



vis 2013
VAST • INFOVIS • SCIVIS
BIOVIS • LDAV

Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw

Carsten Görg, Zhicheng Liu, Jaeyeon Kihm, Jaegul Choo,



Haesun Park, and John Stasko



Text documents are everywhere ...

Academic Papers

Product Reviews

Health Forums

Police Reports

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 19, NO. X, XXXXXXXX 2013

Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw

Carsten Görg, Member, IEEE, Zhicheng Liu, Jaeyoon Kihm, Jaegul Choo, Student Member, IEEE, Haesun Park, Member, IEEE, and John Stasko, Senior Member, IEEE

Abstract—Investigators across many disciplines and organizations must sift through large collections of text documents to understand and piece together information. Whether they are fighting crime, curing diseases, deciding what car to buy, or researching a new field, inevitably investigators will encounter text documents. Taking a visual analytics approach, we integrate multiple text analysis algorithms with a suite of interactive visualizations to provide a flexible and powerful environment that allows analysts to explore collections of documents while sensemaking. Our particular focus is on the process of integrating automated analyses with interactive visualizations in a smooth and fluid manner. We illustrate this integration through two example scenarios: An academic researcher examining InfoVis and VAST conference papers and a consumer exploring car reviews while pondering a purchase decision. Finally, we provide lessons learned toward the design and implementation of visual analytics systems for document exploration and understanding.

Index Terms—Visual analytics, information visualization, sensemaking, exploratory search, information seeking, document analysis

1 INTRODUCTION

EVERYDAY, analysts and investigators confront large collections of data as they make decisions, solve problems, or simply seek to understand a situation better. Frequently, the data collections include text documents or documents with key text components. While numerical or structured data are more amenable to statistical and computational analysis, text data are conversely often messy and noisy, requiring a very sequential, slow processing (reading documents one-at-a-time, in order).

Investigators working with such document collections gather bits of information as they explore the data, hoping to form new insights about the issues at hand. Large, unstructured document collections make this task more difficult; the investigator may not know where to begin, what is important, or how concepts/events are related. The following situations are examples of these kinds of tasks:

- An academic researcher moves into a new area and seeks to understand the key ideas, topics, and trends of the area, as well as the set of top researchers, their interests, and collaborations.
- A consumer wants to buy a new car but encounters a large variety of possible models to choose from, each of which has 10 to 20 “professional” reviews and a web forum with hundreds of postings.
- A family learns that their child may have a rare disease and scours the web for documents and information about the condition, easily encountering many articles.
- A police investigator has a collection of hundreds of case reports, evidence reports, and interview transcripts and seeks to “put the pieces together” to identify the culprits behind a crime.

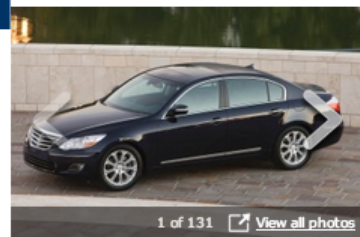
Such processes, sometimes called Sensemaking [39], [50], [54], Information Seeking Support [44], or Exploratory Search [43], [66], go beyond the initial retrieval of data or the simple return of the “right” document. Instead, they involve analysts browsing, exploring, investigating, discovering, and learning about the topics, themes, concepts, and entities within the documents, as well as understanding connections and relationships among the entities.

One approach to this problem is the computational analysis of document text, including text mining [3], [22]. However, as many researchers have noted [37], [58], simply performing computational analysis of the documents may not be sufficient for adequate understanding of a document collection—the investigator inevitably will think of some question or perspective about the documents that is either not addressed by the computational analysis or not represented accurately enough to draw a conclusion.

C. Görg is with the Computational Bioscience Program, University of Colorado, Mail Stop #303, 12861 E 17th Ave, Aurora, CO 80045. E-mail: Carsten.Gorg@colorado.edu.
 Z. Liu is with the Department of Computer Science, Stanford University, 379 Gates Hall, Stanford, CA 94305. E-mail: zlliu@cs.stanford.edu.
 J. Kihm is with Cornell CIS, 301 College Ave., Ithaca, NY 14850. E-mail: jk244@cornell.edu.
 J. Choo and H. Park are with the School of Computational Science and Engineering, College of Computing, Georgia Institute of Technology, 266 Ferst Drive, Atlanta, GA 30332. E-mail: {jchoo, hpark}@cc.gatech.edu.
 J. Stasko is with the School of Interactive Computing, Georgia Institute of Technology, 85 5th St., NW, Technology Square Research Building, Atlanta, GA 30332. E-mail: stasko@cc.gatech.edu.
 Manuscript received 8 May 2012; revised 3 Oct. 2012; accepted 5 Dec. 2012; published online 22 Dec. 2012.
 For information on obtaining reprints of this article, please send e-mail to: reprints@computer.org, and reference IEEECS Log Number TVCG-2012-05-0082. Digital Object Identifier no. 10.1109/TVCG.2012.324.

2010 Hyundai Genesis Sedan - Consumer Reviews

Overview Inventory Appraise Photos MPG More...



Average Consumer Rating: 4.5 stars (64 Total Reviews)

View reviews for a specific Genesis model: All Genesis models Sort by: Most Recent

3 of 3 people found this review helpful

Great value and excellent performance

By Jayz on 11/30/10 00:00 AM (PST)

Vehicle
2010 Hyundai Genesis 3.8 4dr Sedan (3.8L 6cyl 6A) 4.6 4dr Sedan (4.6L 8cyl 6A)

Review

After owning 4 LS series Lexus's I decided on a change due to the Lexus lackluster styling and high price. Toyota/Lexus was having all the recall problems at the time and upon reading the various automotive reviews I took a Genesis for a test drive. What a pleasant surprise. After 3000 miles I can say that the car handles beautifully in all driving conditions, has excellent acceleration and a virtually sound proof cabin. I would highly recommend this car to anyone interested in a high quality luxury sedan like ride at a very affordable price.

Favorite Features
Solid driving experience. Nicely designed.

Suggested Improvements

Driver's seat could be a little more comfortable, similar to the higher priced luxury sedans. Passenger seat should be fully powered so it can be raised or lowered.

