VISUAL ANALYTICS

APPLICATION AREAS

LECTURE 8

Petra Isenberg

informatics mathematics
Inria

TOC

- Business Intelligence
- Legal Matters
- Work at Aviz

ANALYSIS IN THE BUSINESS SECTOR

BUSINESS INTELLIGENCE

"An interactive process for exploring and analyzing structured and domain-specific information to discern trends or patterns, thereby deriving insights and drawing conclusions.

The business intelligence process includes communicating findings and effecting change."

Gartner

→ there are many more definitions and they don't all say the same thing

BUSINESS INTELLIGENCE

- Typical Goals:
  - increase profitability
  - decrease costs
  - improve customer relationship management
  - decrease risks (e.g. credit risk analysis)
- Main goal: aid in making decisions

David Loshin: **Business Intelligence: The Savvy Manager's Guide, Getting Onboard with Emerging IT (2003)**

ANALYSIS IN ENTERPRISES IN GENERAL

- business analysts
- data analysts
- data scientists

→ typical user population for analysis & visualization tools

Sean Kandel, Andreas Paepcke, Joseph M. Hellerstein, Jeffrey Heer: Enterprise Data Analysis and Visualization: An Interview Study. IEEE Trans. Vis. Comput. Graph. 18(12): 2917-2926 (2012

A STUDY

researchers conducted semi-structured interviews in enterprises:

- 35 analysts (26 male)
- 25 organizations: healthcare, retail, finance, social networking, insurance, …

Sean Kandel, Andreas Paepcke, Joseph M. Hellerstein, Jeffrey Heer: Enterprise Data Analysis and Visualization: An Interview Study. IEEE Trans. Vis. Comput. Graph. 18(12): 2917-2926 (2012

QUESTIONS

- What tasks do analysts perform?
- What kinds of data sources and formats do they work with?
- What tools do they regularly use and how do they use them?
- How do analysts vary in terms of programming proficiency?
- How do analysts vary in terms of statistical proficiency?
- What are the "results" of analysis?
- What happens to these results "downstream"?
- What are recurring bottlenecks and pain points?
- How important is scalability?
- How important is sociability?
- What is the relationship between analysts and other business units?
- Where are analysts situated within their corporate hierarchy?

THE ANALYST

Three archetypes found

1) hackers

- most comfortable manipulating data
- used 3+ programming languages (R, Python, SQL, …)
- complex workflows
- work quite isolated (don't need a lot of help)
- used visualizations: Tableau, Excel, PPT, D3, …

THE ANALYST

Three archetypes found

2) scripters

- most analysis done in R/Matlab
- not versed in custom operations (parsing, scraping)
- generally worked on data from data warehouse (with help from IT staff)
- applied models to data
- did exploratory analysis with visualization

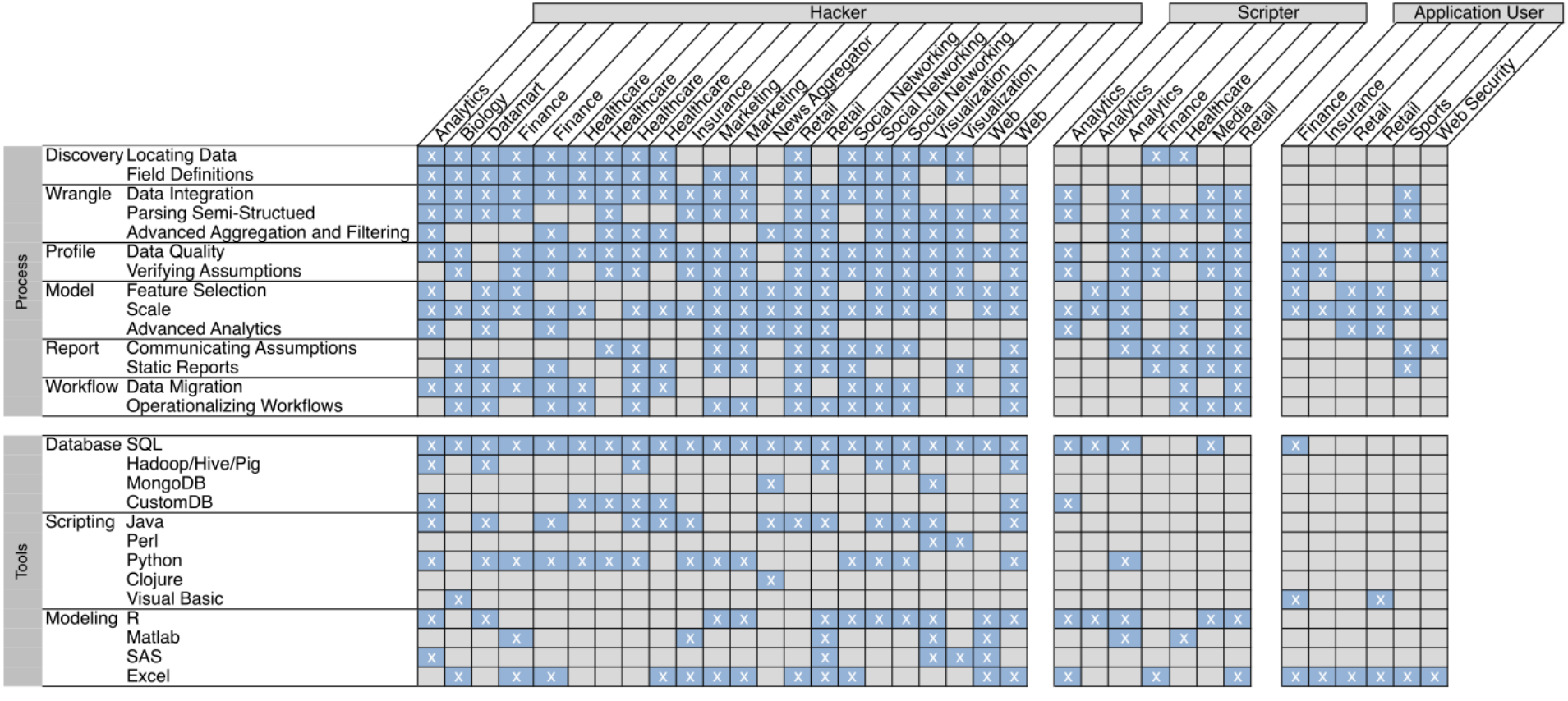
THE ANALYST

Three archetypes found

3) application user

- performed most operation in spreadsheet or analysis app (SAS/JMP, SPSS, …)
- needed help preparing data
- typically worked on smaller datasets

ANALYSTS


Hacker
Scripter
Application User
Analytics
Biology
Datamart
Finance
Finance
Healthcare
Healthcare
Healthcare
Healthcare
Insurance
Marketing
Marketing
News Aggregator
Retail
Retail
Social Networking
Social Networking
Social Networking
Visualization
Visualization
Web
Web
Analytics
Analytics
Analytics
Finance
Healthcare
Media
Retail
Finance
Insurance
Retail
Retail
Sports
Web Security
Process
Discovery
Locating Data
Field Definitions
Wrangle
Data Integration
Parsing Semi-Structued
Advanced Aggregation and Filtering
Profile
Data Quality
Verifying Assumptions
Model
Feature Selection
Scale
Advanced Analytics
Report
Communicating Assumptions
Static Reports
Workflow
Data Migration
Operationalizing Workflows
Tools
Database
SQL
Hadoop/Hive/Pig
MongoDB
CustomDB
Scripting
Java
Perl
Python
Clojure
Visual Basic
Modeling
R
Matlab
SAS
Excel

