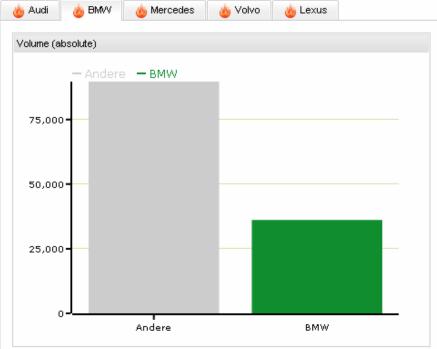


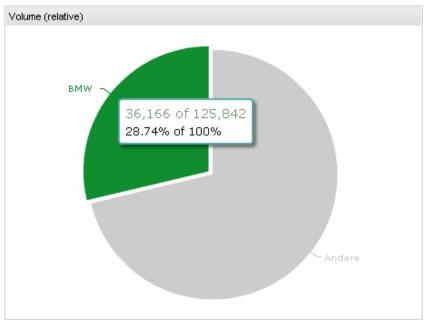


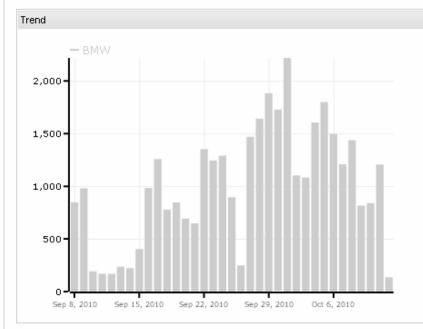






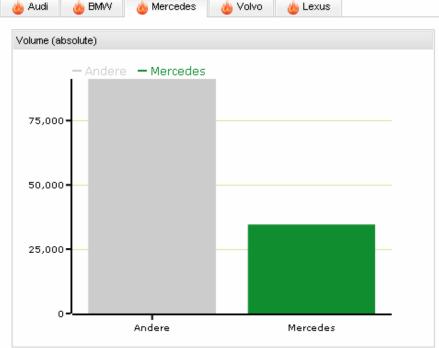


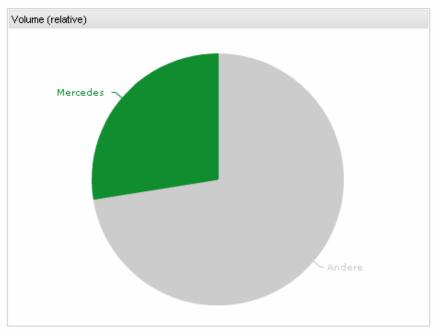


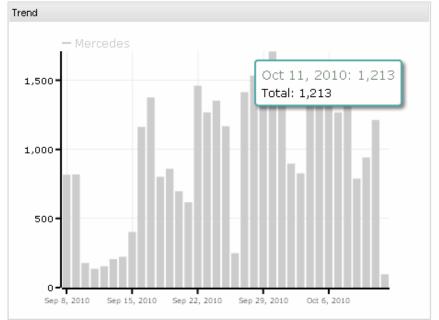






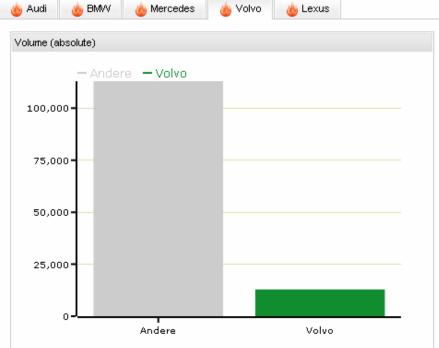


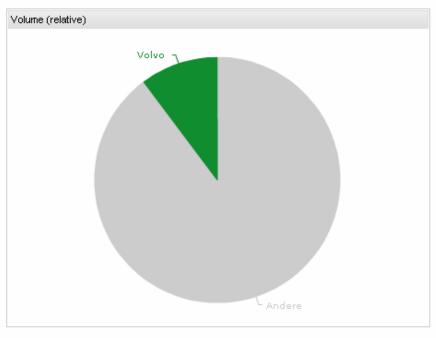


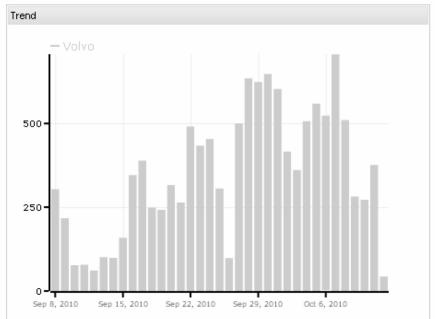






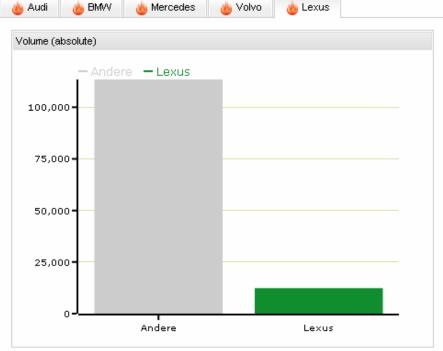


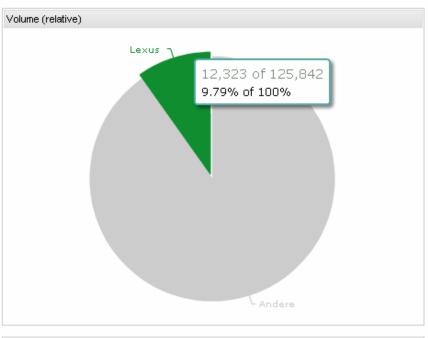


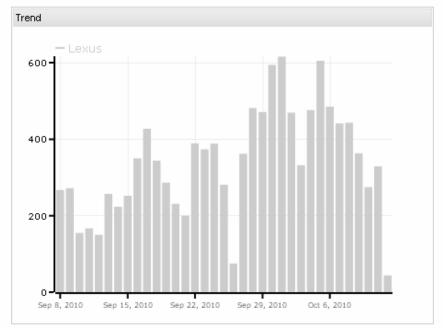


Date	Title
10/12/10 3:28	A Peugeot Looking to Dethrone Audi Cars Is Your Niche
10/12/10 3:27	A Nägel unter Autoreifen
10/12/10 3:23	A SailRaceWin: VOR: Lady Pippa Bake is Ambassador to the Volvo Oc
10/12/10 3:18	A Steps To Saving Money On Volvo Repair Pima County « Quincy Auto
10/12/10 3:15	A Inggitrust » camion frigo
10/12/10 3:06	A Fin de semaine hésitante à Paris
10/12/10 3:00	A blanche » Blog Archive » So consume a salad, own a piece of fruit
10/12/10 2:47	A Pickle Lintas beefs up its creative with two senior appointments
10/12/10 2:45	A Kuwaiti company inaugurates first green fuel station in Sweden
10/12/10 2:45	A Weet-Bix takes to snow and space in NZ and Australia Latest Auto
10/12/10 2:40	A 1972 Volvo P1800 ES Estate Classic Cars Pictures
10/12/10 2:33	A Nägel unter Autoreifen
10/12/10 2:26	A Nägel unter Autoreifen