Recommend Report it

Register now! For updates, forums and more! Sign in email or nickname Password

Text size: A A A Accessibility / Privacy Policy / Site Map



A safe and friendly online community where you can share experiences, get support and discuss autism

Home > Forums

- Home
- Forums
- Latest Posts
- Community guidelines
- How to use the site
- Online Q & A events
- News
- Support
- Members directory
- About us
- Contact us



Discussion Forums

Sign in to start talking!

Topic	Posts
Latest Posts	75
Introductions	117
Off Topic	318
General Autism	164
Diagnosis	176
Education	604
Adulthood	49

FORT WORTH POLICE DEPARTMENT WANTED PERSON REPORT

Distribution: 1. Dispatcher-Records, 2. Identification, 3. Detective

Name: Karen Lynn Bennett Address: Unknown Phone: Unknown

Color: White Sex: Female Age: 19 Birthdate: Alias: Nickname-Little Lynn

Ht. 5 in 1 Wt. 115 Hair: blond Eyes: blue Teeth: Comp: Pair

Speech or Voice: Scars, etc: None

Reason Wanted: MISSING Offense No: U-91

Ident. No:

Warrant issued: Yes No Warrant No:

Wanted By: Department:

Reported By: Date: Time:

Reported By: Mrs. Ruth Welobel How: Ph no Relationship: Grandmother

Probable Destination: Unknown Last Seen: 4228 W Vickery 1-10-64 Time: 8:30 PM

Mental Condition: Doubtful Cause of Absence: Unknown

Sex: Female Year: Make: Body Style:

Color: Year: Make: Body Style:

License: Year: State: Number: Condition:

Hat: Coat: black wool Trousers: Shoes: black heels

Tie: Gloves: black Sweater: Vest: Glasses:

Shirts: white Skirt: black cotton O'Coat: Luggage: Dress:

Socks: hose Condition of Clothing:

ADDITIONAL INFORMATION

Jewelry worn-writ watch-make unknown Large diamond ring on left hand Money carried-Unknown-Subject usually carries large amounts

Subject, Karen Lynn Bennett, came to her grandmother's house on the evening of 1-10-64, and left all of her clothing and property. The subject stayed with her grandmother for a few minutes and then left with a white male, Bruce Carlin, grandmother unknown. Subject told her grandmother that she would be back the next day to pick up her clothes and items. The subject hasn't returned this date. Mrs. Welobel stated that this is the Little Lynn, the strip-tease artist who worked for Jack Ruby in Dallas. She stated the girl may be somewhere in Dallas at this time. The meydana L-21 2-2-64 Time: 9:40 AM

Typed by: Location: Date: Time:

Located By: 36 30900-016 8/5

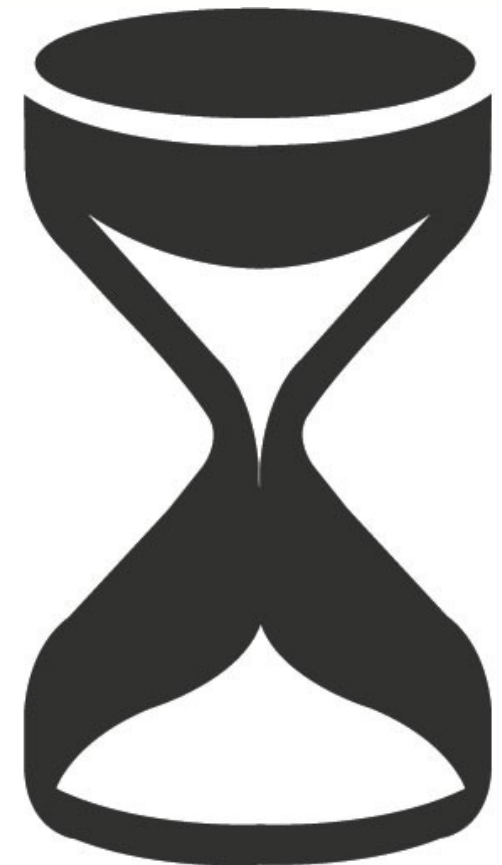
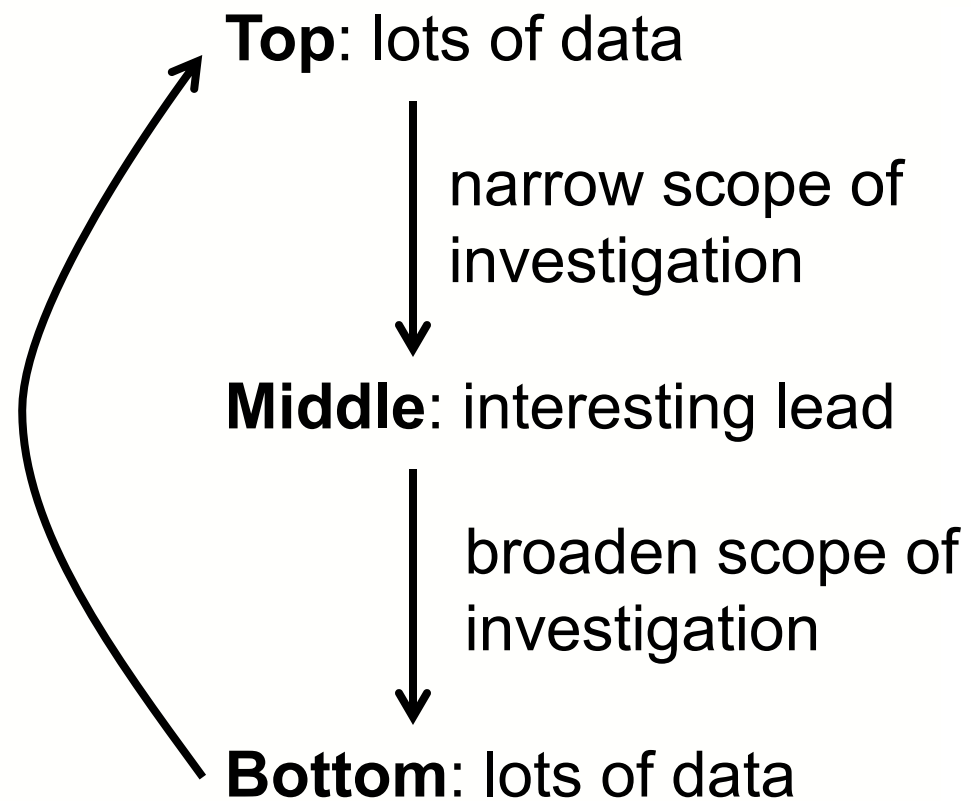
Investigative Analysis