IT TEAM

- Crucial role in process
  - getting, maintaining, accessing data
  - operationalize workflows
  - provide documentation

OTHER FINDINGS

- data:
  - stored in variety of repositories and formats
- consumers of analysis:
  - many different departments, also other analysts
  - static reports shared most commonly, sometimes dynamic dashboards

OTHER FINDINGS

- collaboration
  - work on analysis task mostly done individually
  - resources shared, however: data, scripts, results, documentation

REPORTED CHALLENGES

- discovery
  - where is my data?
  - what does my data mean? (unclear field names, missing units, timezones, …)

REPORTED CHALLENGES

- wrangling
  - processing semi-structured data
  - data integration from multiple sources
  - advanced aggregation and filtering

> *I spend more than half of my time integrating, cleansing and transforming data without doing any actual analysis. Most of the time I'm lucky if I get to do any analysis. Most of the time once you transform the data you just do an average... the insights can be scarily obvious. It's fun when you get to do something somewhat analytical.*

REPORTED CHALLENGES

- profiling
  - assessing & improving data quality (missing data, wrong formats, …)
  - checking assumptions (data distribution, semantics of data, …)

REPORTED CHALLENGES

- modeling
  - finding the right features to analyze
  - scale of data
  - visualization of statistical models missing

REPORTED CHALLENGES

- reporting
  - how to document assumptions
  - flexibility in reports missing

REPORTED CHALLENGES

- workflows are non-linear
  - moving data between tools necessary (creates formatting issues)
  - creating repeatable, reliable, and scalable workflows

BUSINESS ANALYSIS TOOLS

a research view

TOOLS PREDECESSORS

- Management Information Systems (MIS)
- Management Decision Systems (MDS)
- Decision Support Tools (DSS)
- Executive Information Systems (EIS)
- Analysis Information Systems (AIS)
- …

TOOLS

Warning: I haven't tried any of them, so can't make recommendations

- Professional
  - SAP Business Intelligence
  - IBM Cognos
  - …
- Open Source
  - SpagoBI
  - Pentaho
  - …
- Research
  - see next slides

DOTLINK360

Basole, R.C.; Clear, T.; Mengdie Hu; Mehrotra, H.; Stasko, J., "Understanding Interfirm Relationships in Business Ecosystems with Interactive Visualization," *Visualization and Computer Graphics, IEEE Transactions on* , vol.19, no.12, pp.2526,2535, Dec. 2013
doi: 10.1109/TVCG.2013.209

2013

dotlink360

Understanding Interfirm Relationships in Business Ecosystems with Interactive Visualization

Rahul C. Basole | Trustin Clear | Mengdie Hu | Harshit Mehrotra | John Stasko
contact: basole@gatech.edu

Georgia Institute of Technology

BUSINESS ECOSYSTEM INTELLIGENCE

- market analysts
  - understand competitive trends/strategies/threats/opportunities
- executives
  - identify strategic collaborations & customers, find areas for innovation
- venture capitalists
  - identify investment opportunities, see how they fit in the business landscape

DOTLINK360


Navigator
Ecosystem Segments
Application & Software Providers
Cable Providers
Content Providers
Contract Electronics Manufacturers
Device Manufacturers
Gaming Providers
Internet Service Providers
Media & Entertainment Providers
Mobile Network Operators
Network & Infrastructure Providers
Photography & Digital Imaging
Platform Providers
Resellers & Retailers
Service & Billing Providers
Silicon Vendors & Other Component Providers
Companies
QNX Software Systems Ltd
QSC AG
QSound Labs Inc
QuadraMed Corp
QUALCOMM Inc
Qualmark Corp
Qualstar Corp
Qualys Inc
Quanta Computer Inc
Quantum Corp
Quantum Fuel Systems Tech Inc
Quark Inc
Quebecor Media Inc
Quepasa Corp
Quest Software Inc
QuickLogic Corp
Countries and Regions
Argentina
Australia
Austria
Belgium
Bermuda
Brazil
Canada
China
Czech Republic
Denmark
Egypt
Finland
France
Germany
Home
Path
Segment
Scatternet
Composition
Timeline
Geography
Content Providers
System Integrators
Contract Electronics Manufacturers
Service & Billing Providers
Mobile Network Operators
Internet Service Providers
Platform Providers
Device Manufacturers
Cable Providers
Gaming Providers
QUALCOMM Inc
Photography & Digital Imaging
Nokia Oyj
Resellers & Retailers
Apple Inc
Media & Entertainment Providers
Application & Software Providers
Intel Corp
Texas Instruments Inc
Silicon Vendors & Other Component Providers
Network & Infrastructure Providers
Agreement Details
EXPAND ALL
SORT BY
SELECT
Apple Inc
10/29/2012 Rumor
Partners: Sharp Corp, Google Inc, Microsoft Corp
Description: Sharp Corp (Sharp), Apple Inc (Apple), Google Inc (Google) and Microsoft Corp (Microsoft) was rumored to be planning to form a strategic alliance under which Sharp will provide Apple, Google and Microsoft with small and medium-sized Igzo power-saving LCD panels.
06/09/2008 Completed/Signed
05/00/2008 Completed/Signed
05/07/2008 Completed/Signed
12/03/2007 Completed/Signed
05/03/2006 Completed/Signed
05/03/2006 Completed/Signed
05/03/2006 Pending
12/06/2005 Completed/Signed
10/17/2005 Terminated
09/07/2005 Completed/Signed
07/06/2004 Pending
04/05/2004 Rumor
01/08/2004 Pending
07/03/2002 Pending
07/03/2000 Pending
05/04/2000 Completed/Signed
09/09/1999 Completed/Signed
07/09/1999 Completed/Signed
07/09/1999 Pending
01/02/1999 Completed/Signed
01/08/1999 Pending
12/04/1998 Completed/Signed
08/00/1998 Completed/Signed
03/07/1998 Completed/Signed
Agreement Type
Strategic
R &D
Manufacturing
Supply
Marketing
Licensing
Other
Company Information
Agreement Details
Structural Properties

DESIGN PROCESS

- field study of analysts
- derive set of design requirements
- develop initial version
- user testing
- redesign

DATA SOURCES

- Thomson Reuters SDC Platinum database
  - commercial database
  - ca. 700,000 global alliances, agreements, joint-ventures, since 1985
- Capital IQ Compustat
  - contains e.g. quarterly financial and accounting data for active and inactive publicly listed companies

DATA TRANSFORMATION

- turned data into a network
  - nodes = companies
  - edges = agreements
- nodes have multiple attributes
- edges have multiple types
- +temporal data

→ time-varying multivariate network

DESIGN REQUIREMENTS

- Field study
  - online survey + interviews
  - 24 senior industry individuals (market analysts, executives, venture capitalists)
  - each >10 years of experience