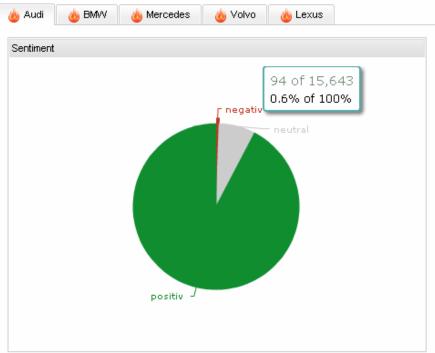


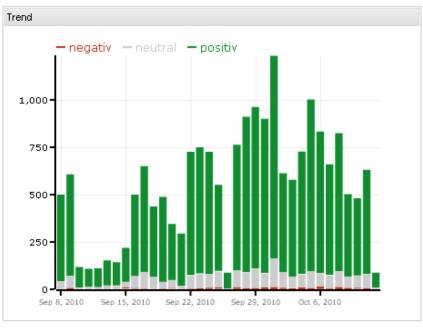




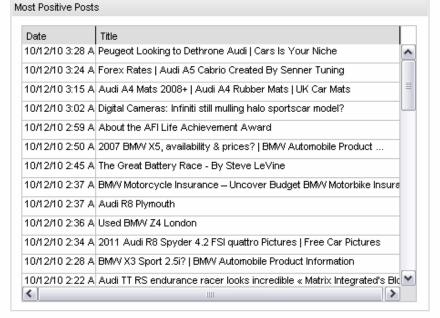






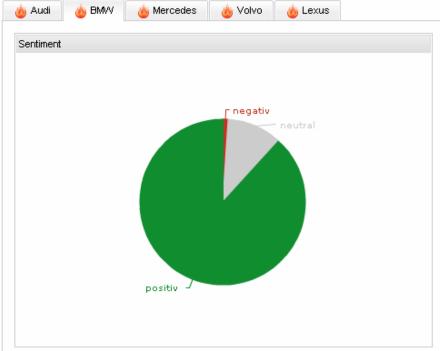


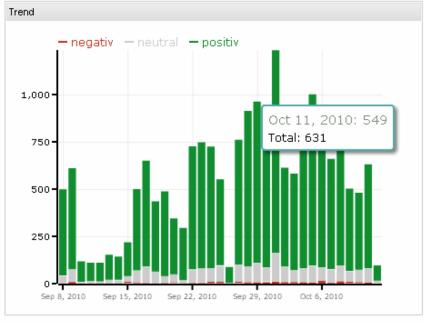
Date	Title	ı
10/12/10 12:59 /	Consider the Benefits of Going Online To Get Your Audi Wheels	1
10/11/10 6:28 Pt	girls, shoul Howered my standards??w/ girls? Audi Auto	
10/10/10 11:25 F	Driver arrested after motorcyclist killed in crash	1
10/9/10 11:12 Pf	Hot Video 2010 Audi RS5 Latest Spy Photos HD Alloy Wheels With 1	L
10/9/10 8:29 PM	Davies says he wasn't driver of car going 125 mph	
10/9/10 7:04 PM	Charlie Davies says he wasn't driving speeding car, took blame to sp	
10/8/10 7:13 PM	How Much Does Engine Swap Job Cost For Audi A4? b2b b2c infor	
10/8/10 1:11 PM	Audi S5 SP - sc16 « DPE Wheels, USA	
10/8/10 10:50 A	Man charged over road smash death	
10/7/10 9:56 PM	Taunton house hit by car, but who was driving?	
10/7/10 8:22 PM	Taunton house hit by car, but who was driving?	
10/7/10 11:44 A	Kill plot dad ducks bullet	
10/7/10 7:02 AM	Cops: True target found	1



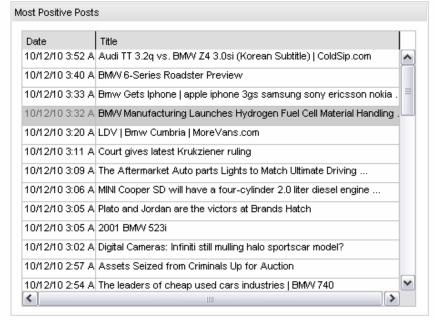
About RapidSentilizer I Rapid

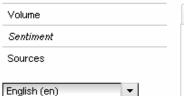


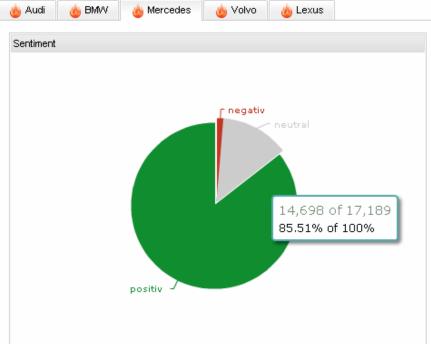


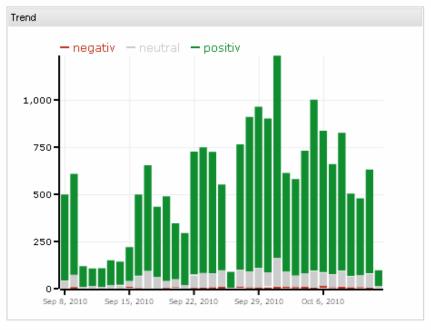


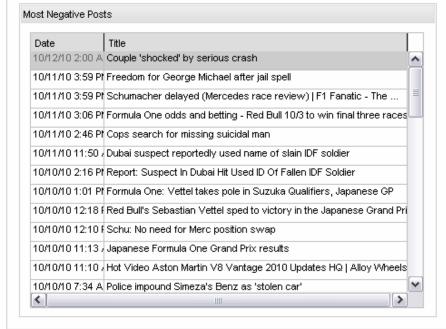
Date	Title	L
10/9/10 3:17 PM	BMW seized in drug arrest after traffic stop	1
10/9/10 1:30 PM	Tributes paid to gran killed in car crash in South Africa	
10/9/10 12:10 Pt	Tributes paid to gran killed in car crash in South Africa	
10/9/10 11:10 A	Hot Video 2004 BM/V 745LI VEHICLEMAX.NET White #29942A Miami	
10/9/10 10:21 A	Cocaine gang trio jailed for 11 years after police undercover op	
10/9/10 10:14 A	Cabbie leaves boy to hobble home after collision	=
10/9/10 12:19 A	Police Service Aide Busted for DUI By Buddies	۲
10/8/10 11:44 Pt	FORT LAUDERDALE —	
10/8/10 11:38 Pt	COPS: Drunk woman hits man, drives home with scooter stuck under	
10/8/10 4:27 PM	BMW drink-driver jailed after trying to bribe police	
10/8/10 2:29 PM	Jailed: Villain who pretended to be 'hardman' Paul Massey in £1m bla	
10/8/10 12:56 Pf	BUCKHURST HILL: Father who had £70k worth of tools stolen is sho	