- Assist analysts (police/intelligence, lawyers, reporters, scientists) confronting “information overload” problems while doing exploratory data analysis
- Given a collection of textual documents, assist analysts in
 1. Information foraging
 2. Sense-making

Two Types of Scenarios

1. Targeted analysis scenarios
2. Open ended, strategic analysis scenarios

Hourglass-shaped investigations



Text Analytic Approaches

Text Mining &
Computational Analysis

Interactive Text Visualization

contribution
of this paper

Integration

Visual Text Analytics

Jigsaw

Jigsaw
Infovis & VAST Papers '95-'11
578 documents

Color Legend:

- author (1139)
- concept (78)
- conference (2)
- indexterm (1915)
- journal (18)
- keyword (1197)
- year (17)

Search

Calendar View

Edit View Bookmarks Show A

Use Doc Date
 Use Date Entities

Begin: 2008
End: 2011

Jan Feb Mar Apr May Jun Jul Aug Sep Oct

model (0)
overall_rating (0)
performance_rating (0)
reliability_rating (0)
 Document (231)
123456701
123456702

List View

Edit View Bookmarks Lists Options Export

author: Keim, D.A., Oelke, D., Schneidewind, J., Dayal, U., Hao, M.C., Mansmann, F., North, S., Panse, C., Sips, M., Bak, P., Janetzko, H., Rohrdantz, C., Schreck, T., Stoffel, A., Albuquerque, G., Ankerst, M., Berchtold, S., Danon, G., Deussen, O., Eisemann, M., Gruse, T., Haug, L.-E., Heilmann, R., Hsu, M., Jenny, M.

concept: insight, text, pixel, distortion, document, geographic, hierarchy, interaction, parallel coordinates, case study, clustering, color, evaluation, network, time series, treemap, animation, business, cluster, financial, geospatial, high-dimensional da..., overview, radial, security

Document Cluster View

Edit View Bookmarks Export Options

Highlight Viewed Documents

Filters: All Filters

Group by Filters
Undo Filters
Hide Unfiltered

Clusters: Text Seed (20)

3d, displays, navigation

analytics, anomalies, detect

history, mining, patterns

querying, series, temporal

state, displayed, explored

collaborative, uses, framework

video, explorer, stories

text, features, topic

interact, understand, cognition

diverse, environments, toolkit

spaces, internet, search

network, graph, social

graphs, edge, algorithm

classification, geographic, statistics

transform, quality, studied

insight, genes, expression

animation, trends, causality

tables, database, interfaces

treemaps, coloring, hierarchically

dimensions, coordinates, parallel

Word Tree View

Edit Bookmarks

ride

176 matches

and: comfort in driving, listening to Lexicon, and the way the car is put together. Hyundai has a hit for my standards is just perfect. handling prowess. Suggested Improvements: The badge will not sell this car unfortunately. Power folding outside mirrors, Full power passenger seat with memory, chrome wheel option, illuminated garage, the power, the mileage on regular gas, and the luxury looks inside and out. acceleration, and interior size (the trunk is tremendous and the back seats are very comfortable) Suggested Improvements as some people say is on the stiffer end auto turn signal, blind spot sensor, available 19" wheels. comfortable and quiet, acceleration is superb for a big car, and the price is right. rough and jittery with a tendency towards porpoising on almost all surfaces, even those that appear to be smooth a little bumpy, suspension needs work. is: smooth and quiet, power is plentiful, comfort outstanding, all kinds of bells and whistles, done very tastefully. smoother and quieter, the quietness being my first impression when I test drove the car. amazing. busy. excellent, handles extreme, y well in tight spots. and: handling of this car. quietness are excellent. quality: is bad. craftsmanship, ergonomics, handling, and the premium leather seats. BMW's and Audi's. of the: E Class MB (realistically the benchmark of "smooth") I dumped but real close on the open road--definitely not kind of similar to my father's 2001 BMW 528i, but more smooth overall. comfort of the Genesis? could be improved - there has to be a better compromise on maintaining sportiness and handling with a bit more ride comfort in luxury. Suggested Improvements: Maybe fuel economy but you can not expect 4-cylinder fuel economy from this car. but the sound system was not as good as Genesis on this car. remains a nice balance between soft and hard. (smooth), features (lots).

Showing 35 / 135 (25.926%) leaves in current branch, of 135 leaves in tree. Prune Percentage: 25.926 Prune to Window

Document Grid View

Edit View Bookmarks Export

Add All Entity (10) Freq Words Unique Words Sort by Sentiment Color by Sentiment

178461953 : -0.26 Document to compare similarity to Compute Similarity based on Text

- controls, need, works
- package, drivers, passenger
- look, nice, dealership
- buy, value, 6cyl
- tire, fuel, economy
- comfortable, stereo, navigation
- drive, comfort, buy
- improvement, rear, trunk
- loving, infiniti, looks
- passenger, package, gas

Favorite Features: Nothing Suggested Improvements: Test the car for more that 10 miles before trying to sell it.

Document View

Edit View Bookmarks Export

Only Entities

Chrysler design drive driver drove favorite features Genesis Hyundai improvements interior luxury miles people price problem rides rough seat seats smooth steering suggested suspension trunk

Documents: 112521695, 144901743, 154441885, 154442215, 155461819, 156541729, 160201695, 164581699, 178461953, 179341707, 181041695, 182101705, 182601721, 184482141, 184482337, 184561861, 184901719, 185021703

Summary: It is a gorgeous car, but after I drove it 850 miles it reminds me of the gorgeous, brainy woman who can't boil water.