DESIGN REQUIREMENTS

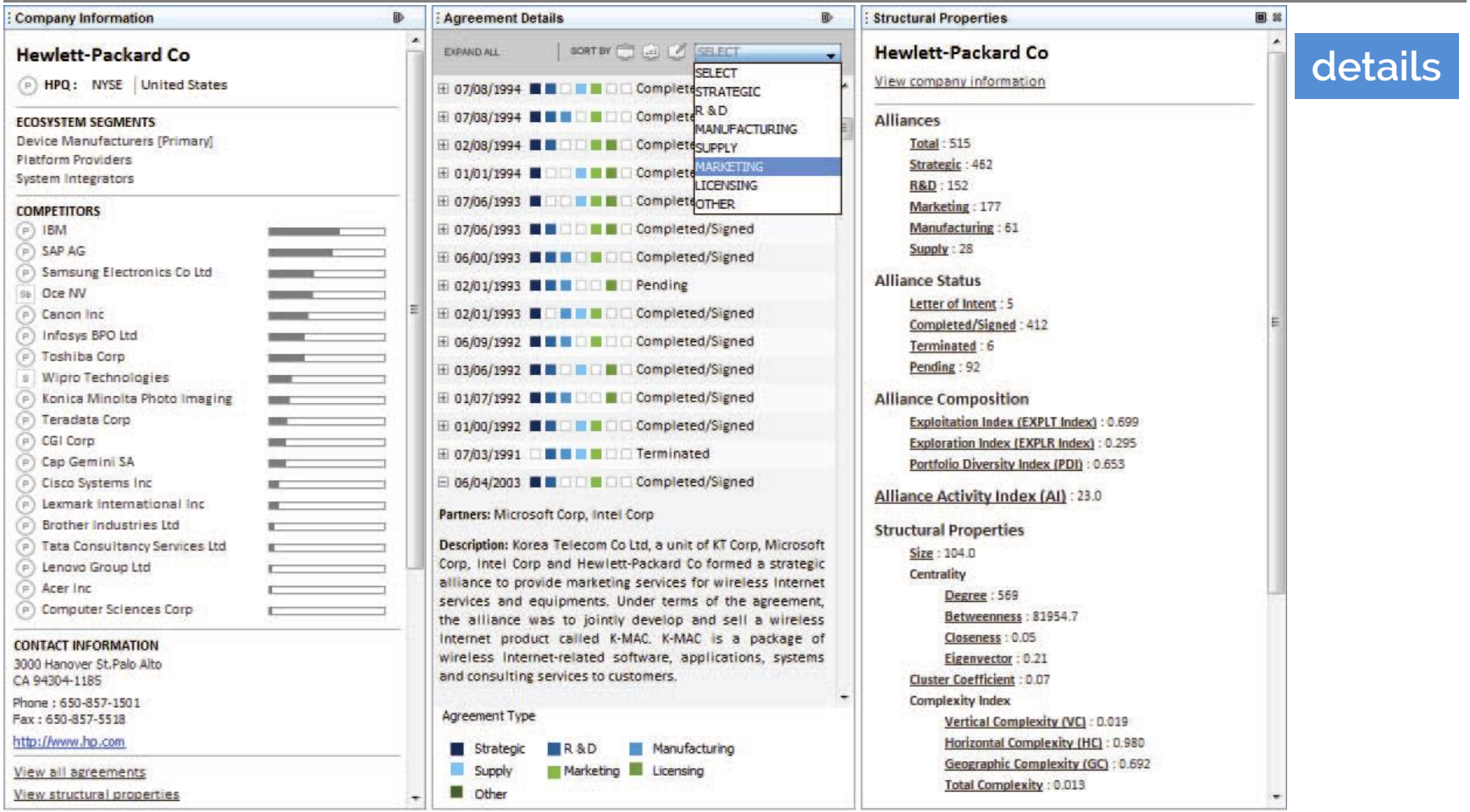
top-down & bottom-up examination of an ecosystem are critical

"*it is helpful to have a birds-eye view of the ecosystem, but at the same time [the system] should be able to allow users to drill down into individual companies and segments*."

DESIGN REQUIREMENTS

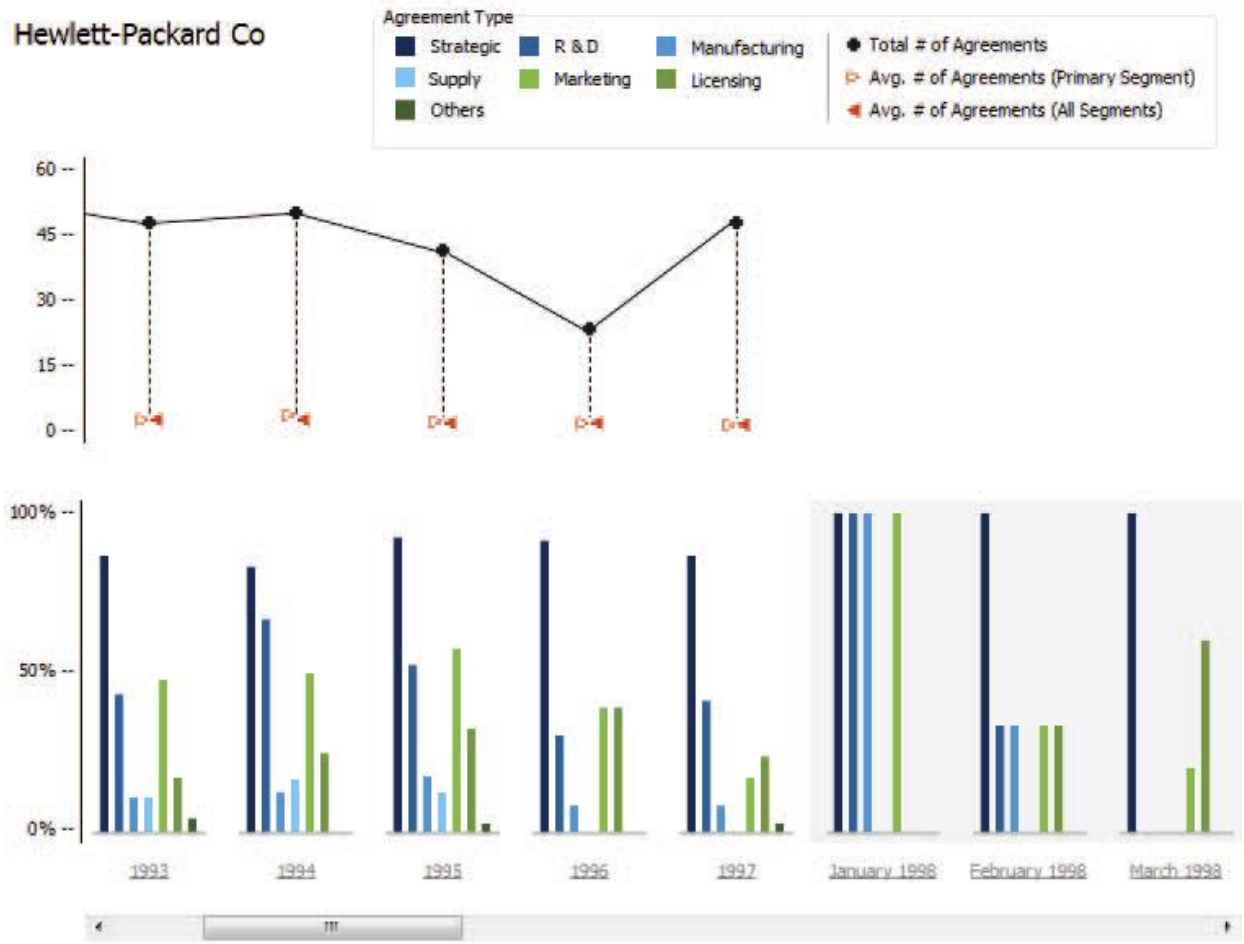
- understanding interfirm connectivity, composition, and temporality is vital
- comparative perspectives drive insight
- first: communicate agreement summaries (structural information) → then: offer details
- provide a familiar metaphor while supporting direct and prompt interaction, not complex queries and commands
- add common network-related analysis tasks (see the InfoVis lecture on graphs and networks)