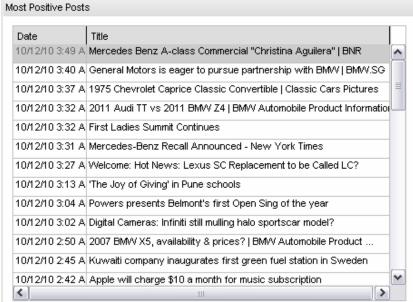




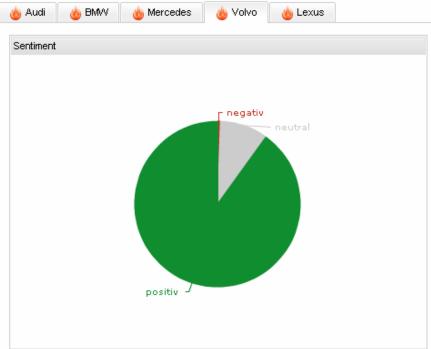


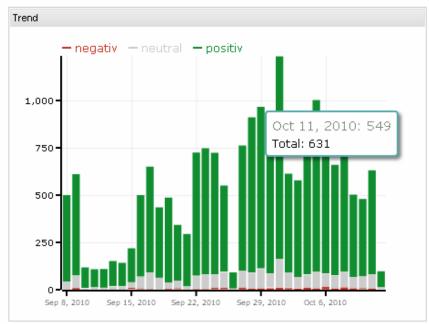


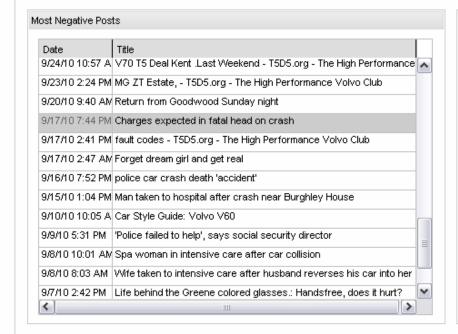










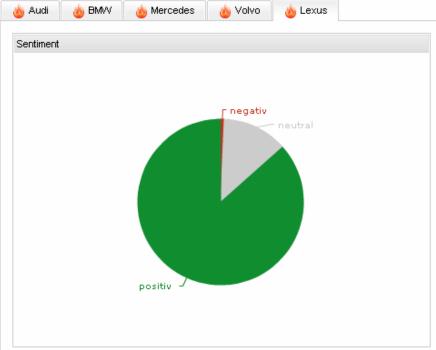


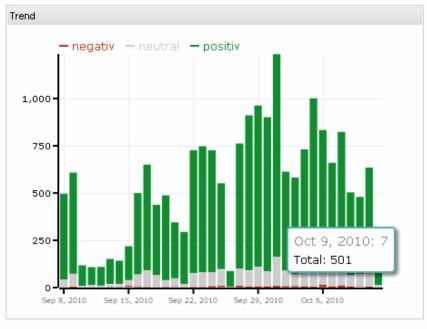


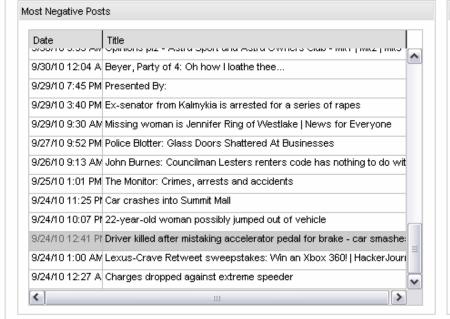
About RapidSentilyzer I Rapid-

Welcome, Ralf Klinkenberg | Switch project | Logout



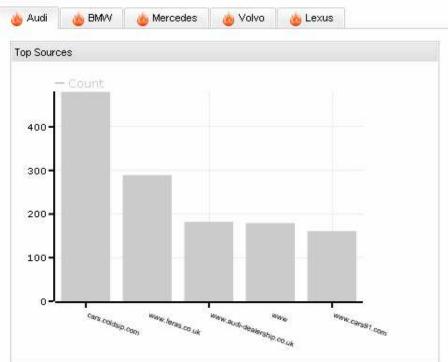












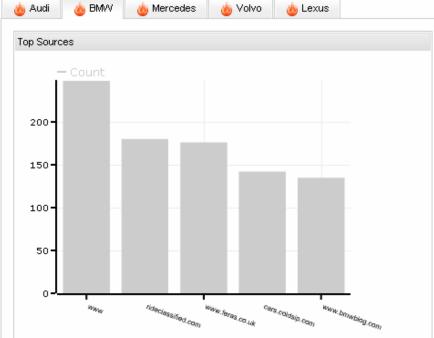
ources				
Source	negative	neutral	positive	
cars.coldsip.com	0	4	476	^
www.feras.co.uk	0	2	287	
www.audi-dealership.co.uk	0	11	171	
www	1	13	165	1
www.cars91.com	0	0	161	i
bmwautomobile.ditttos.com	0	2	117	
www.carswikipedia.com	0	3	113	
Autoblog (blog)	0	3	102	1
www.roadstercollectibles.com	0	6	90	
www.autoblog.com	0	1	89	1
carnews.gossipblender.com	0	0	90	1
www.carzbuzz.com	0	1	87	1
babewash.net	0	1	82	~
<	Ш		>	

Welcome, Ralf Klinkenberg | Switch project | Log







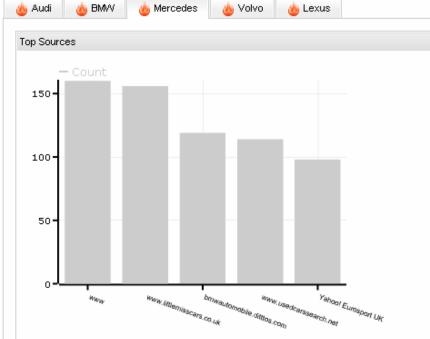




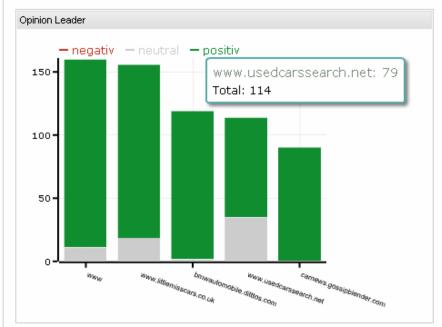
Sources negative Source neutral positive 28 220 www 175 rideclassified.com 175 www.feras.co.uk 0 1 cars.coldsip.com 0 2 140 www.bmwblog.com 135 bmwautomobile.ditttos.com 117 99 www.tsikot.com www.roadstercollectibles.com 6 90 usedcars4auction.info 22 74 BMW BLOG (blog) 95 1 carnews.gossipblender.com 90 85 Motorsport.com AutoSpies.com 78 >

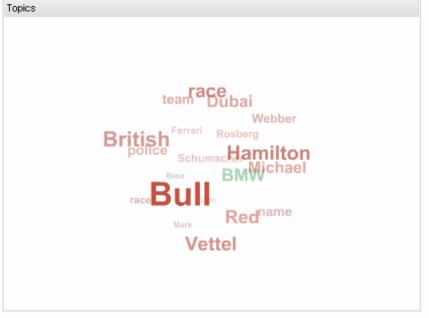


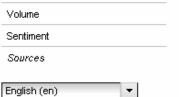


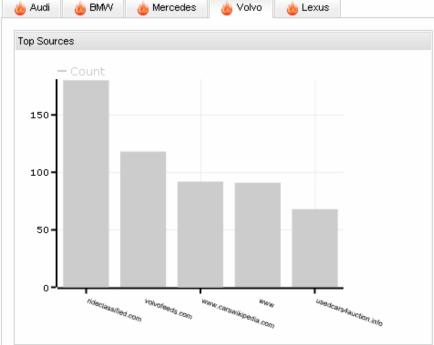


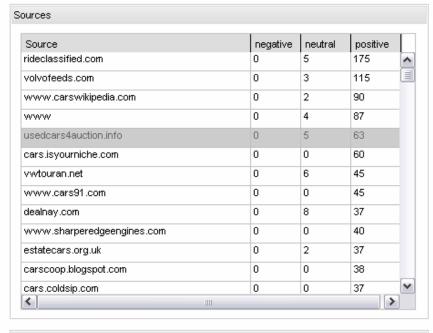
Source	negative	neutral	positive
www	0	11	149
www.littlemisscars.co.uk	0	19	137
bmwautomobile.ditttos.com	0	2	117
www.usedcarssearch.net	0	35	79
Yahoo! Eurosport UK	3	38	57
bettor.com (blog)	2	24	67
carnews.gossipblender.com	0	0	90
Bloomberg	1	8	74
www.cargeni.com	0	0	79
www.mercedes-benz-blog.blogspot.com	0	4	74
www.atlanticoptimize.com	0	30	43
babewash.net	0	0	70
sports-car.limewebs.com	0	4	65

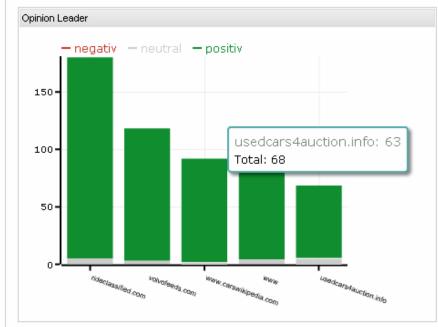


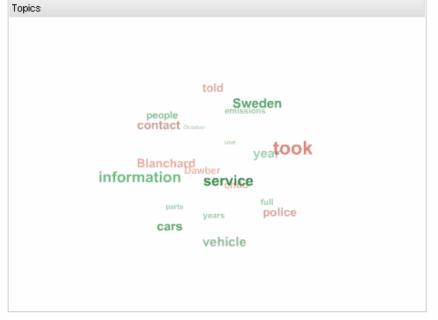




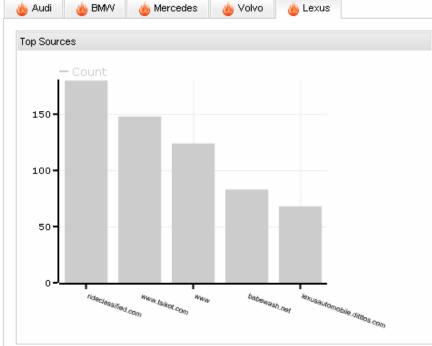




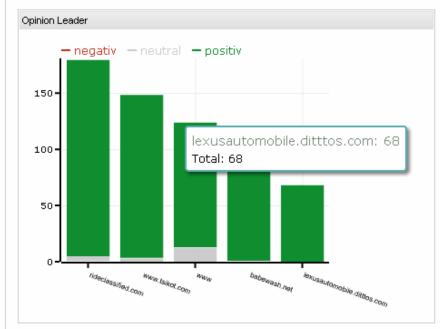








Source	negative	neutral	positive	
rideclassified.com	0	5	175	^
www.tsikot.com	0	3	145	
www	0	13	111	1
babewash.net	0	1	82	1
lexusautomobile.ditttos.com	0	0	68	П
cars.coldsip.com	0	2	63	1
dailydriverblog.com	0	0	64	
cargalaxy.info	0	0	62	
cars.isyourniche.com	0	0	60	1
dealnay.com	0	5	54	1
rkchevytv.com	0	1	55	
www.usedcarssearch.net	0	46	9	
www.feras.co.uk	0	1	54	~