Source: Rides too rough..

Date: Sep 4, 2009

I bought my pre-owned 3.8 from a man who is known to be less than frugal who really "unloaded" this beauty upon me. It is a gorgeous car, but after I drove it 850 miles it reminds me of the gorgeous, brainy woman who can't boil water. Only the BMW 318i that I just sold rode worse. The suspension system is horrible. When the tires are cold you can feel every tar strip in the highway. The guys who designed the suspension should not be allowed to graduate fro high school this year. The steering is also numb and it becomes very tiring to be driving with the intent to find a good centerline feel and have that suspension fighting me. There is too much road noise being transmitted into the car.

Favorite Features: The design statement, interior feels great.

Suggested Improvements: Redesign the whole suspension system. It rides about as bad as the BMW. The cup holders are too few in number and poorly placed. It could use a Bluetooth that will download my entire address book. The one in there is a bit low tech.

Affiliated entities: build_quality_rating: 10, comfort_rating: 6, exterior_design_rating: 10, features: bluetooth noise steering suspension tires, fuel_economy_rating: 5, fun_to_drive_rating: 4, interior_design_rating: 10, make: BMW, overall_rating: 7.1, performance_rating: 5, reliability_rating: 7

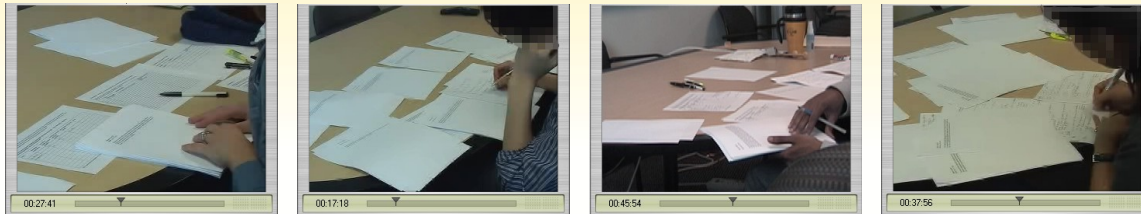


Stasko, Görg, and Liu. Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, 2008.