MULTIPLE VIEWS



details

MULTIPLE VIEWS

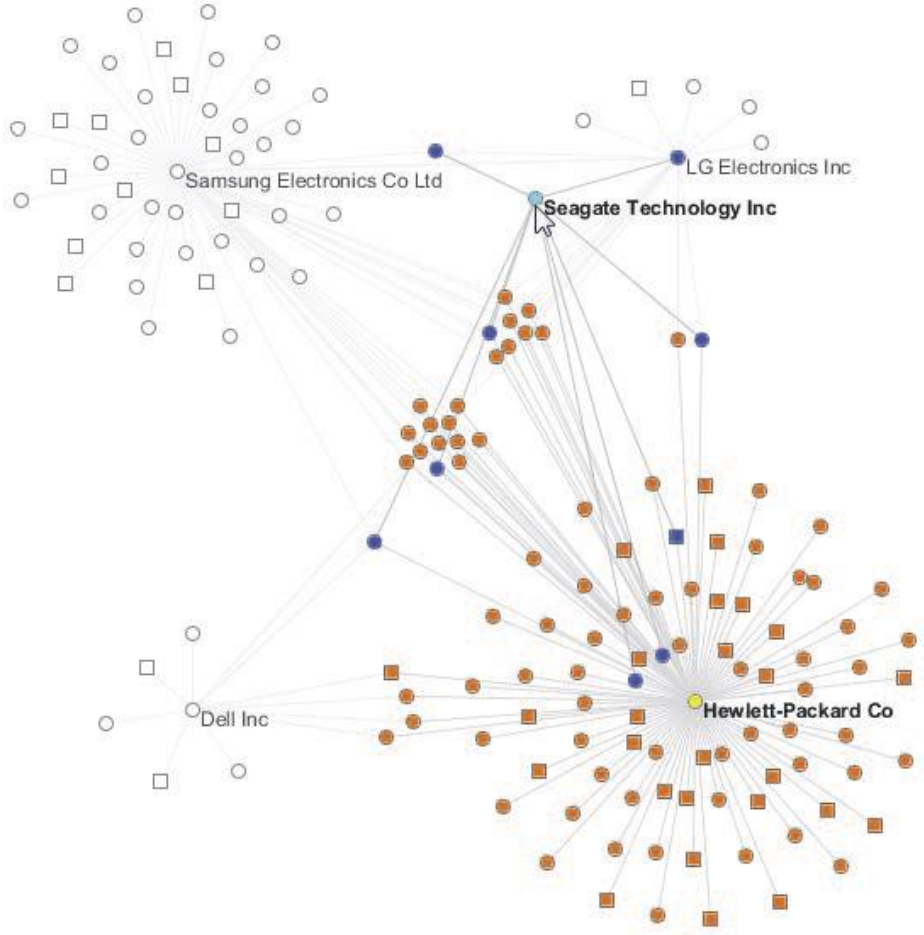


timeline

MULTIPLE VIEWS

portfolio overview

MULTIPLE VIEWS



comparison

... and many more

WIREVIS

Remco Chang, Mohammad Ghoniem, Robert Kosara, William Ribarsky, Jing Yang, Evan Suma, Caroline Ziemkiewicz, Daniel Kern, Agus Sudjianto

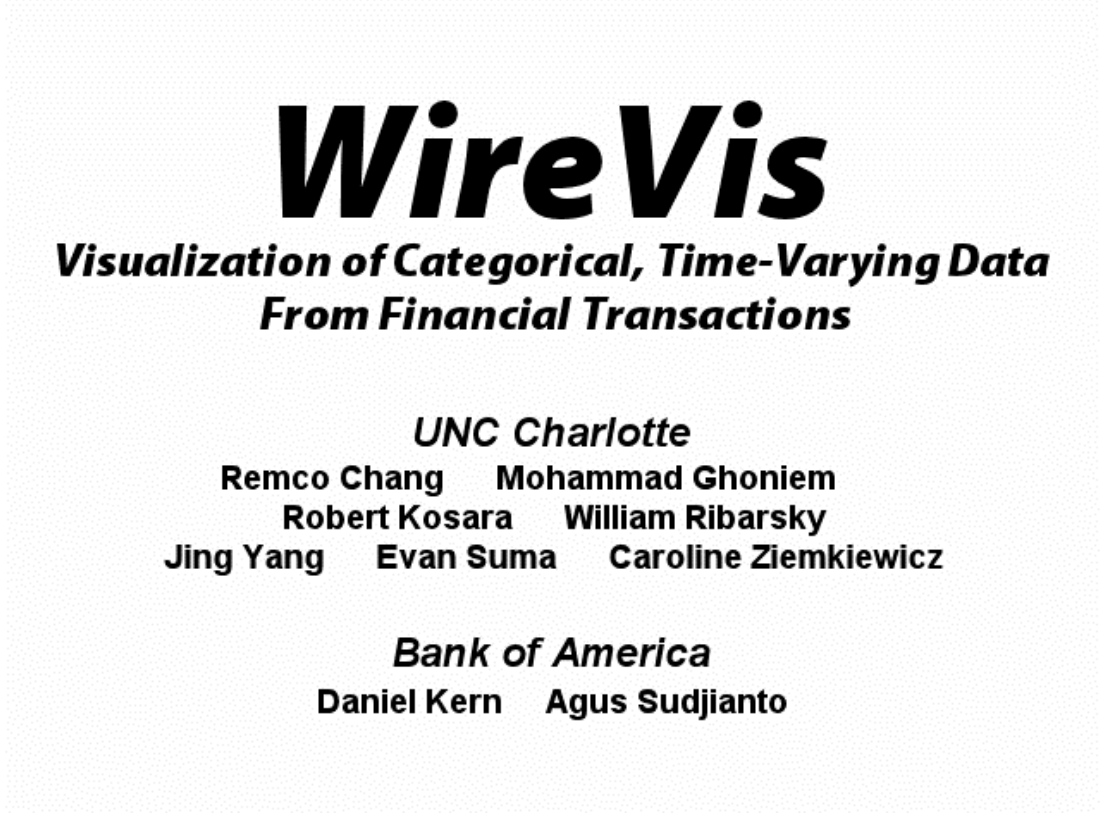
IEEE Visual Analytics Science and Technology (VAST) 2007.