235

224

229

>

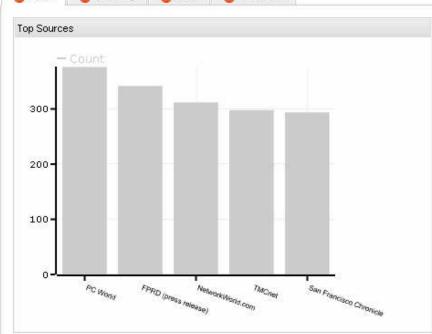
10

Development/Consumer Electronics

🔥 Apple

M Samsung





🔥 Panasonic

& Sony



Source negative neutral positive PC World FPRD (press release) 0 0 342 Network/Vorld.com 20 289 **TMCnet** 13 285 San Francisco Chronicle 0 23 271 Electronista 285 ZDNet (blog) 0 246 OfficialWire (press release) 246 Examiner.com 221 www. 2 23 210

0

Sources

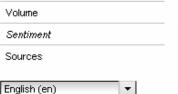
Market/Vatch (press release)

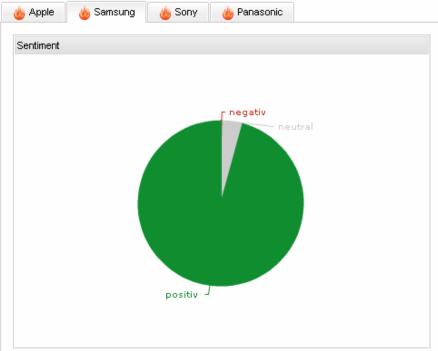
Information/Week

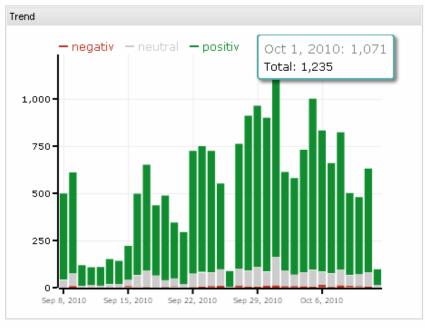
CNET

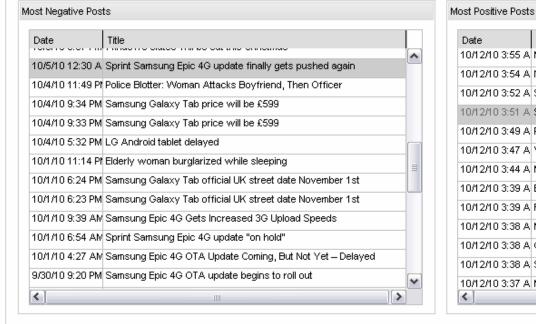


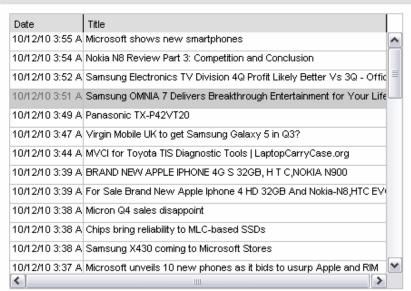
Development/Consumer Electronics



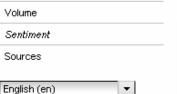


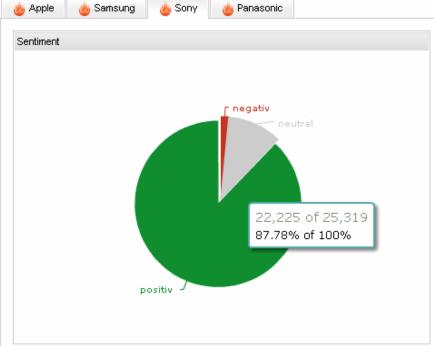


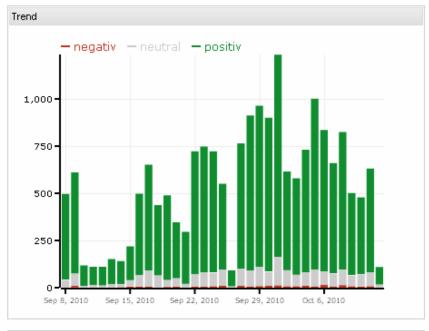