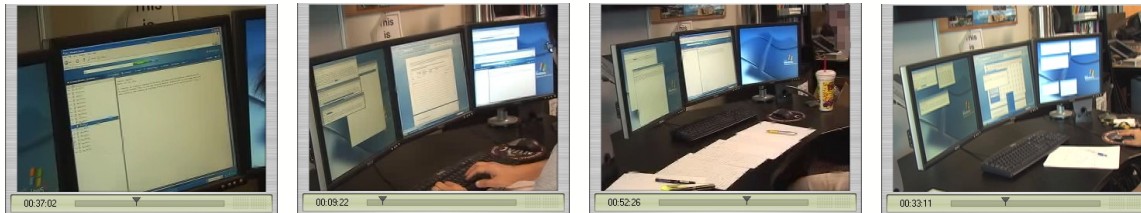
Evaluation Study

4 Settings

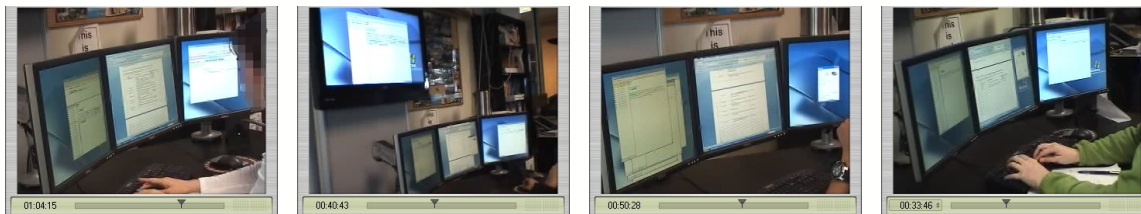
Paper



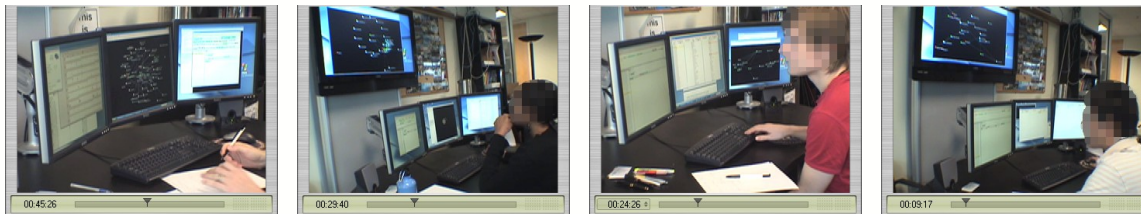
Desktop



Entity



Jigsaw



4 Strategies

Summarization

Doc Metrics

1. Overview, filter & detail

Doc Similarity

Recommend related docs

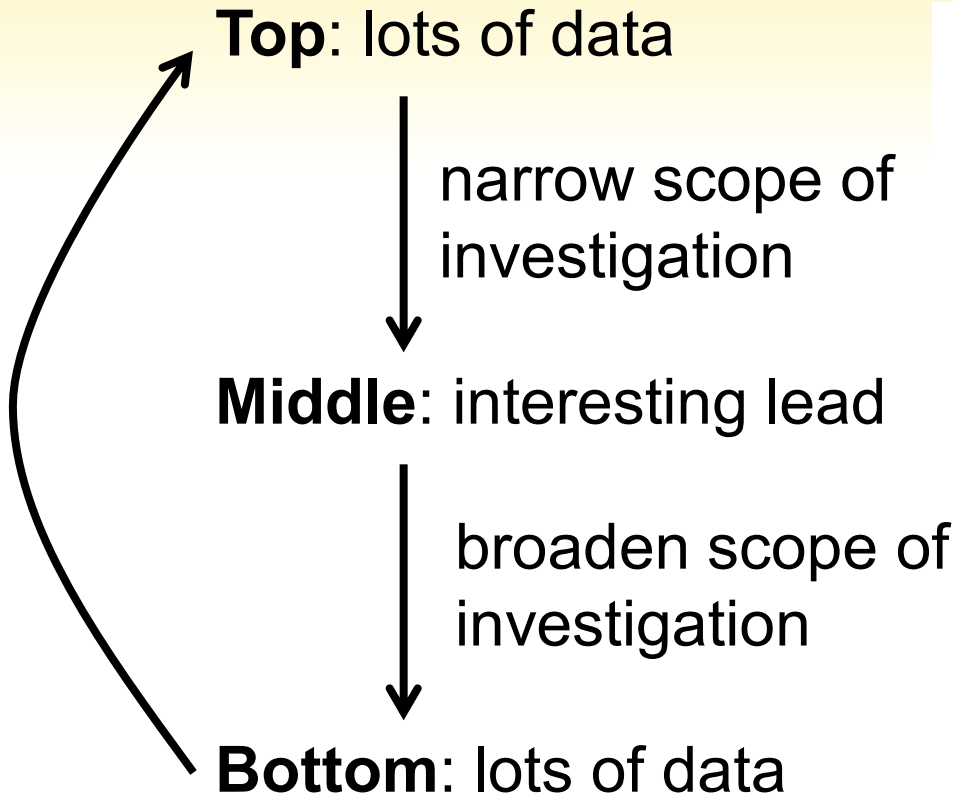
2. Build from detail

3. Hit the keyword

Doc Clustering

4. Find a clue and follow the trail

Computational support for hourglass-shaped investigation



Document clustering
Sentiment analysis

Document similarity
Recommending related documents

Computational Text Analyses

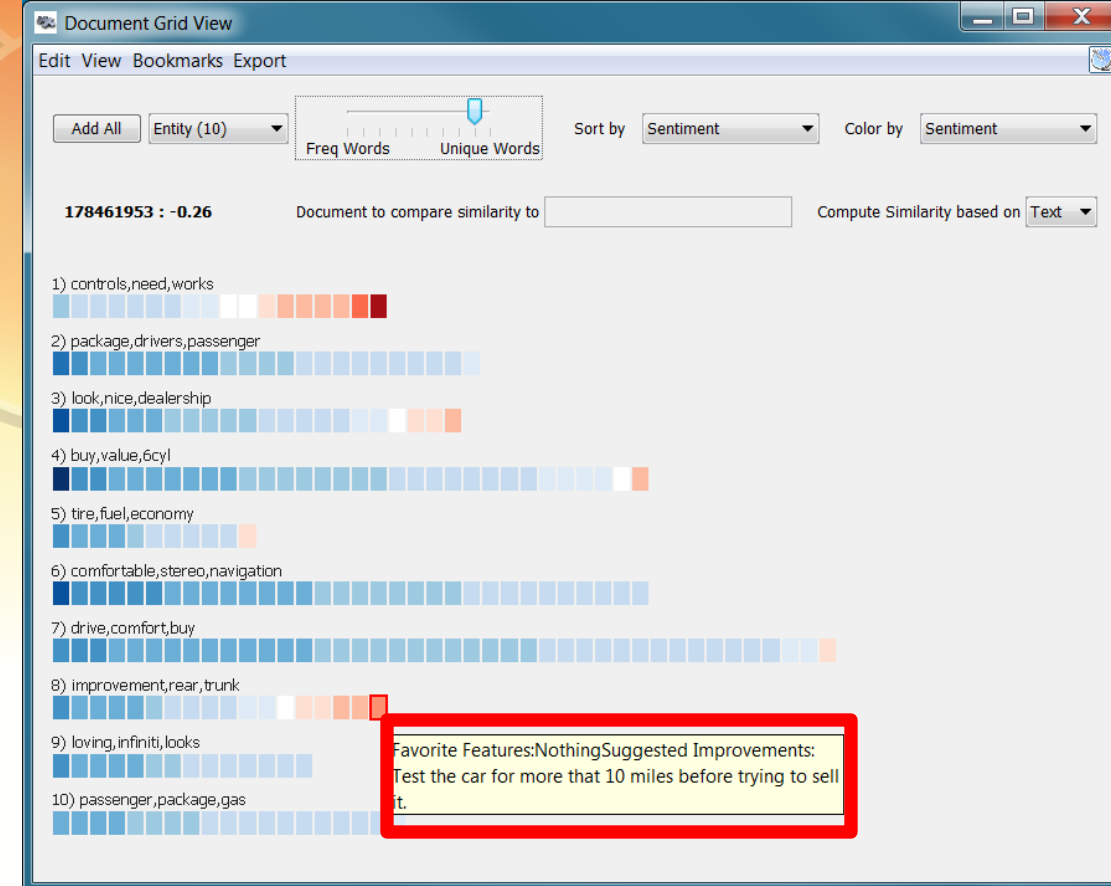
- Document Summarization
(one sentence summaries, word clouds, and keyword summaries)
- Document Similarity
(based on text or entities, cosine similarity)
- Document Clustering
(based on text or entities, k-means clustering)
- Sentiment Analysis
(hierarchical classification using Lingpipe and dictionary-based approach)

Document Summarization

word clouds
(multiple docs)

one sentence
summaries
(single doc)

keyword
summaries
(multiple docs)