- financial institutions have obligation to discover suspicions financial transactions
  - can be fined if not found or shut down
- many transactions are purely digital & banks often are only the middle "man"
- large banks need to monitor hundreds of thousands of transactions per day

WIREVIS

Remco Chang, Mohammad Ghoniem, Robert Kosara, William Ribarsky, Jing Yang, Evan Suma, Caroline Ziemkiewicz, Daniel Kern, Agus Sudjianto

IEEE Visual Analytics Science and Technology (VAST) 2007.



https://www.youtube.com/watch?v=RPKcrLQBqiE

WIREVIS

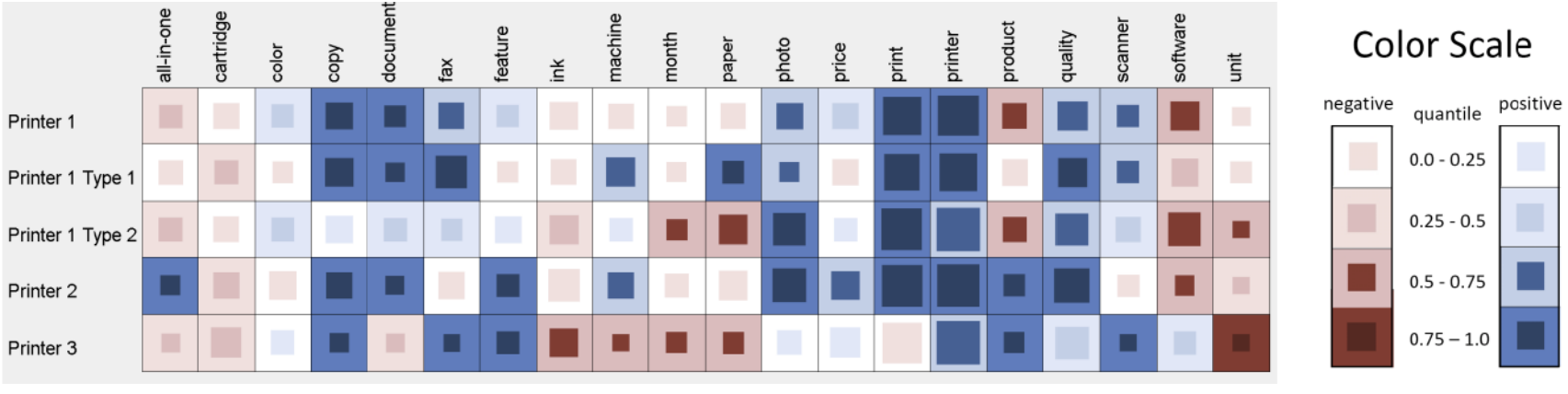
goals

- provide overview of transactions over any period of time → apply to large data
- allow identification of patterns over time and keywords
- replace blind queries with in-place analysis
- provide search-by-example technique

WIREVIS

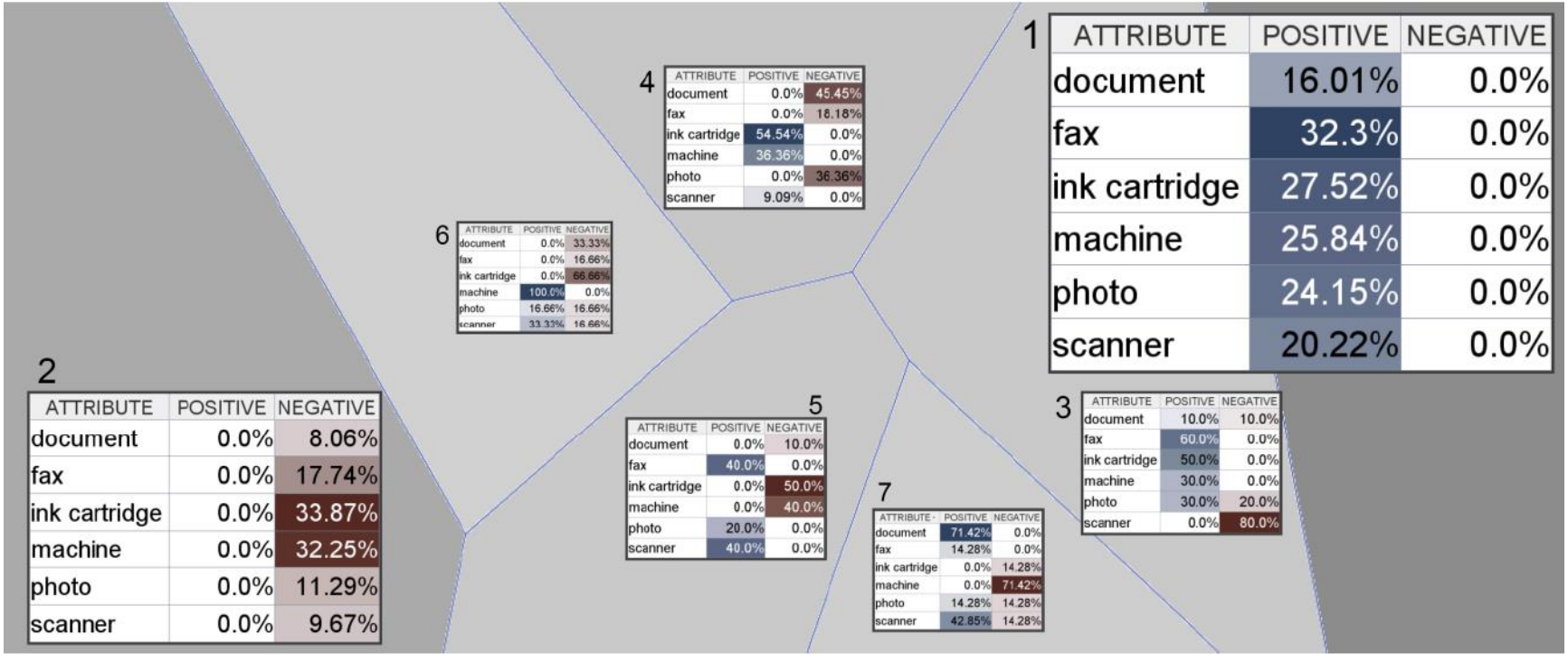
- wire transaction data
  - semi-structured data record
  - fixed data (sender, receiver, etc.)
  - optional free text
  - temporal, categorical, quantitative
- current (2007) procedure
  - filter based on risk assessment (e.g. money exceed threshold, sender/receiver is a high-risk country or organization)
  - look at spreadsheets

VISUAL ANALYSIS OF CUSTOMER FEEDBACK



Oelke, D.; Ming Hao; Rohrdantz, C.; Keim, D.A.; Dayal, U.; Haug, L.; Janetzko, H., "Visual opinion analysis of customer feedback data," *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on* , vol., no., pp.187,194, 12-13 Oct. 2009
doi: 10.1109/VAST.2009.5333919