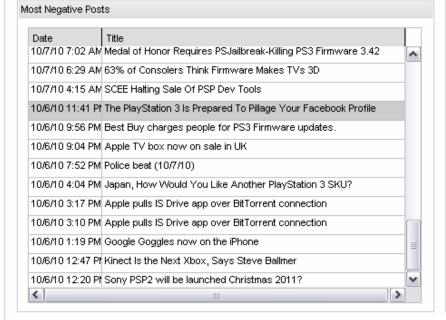


Development/Consumer Electronics











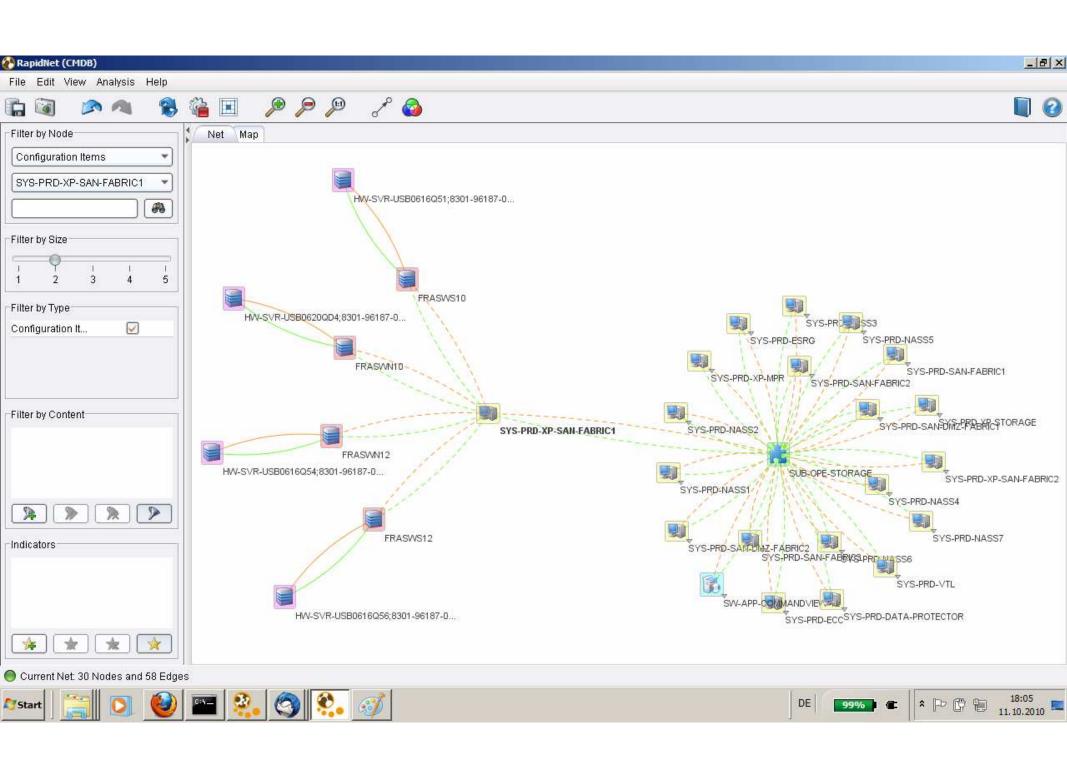


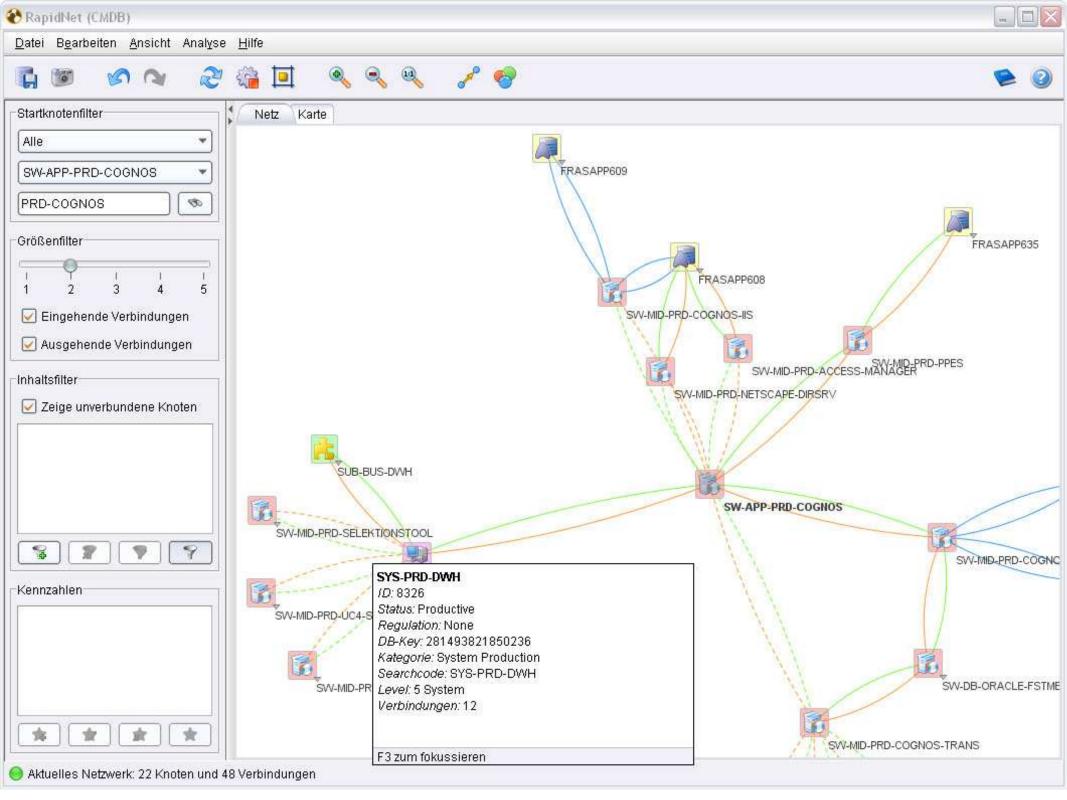
RapidNet -- Visualizing Relations, Networks, Transactions

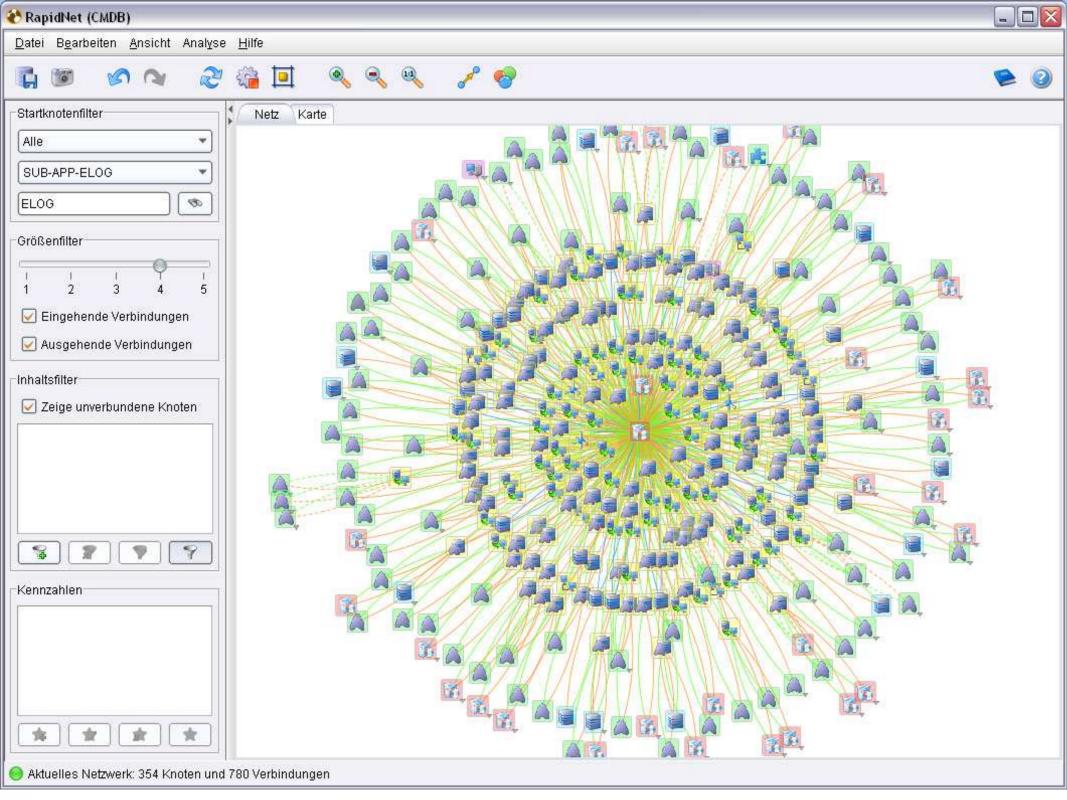
- Examples from several network analysis domains on the following slides:
 - Hierarchical networks of computers hardware and software in a large international corporation with different kinds of dependencies and other relations.
 - Distribution of certain products or money transactions between countries.
 - Different kinds of relations between different kinds of persons (e.g. sales, customers, ...).
 - Relations of persons and companies, underlied with geographical maps.

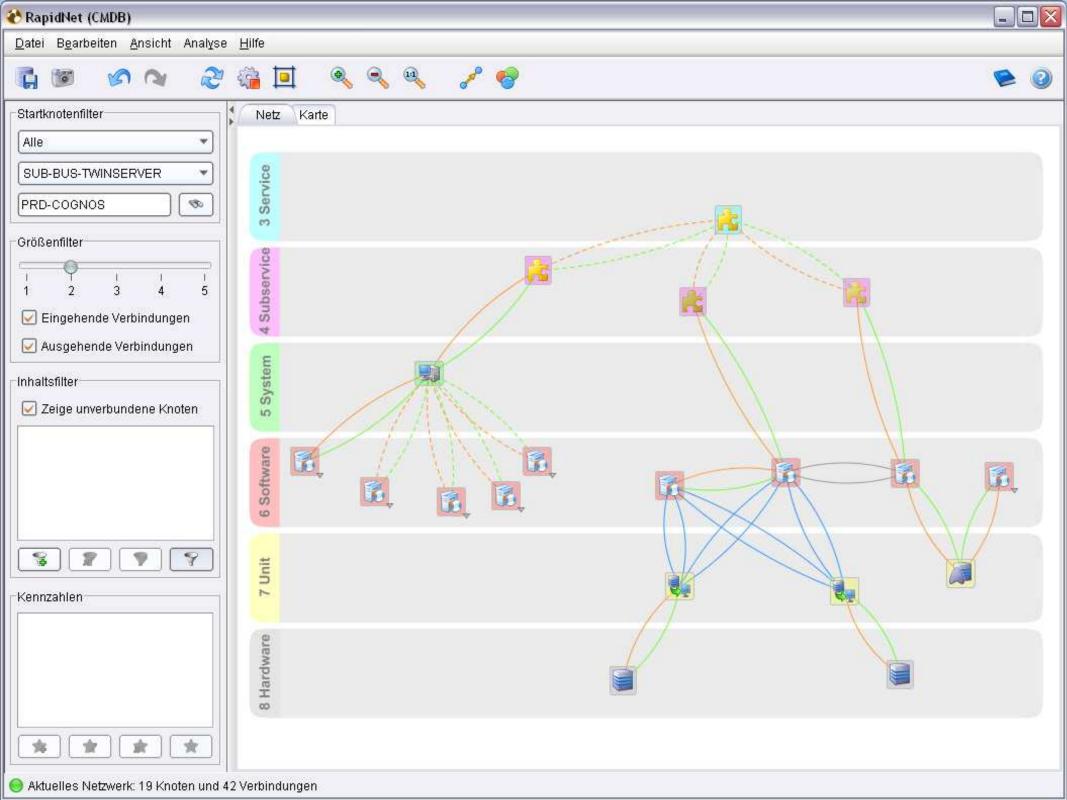
RapidNet elements:

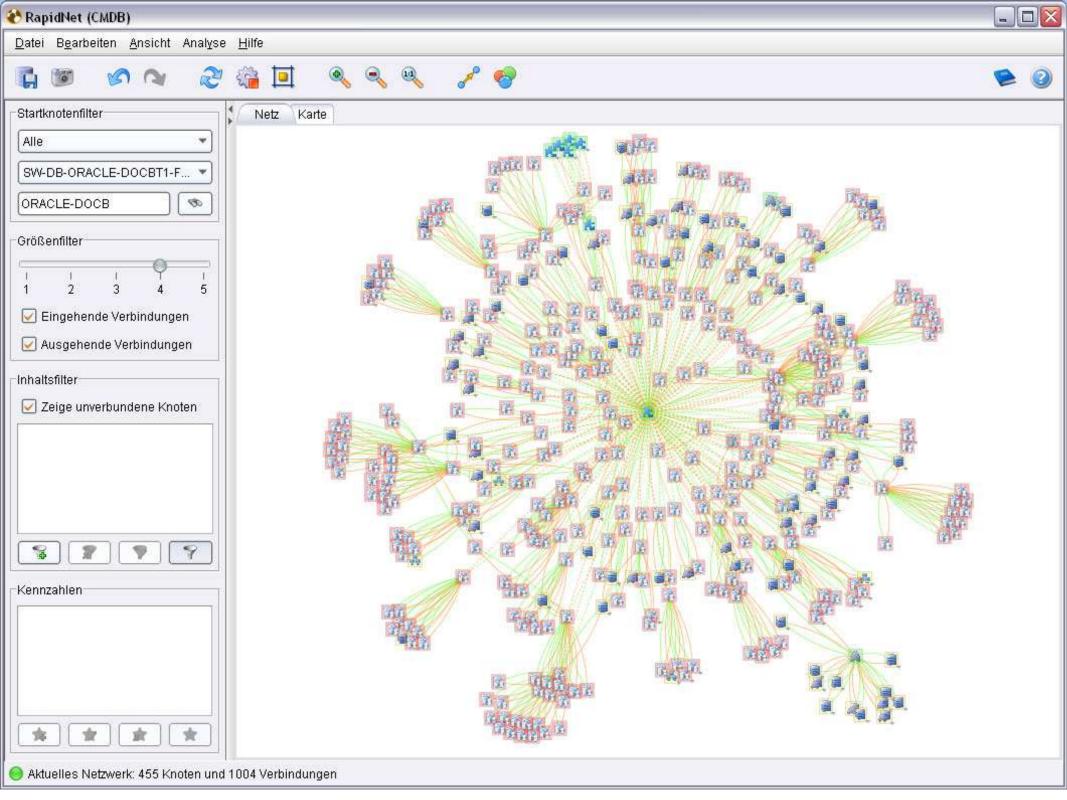
- Selection of start node type and start node for interactive network exploration.
- Number of levels for automated network expansion plus further interactive expansion.
- Definition of filters (e.g. only display nodes with sales > 1 Mio. USD and connecting edges with transactions > 20,000 USD).
- Definition and visualisation of key performance indices (KPIs) or other indicators as bar charts, trend arrows, or text next to the correspodnig nodes.