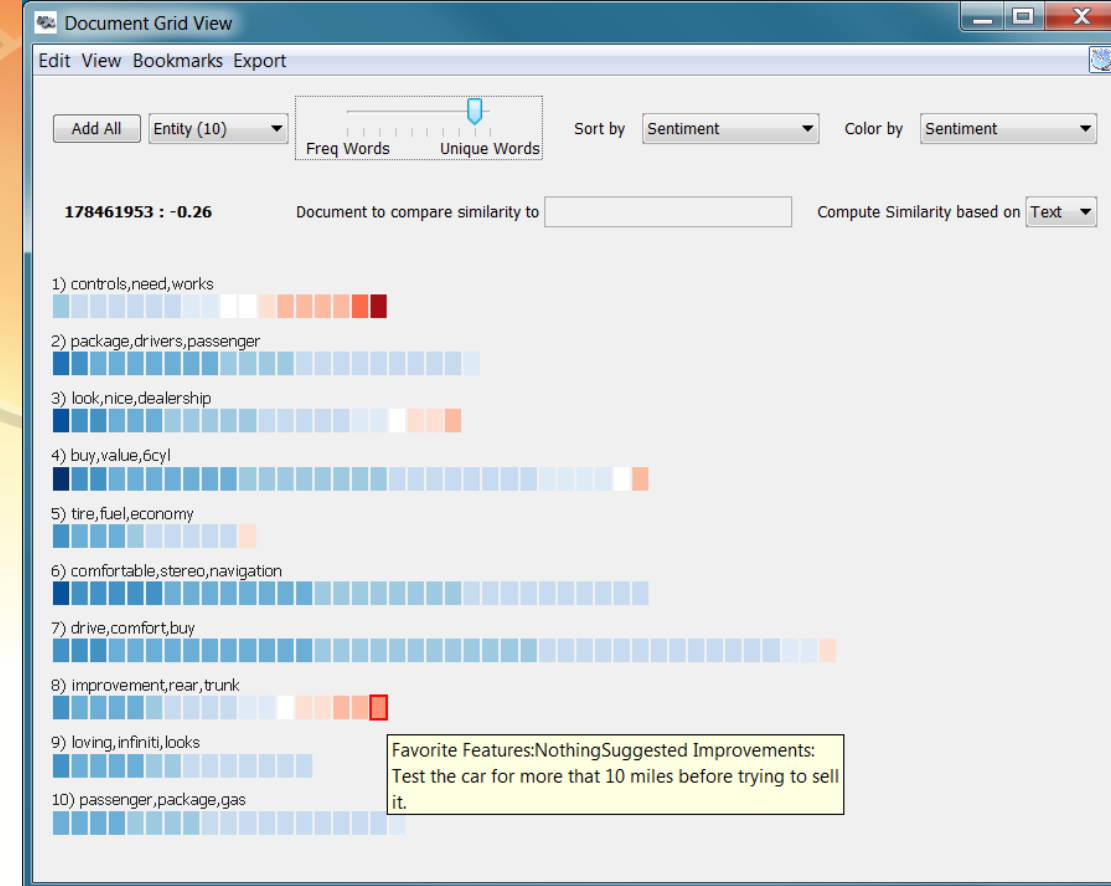
Document View interface showing a document summary and metadata. The interface includes a menu bar (Edit, View, Bookmarks, Export) and a toolbar with a 'Only Entities' checkbox. The main area displays a document titled 'network visualization' with a summary and metadata. A red box highlights the summary text: 'Summary: Powered by an underlying relational algebraic framework, Ploceus supports flexible construction and transformation of networks through a direct manipulation interface, and integrates dynamic network manipulation with visual exploration for a seamless analytic experience.' The metadata includes: 'Source: Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on', 'Date: Oct 26, 2011', 'Network-based visual analysis of tabular data.', and 'Tabular data are pervasive. Although tables often describe multivariate data without explicit network semantics, it may be advantageous to explore the data modeled as a graph or network for analysis. Even when a given table design conveys some static network semantics, analysts may want to look at multiple networks from different perspectives, at different levels of abstraction, and with different edge semantics. We present a system called Ploceus that offers a general approach for performing multi-dimensional and multi-level network-based visual analysis on multivariate tabular data. Powered by an underlying relational algebraic framework, Ploceus supports flexible construction and transformation of networks through a direct manipulation interface, and integrates dynamic network manipulation with visual exploration for a seamless analytic experience.' The 'Affiliated entities' section lists: 'author: Liu, Z. Navathe, S.B. Stasko, J.', 'concept: graph network', 'conference: VAST', 'indexterm: aggregates cities and towns data visualization organizations relational databases semantics visualization', 'journal: Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on', and 'year: 2011'. A 'Documents' list on the left shows several document IDs, with 'vast11-6102440' selected. A red box highlights the 'Add all' button at the bottom left.

Document Summarization

word clouds
(multiple docs)

one sentence
summaries
(single doc)

keyword
summaries
(multiple docs)



network

networks

visualization

Summary: Powered by an underlying relational algebraic framework, Ploceus supports flexible construction and transformation of networks through a direct manipulation interface, and integrates dynamic network manipulation with visual exploration for a seamless analytic experience.

Source: Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on

Date: Oct 26, 2011

Network-based visual analysis of tabular data.

Tabular data are pervasive. Although tables often describe multivariate data without explicit network semantics, it may be advantageous to explore the data modeled as a graph or network for analysis. Even when a given table design conveys some static network semantics, analysts may want to look at multiple networks from different perspectives, at different levels of abstraction, and with different edge semantics. We present a system called Ploceus that offers a general approach for performing multi-dimensional and multi-level network-based visual analysis on multivariate tabular data. Powered by an underlying relational algebraic framework, Ploceus supports flexible construction and transformation of networks through a direct manipulation interface, and integrates dynamic network manipulation with visual exploration for a seamless analytic experience.

Affiliated entities:

author: Liu, Z. Navathe, S.B. Stasko, J.