VISUAL ANALYSIS OF CUSTOMER FEEDBACK



1

ATTRIBUTE	POSITIVE	NEGATIVE
document	16.01%	0.0%
fax	32.3%	0.0%
ink cartridge	27.52%	0.0%
machine	25.84%	0.0%
photo	24.15%	0.0%
scanner	20.22%	0.0%

2

ATTRIBUTE	POSITIVE	NEGATIVE
document	0.0%	8.06%
fax	0.0%	17.74%
ink cartridge	0.0%	33.87%
machine	0.0%	32.25%
photo	0.0%	11.29%
scanner	0.0%	9.67%

3

ATTRIBUTE	POSITIVE	NEGATIVE
document	10.0%	10.0%
fax	60.0%	0.0%
ink cartridge	50.0%	0.0%
machine	30.0%	0.0%
photo	30.0%	20.0%
scanner	0.0%	80.0%

4

ATTRIBUTE	POSITIVE	NEGATIVE
document	0.0%	45.45%
fax	0.0%	18.18%
ink cartridge	54.54%	0.0%
machine	36.36%	0.0%
photo	0.0%	36.36%
scanner	9.09%	0.0%

5

ATTRIBUTE	POSITIVE	NEGATIVE
document	0.0%	10.0%
fax	40.0%	0.0%
ink cartridge	0.0%	50.0%
machine	0.0%	40.0%
photo	20.0%	0.0%
scanner	40.0%	0.0%

6

ATTRIBUTE	POSITIVE	NEGATIVE
document	0.0%	33.33%
fax	0.0%	16.66%
ink cartridge	0.0%	66.66%
machine	100.0%	0.0%
photo	16.66%	16.66%
scanner	33.33%	16.66%

7

ATTRIBUTE	POSITIVE	NEGATIVE
document	71.42%	0.0%
fax	14.28%	0.0%
ink cartridge	0.0%	14.28%
machine	0.0%	71.42%
photo	14.28%	14.28%
scanner	42.85%	14.28%

LEGAL MATTERS

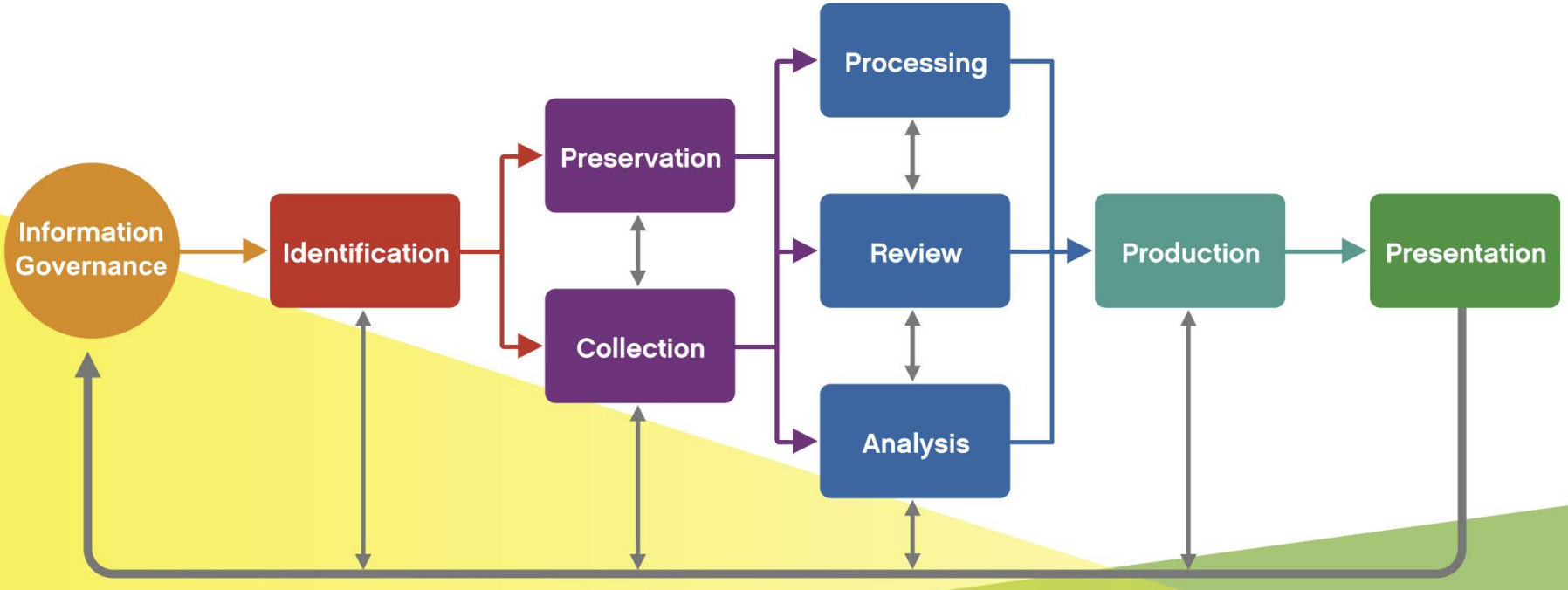
or

working with large corpora of electronic texts

E-DISCOVERY (MOSTLY USA)

- legal electronic document discovery
  - for use in law suits
  - regulatory information requests
  - investigations, audits, freedom of information act, …
- documents typically given to court & opponent
- if complaint received, corporations have to produce all related information