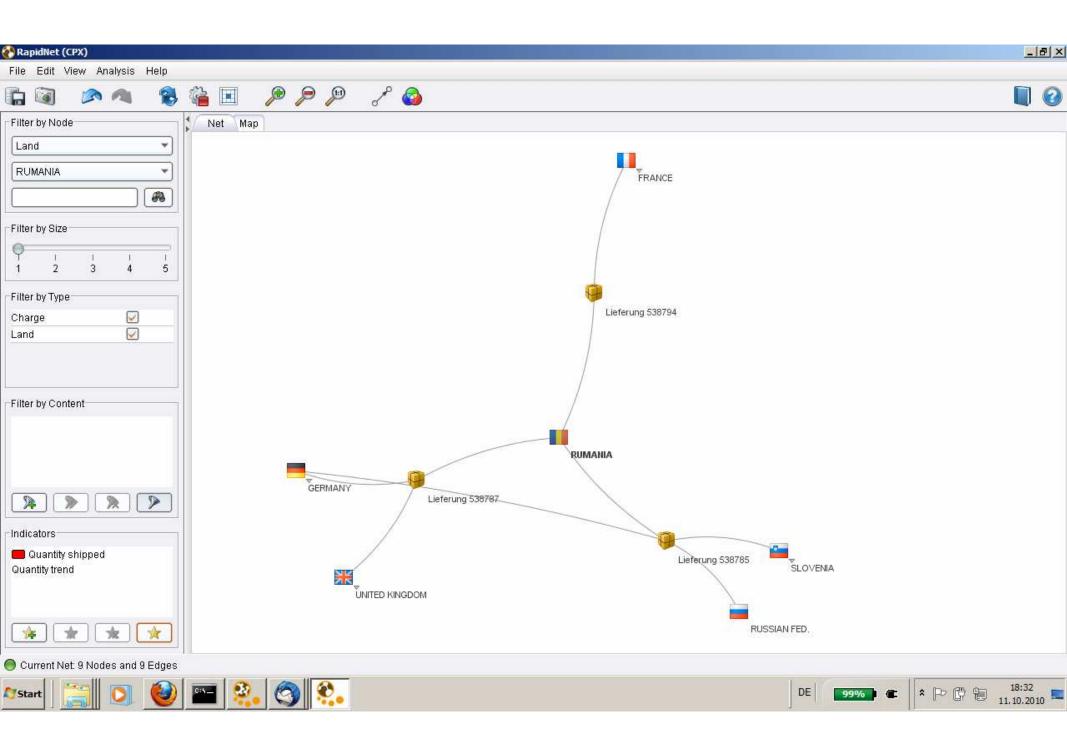


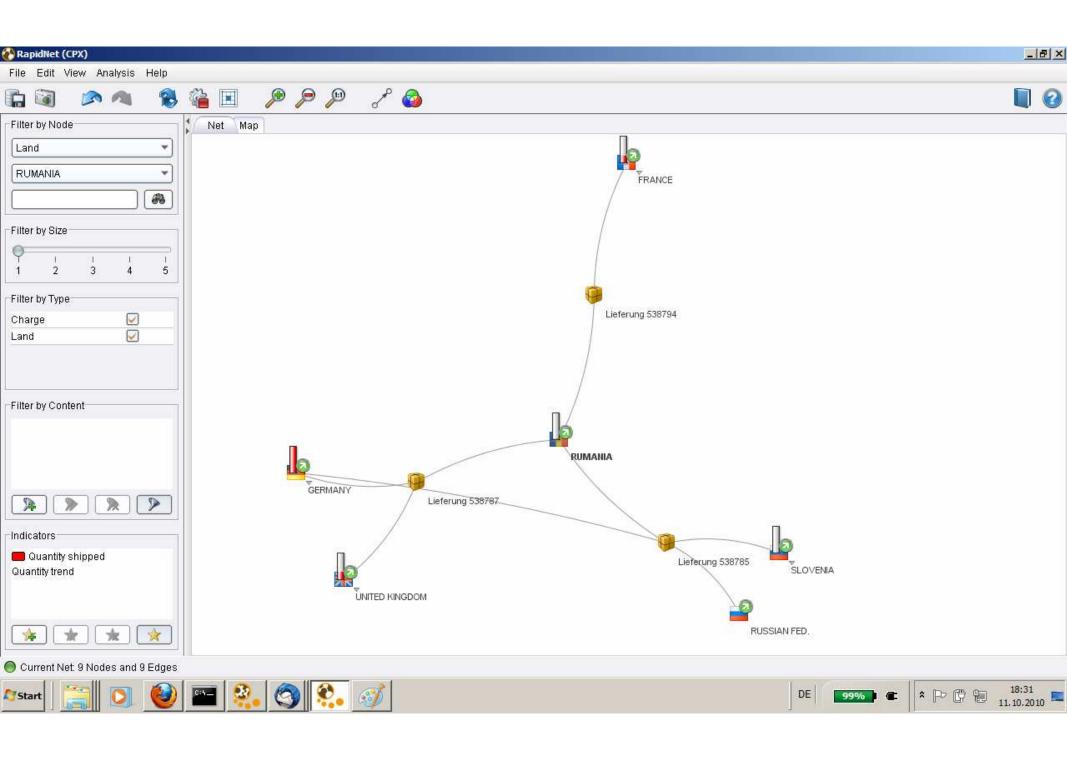


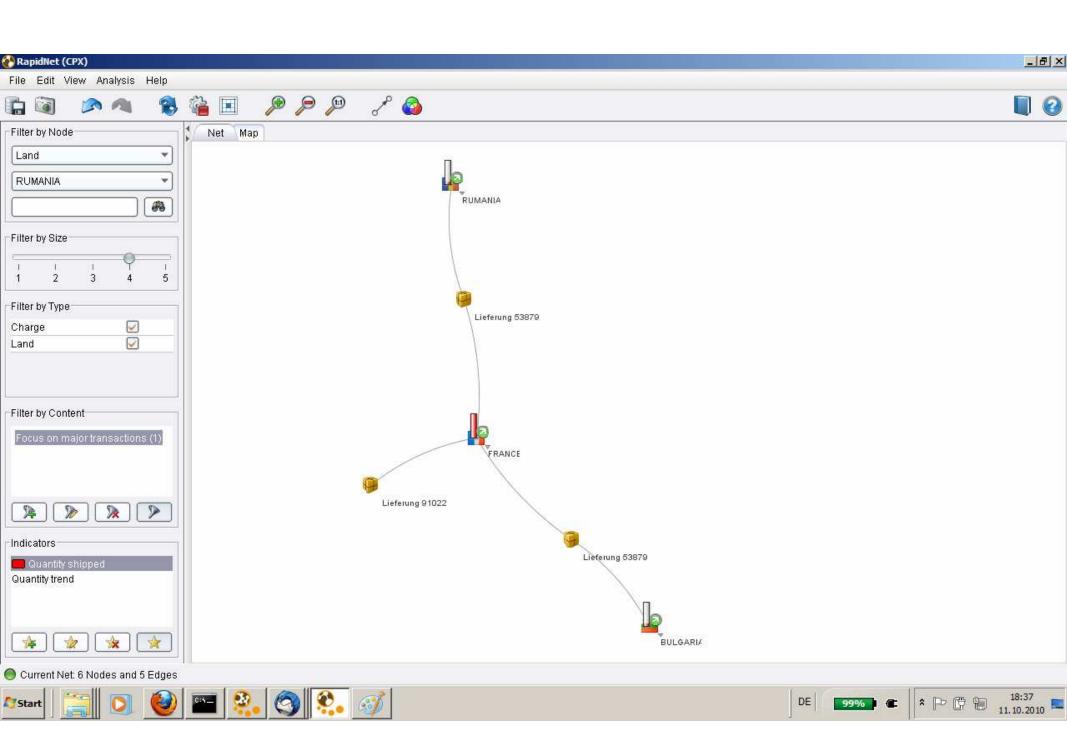


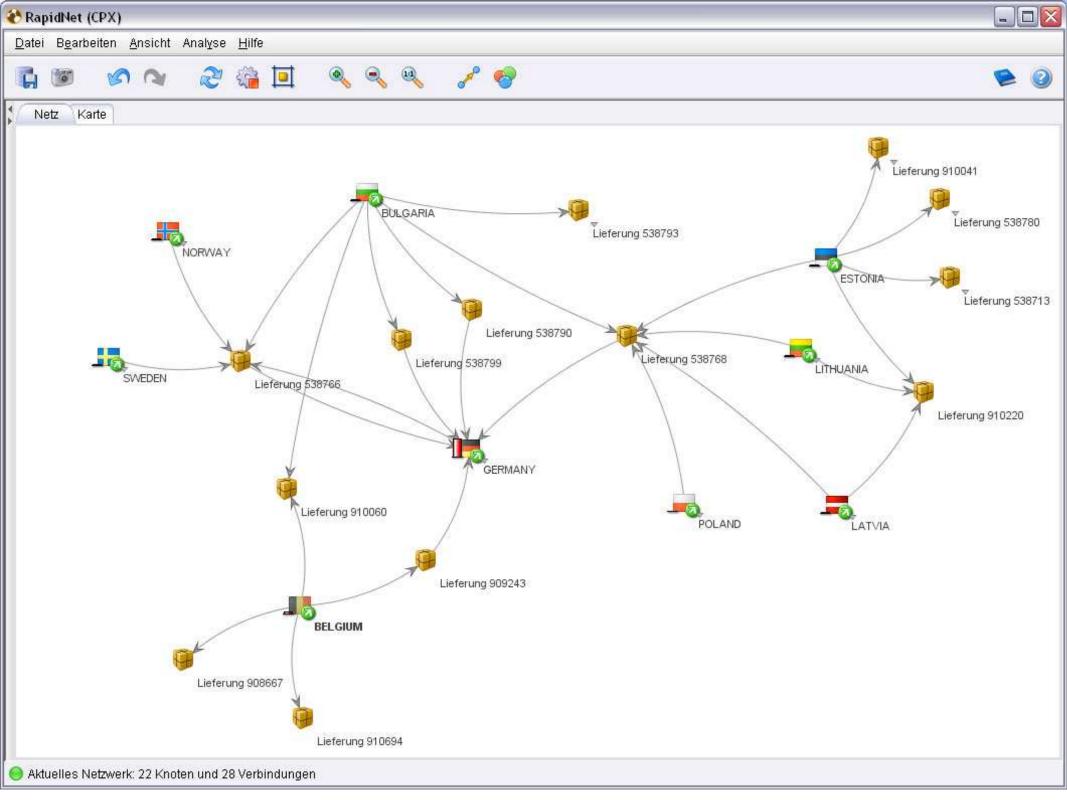


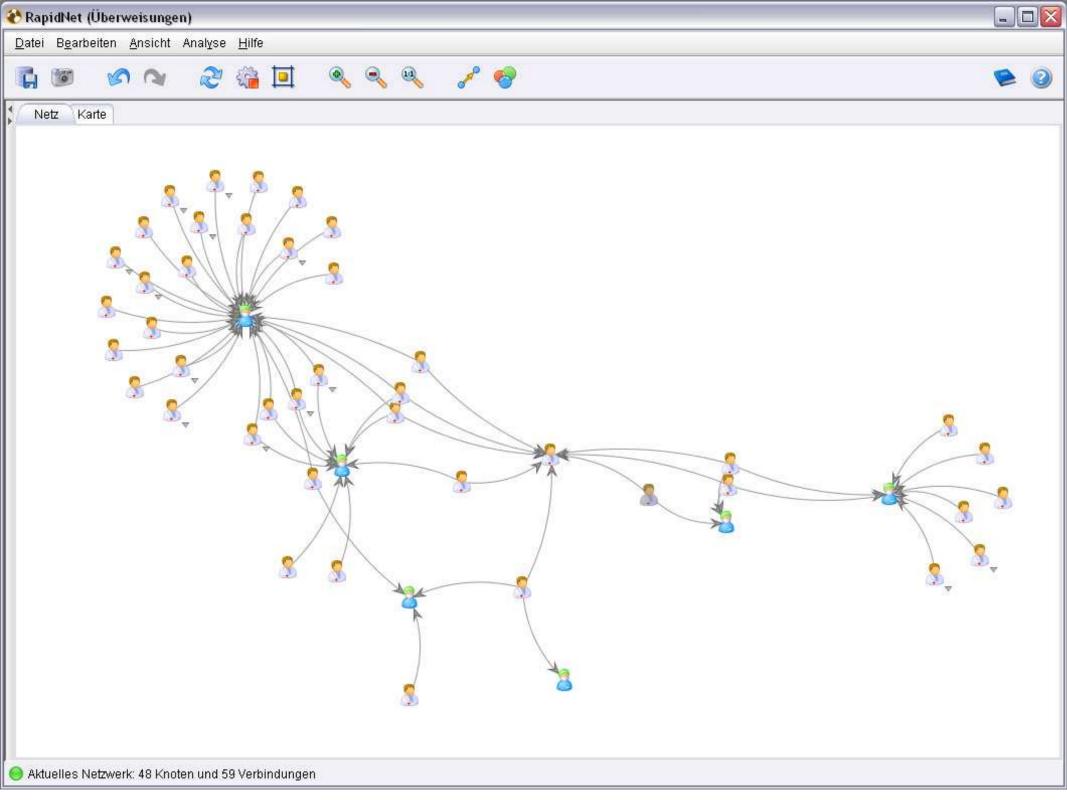


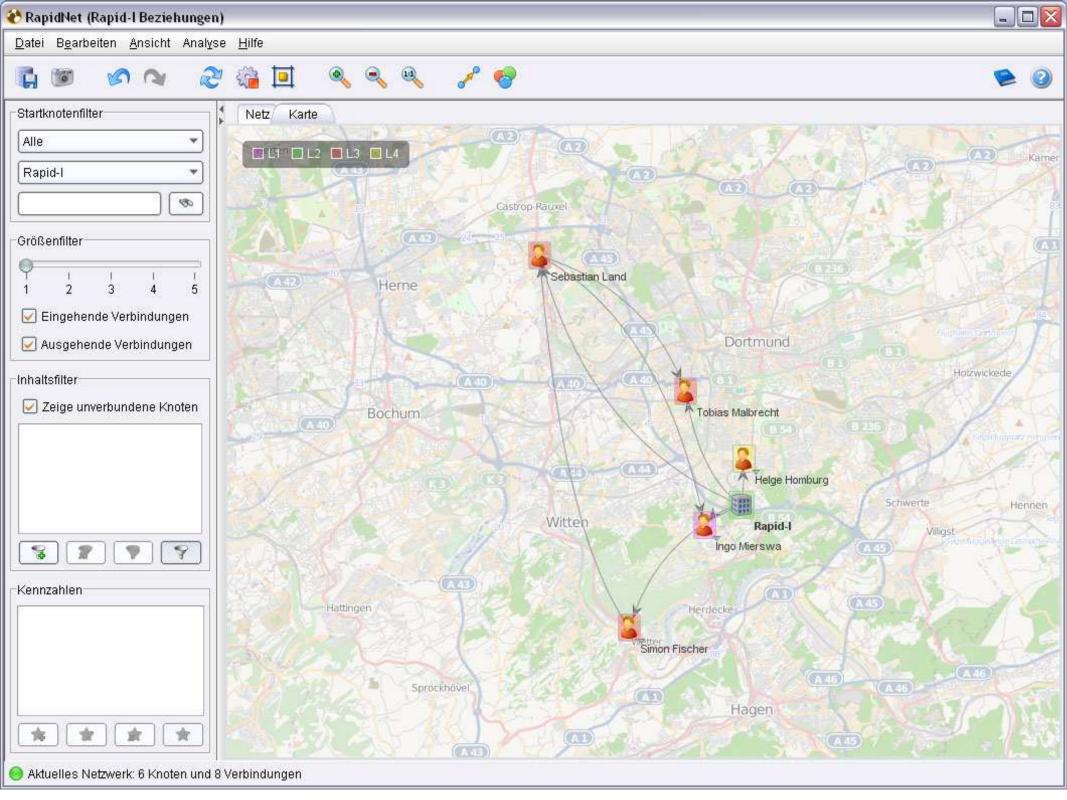


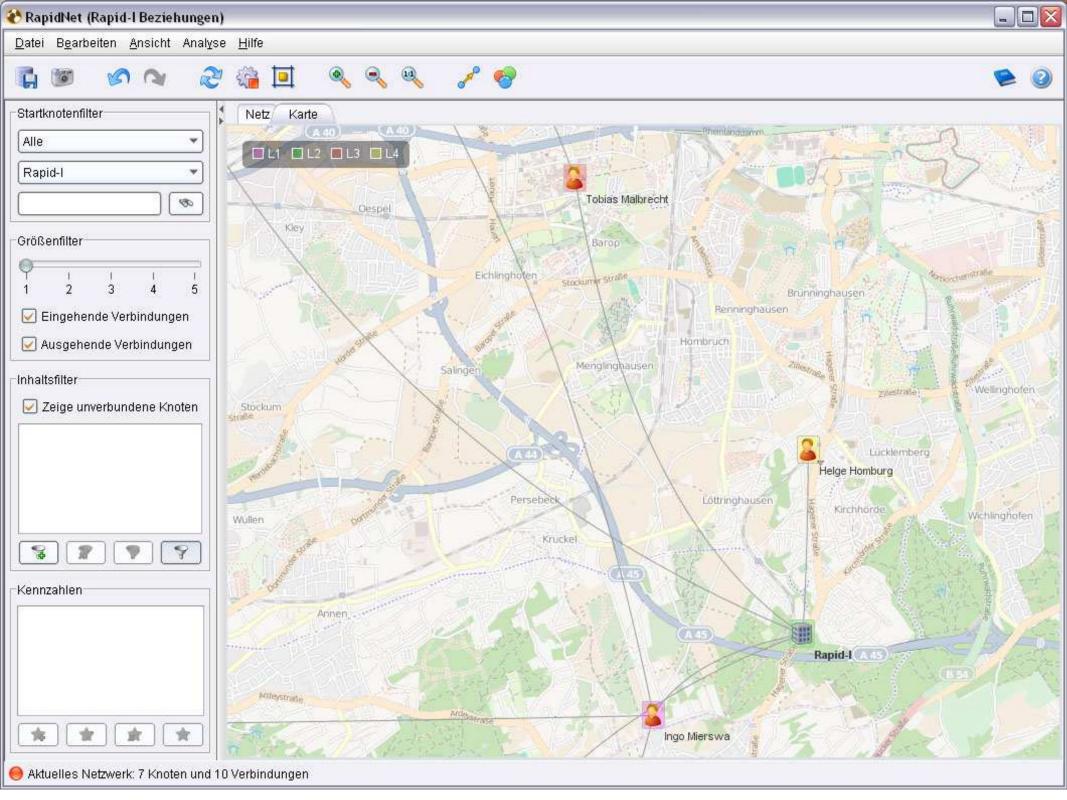








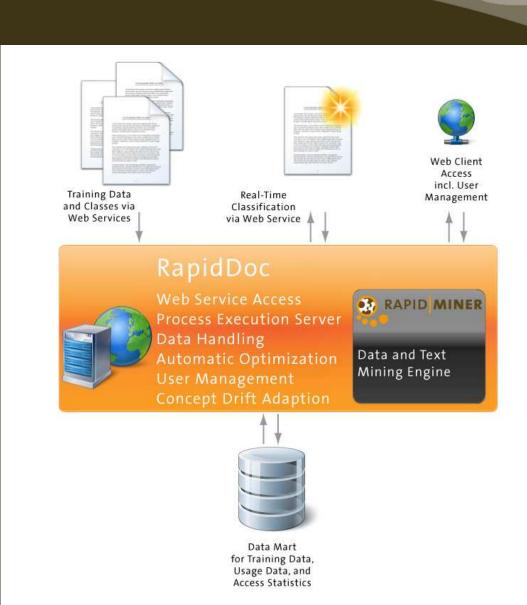






RapidDoc

- Automated text categorization for documents like e-mails, news, web pages, customer feedback, ...
- Automatically score texts according to their content or sentiment
- Real-time scoring of texts and easy integration via web service interface
- Easy to train and to use
- Hosted web service at Rapid-I
 => no effort on your side: configuration, optimization, and maintenance by Rapid-I experts
- Or inhouse at your company:
 RapidAnalytics server plus set-up and configuration project



Rapid-I Market Place



http://marketplace.rapid-i.com/

- Enables developers to share their extensions with the large RapidMiner community
- Allows RapidMiner
 users to easily install
 and deploy these
 extensions with a few
 clicks
- Sharing of free and commercial extensions



Future Directions





- Scalability to even larger data volumes ("Big Data"):
 - Parallelization on multi-processor and multi-core computers
 - => RapidMiner and RapidAnalytics extension (available)
 - In-Database Mining => RapidAnalytics Enterprise Edition
 - Profiler => RapidAnalytics Enterprise Edition
 - Hadoop: Map & Reduce=> Radoop = RapidMiner Hadoop Extension
 - Cluster Computing
 - => RapidAnalytics Enterprise Edition (under development)
- Ease of use and reusability of RapidMiner processes:
 - Building Blocks
 - => RapidMiner and RapidAnalytics (built-in)