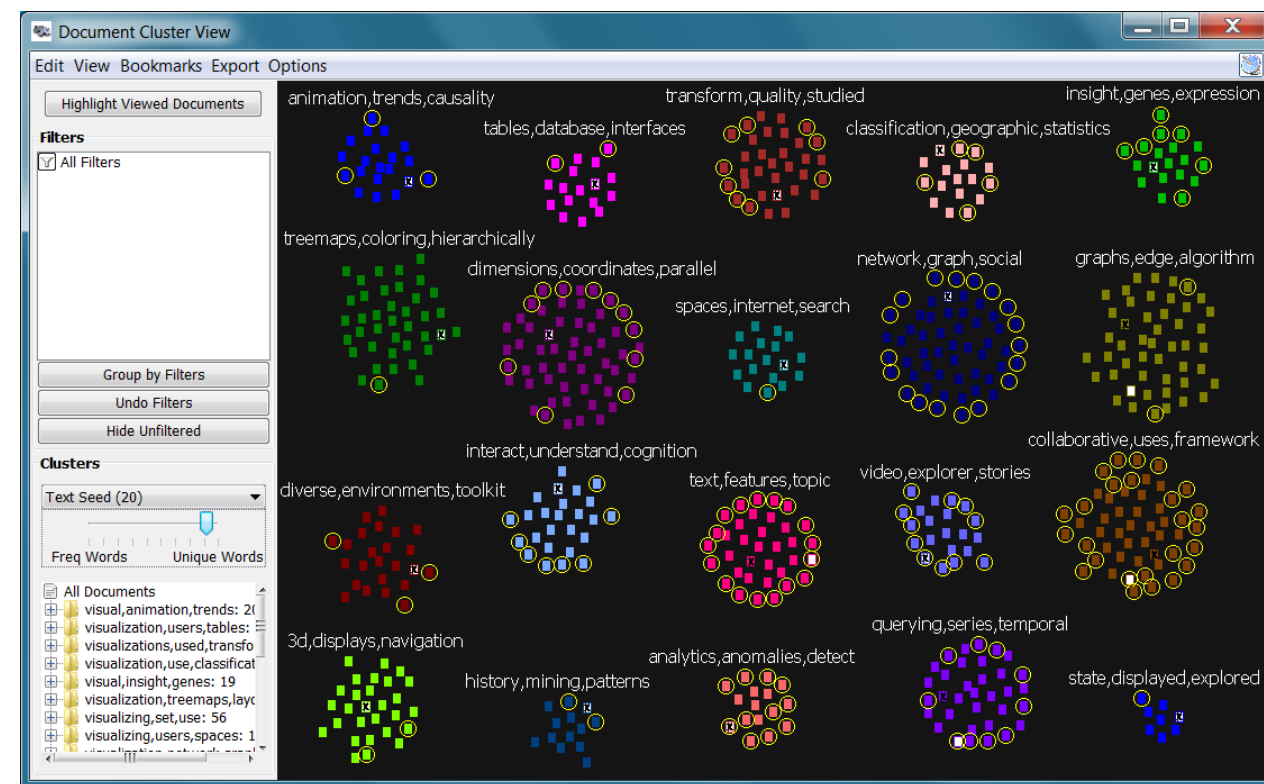
concept: graph network

conference: VAST

index term: aggregates cities and towns data visualization organizations relational databases semantics visualization

journal: Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on

year: 2011

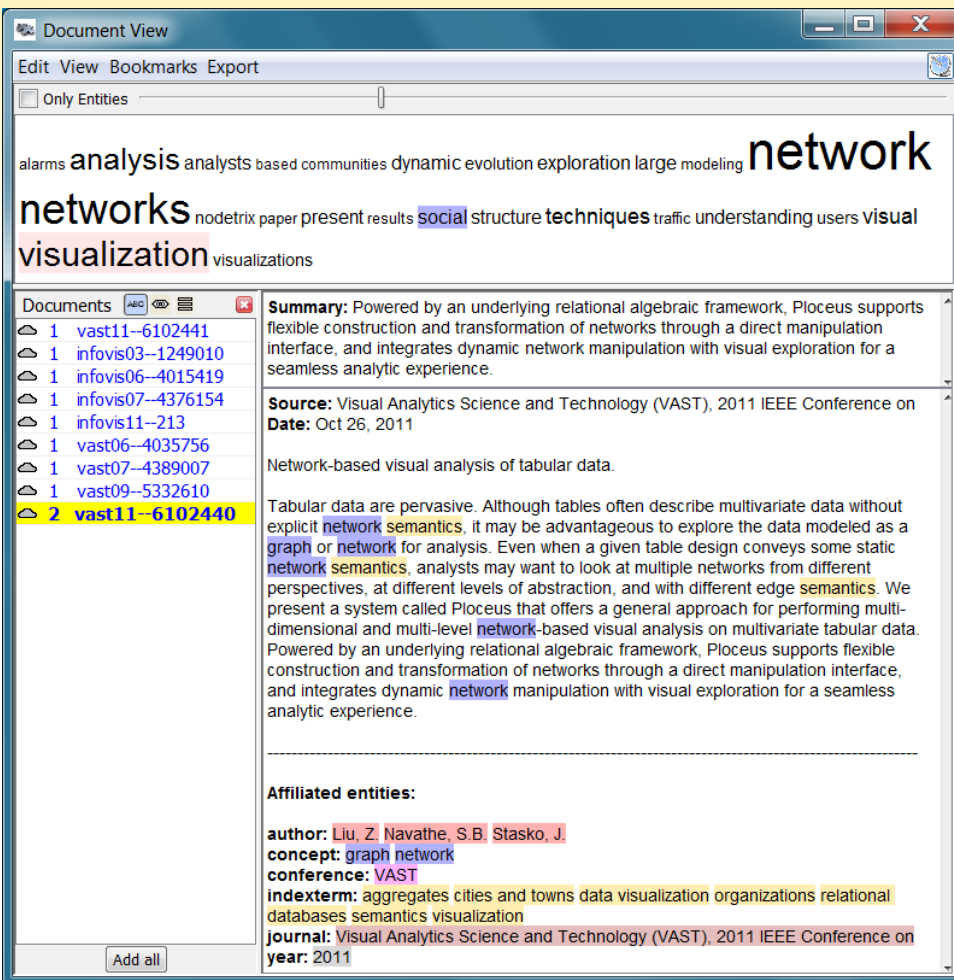
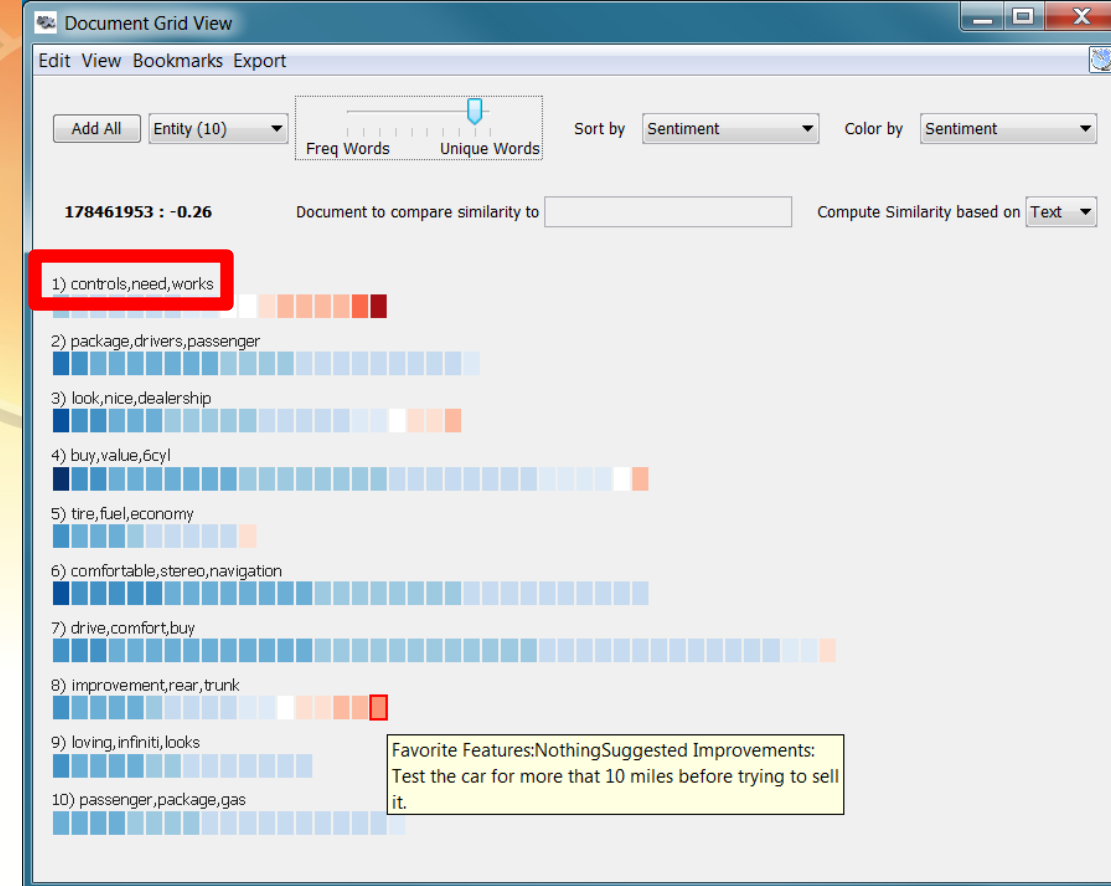