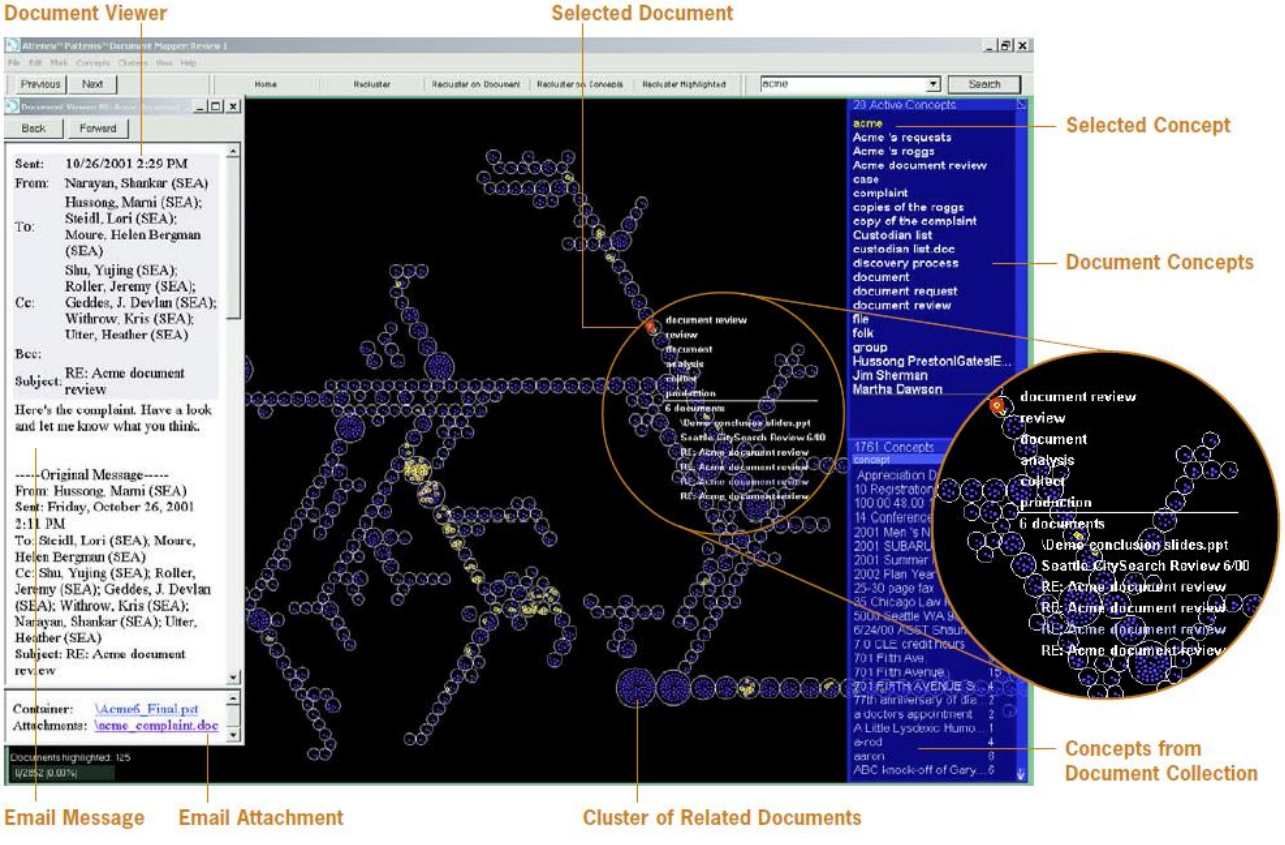
Electronic Discovery Reference Model



https://www.youtube.com/watch?v=eo03DWk4_lU

EXAMPLE TOOLS

Commercial: Attenex Patterns (now FTI Ringtail)



SUCCESS STORIES OF VA TOOL

- Lovells (6th largest law firm worldwide) was investigating 35Gb of email data
  - traditional method: estimated cost: 1 year, 4-5 million $US
  - with VA software: three months, cost 1 million $US
- JMBM law form represented spinal surgeon & inventor in intellectual property suit
  - 20 claims, 50 million pages of documents, contracts, patents, etc.
  - with VA software (used on 7 computers), 44 million pages sifted through in 4 months
  - important doc found, surgeon won, received $570 million in compensation

Figure 1. Magic Quadrant for E-Discovery Software

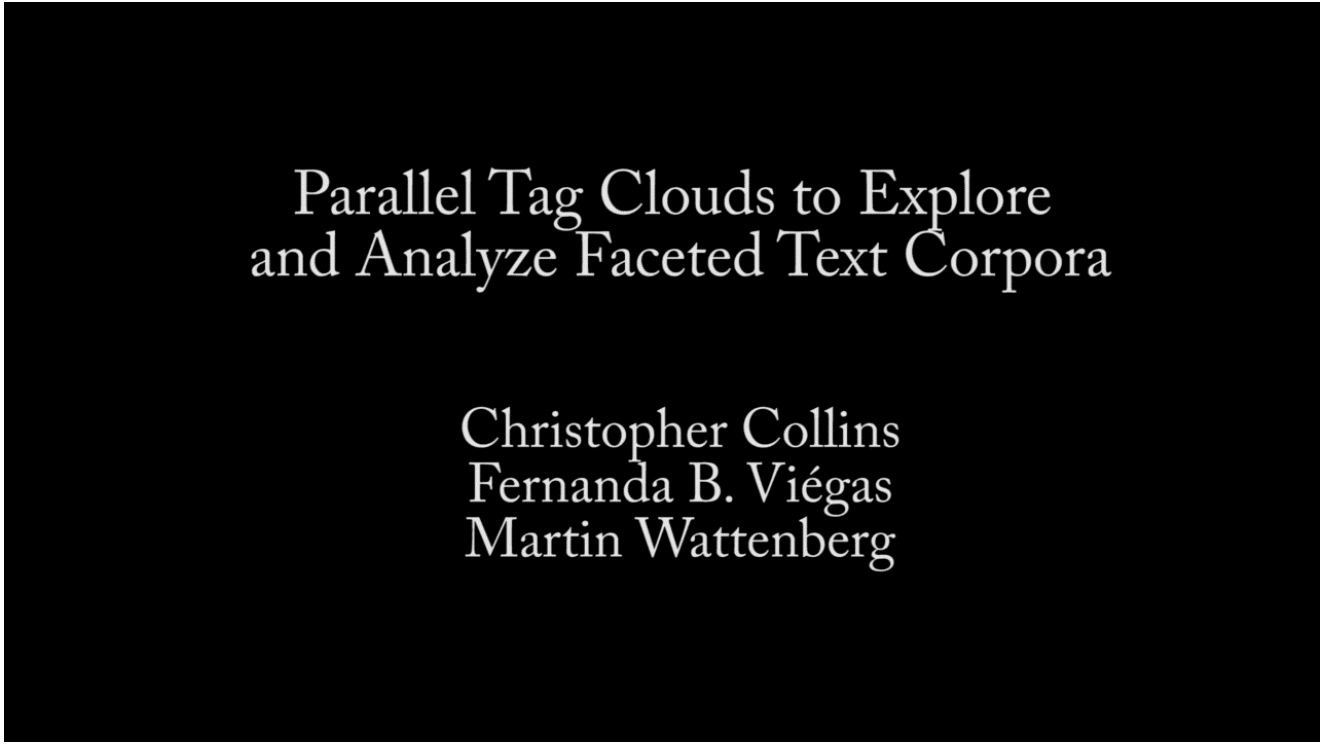
Source: Gartner (June 2014)

RESEARCH

- working on this level in research is difficult
  - man power for providing support
  - “research aspect” needs most attention
  - adoption of prototypes unlikely when stakes are high ($$$)
- thus, next: smaller research projects in regards to document analysis