 - Process Embedder
 - => RapidMiner and RapidAnalytics (built-in)
 - Operator-Libray: Using complete RapidMiner processes like configurable operators in other processes
 - => RapidMiner and RapidAnalytics (under development)
- Your ideas and those of the community...





Data Mining Services by Rapid-I

- Consulting
- Training courses and inhouse workshops
- Professional support (guaranteed response times)
- Individual product adaptations and extensions
- Individual software development
- System integration
- RapidMiner data mining processes as web services
- Complete data analysis (out-sourcing to Rapid-I)
- RapidMiner & RapidAnalytics Enterprise Edition =
 Community Edition + Additions + Support & Services
 + Guarantees + Intellectual Property Idemnification
- RapidMiner OEM Licenses for your products



Selected Customers / RapidMiner Users

Thank you for your attention!

More information at: www.rapid-i.com















































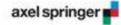
















About Rapid-I

- Founded in 2006 by the developers of RapidMiner
- Best IT start-up in the nation-wide German start-up competition start2grow 2006/2007 for all industries
- Winner of the most highly rewarded European start-up price Open Source Business Award 2008
- RapidMiner awarded as one of the most innovative open source products by the German initiative for small and medium sized businesses (Innovationspreis 2008, Innitiative Mittelstand)
- Rapid-I provides innovative data mining and business intelligence solutions increasing the efficiency and profitability of its customers
- Data mining software development, consulting, and services based on RapidMiner and beyond