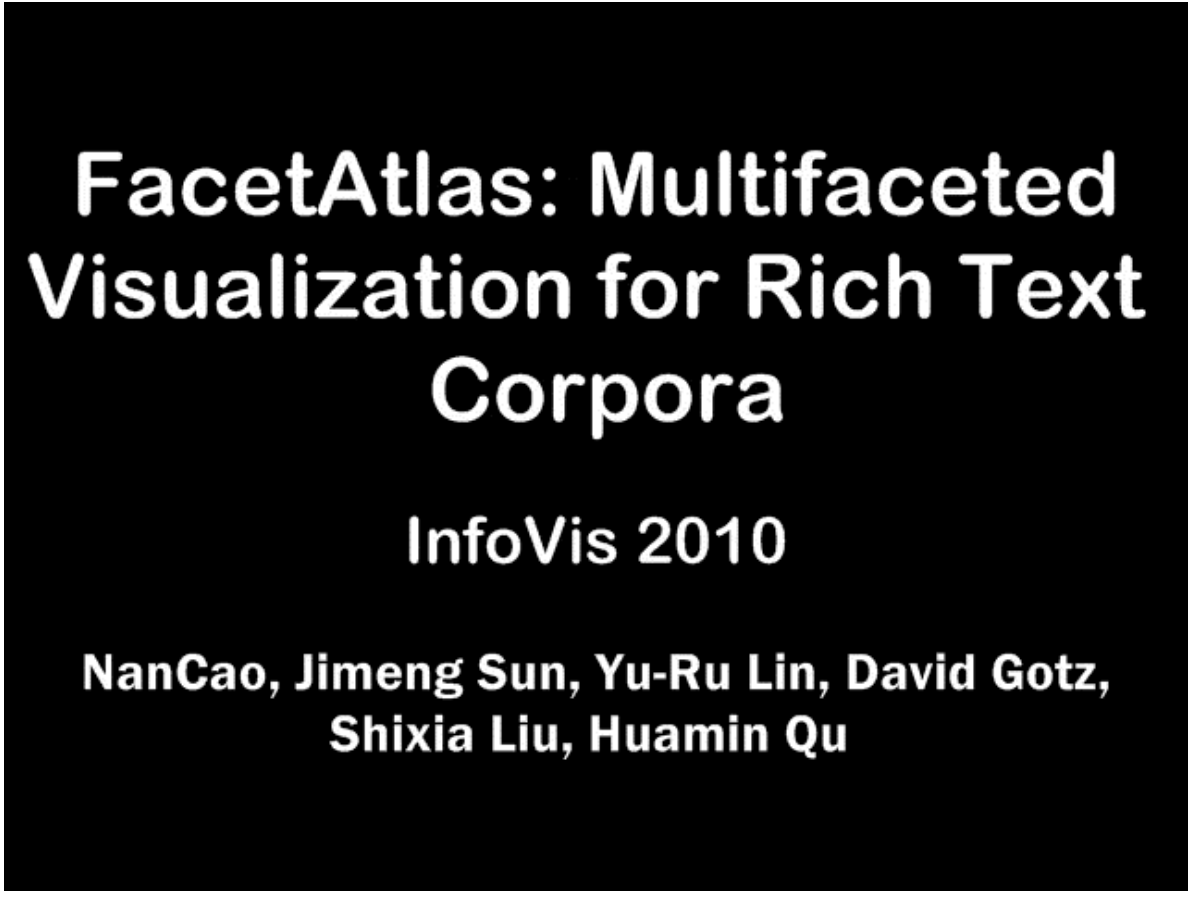
PARALLEL TAG CLOUDS

C. Collins, F. B. Viégas, and M. Wattenberg, “Parallel Tag Clouds to Explore and Analyze Facted Text Corpora,” in *Proc. of the IEEE Symp. on Visual Analytics Science and Technology (VAST)*, 2009.

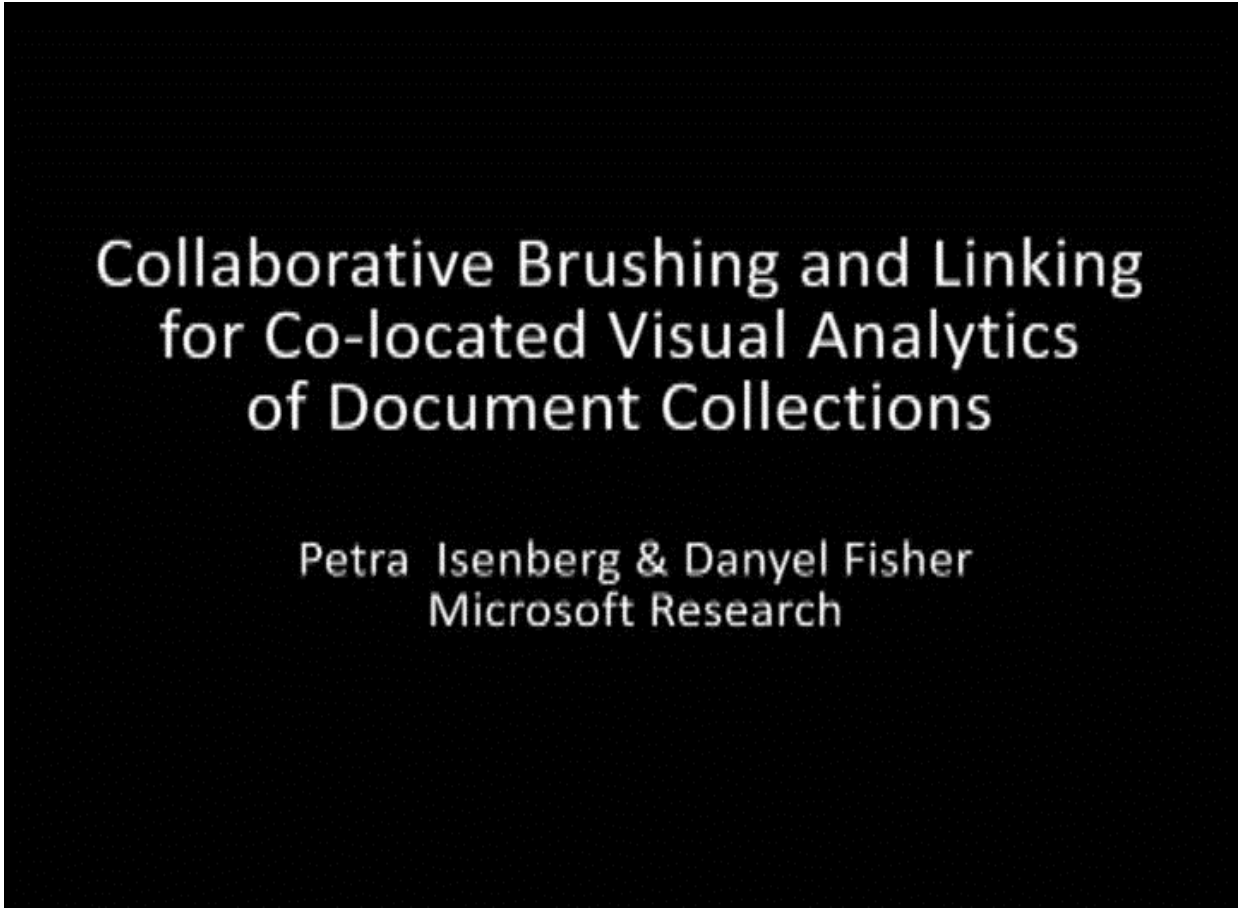


FACET ATLAS

Nan Cao; Jimeng Sun; Yu-Ru Lin; Gotz, D.; Shixia Liu; Huamin Qu, "FacetAtlas: Multifaceted Visualization for Rich Text Corpora," *Visualization and Computer Graphics, IEEE Transactions on* , vol.16, no.6, pp.1172,1181, Nov.-Dec. 2010
doi: 10.1109/TVCG.2010.154



CAMBIERA


Collaborative Brushing and Linking
for Co-located Visual Analytics
of Document Collections
Petra Isenberg & Danyel Fisher
Microsoft Research