Document Summarization

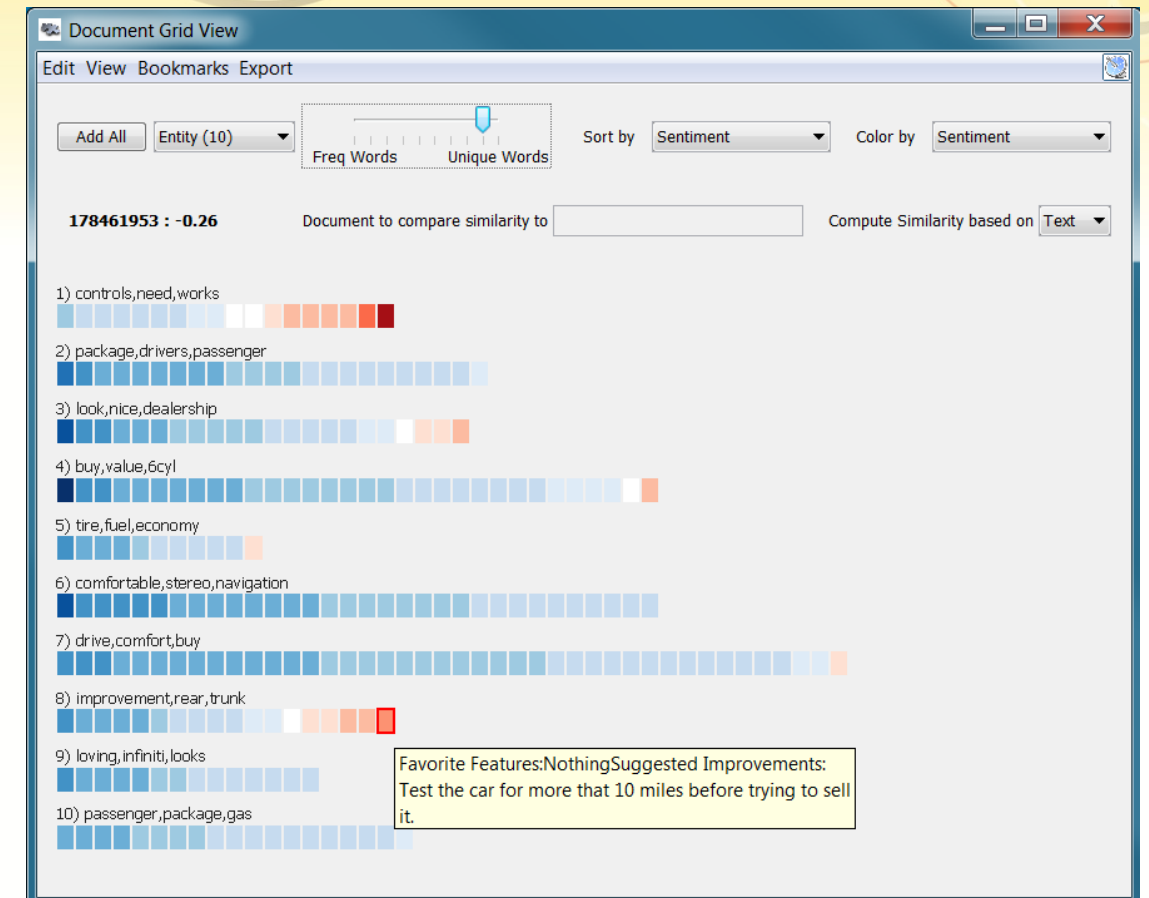
word clouds
(multiple docs)

one sentence
summaries
(single doc)

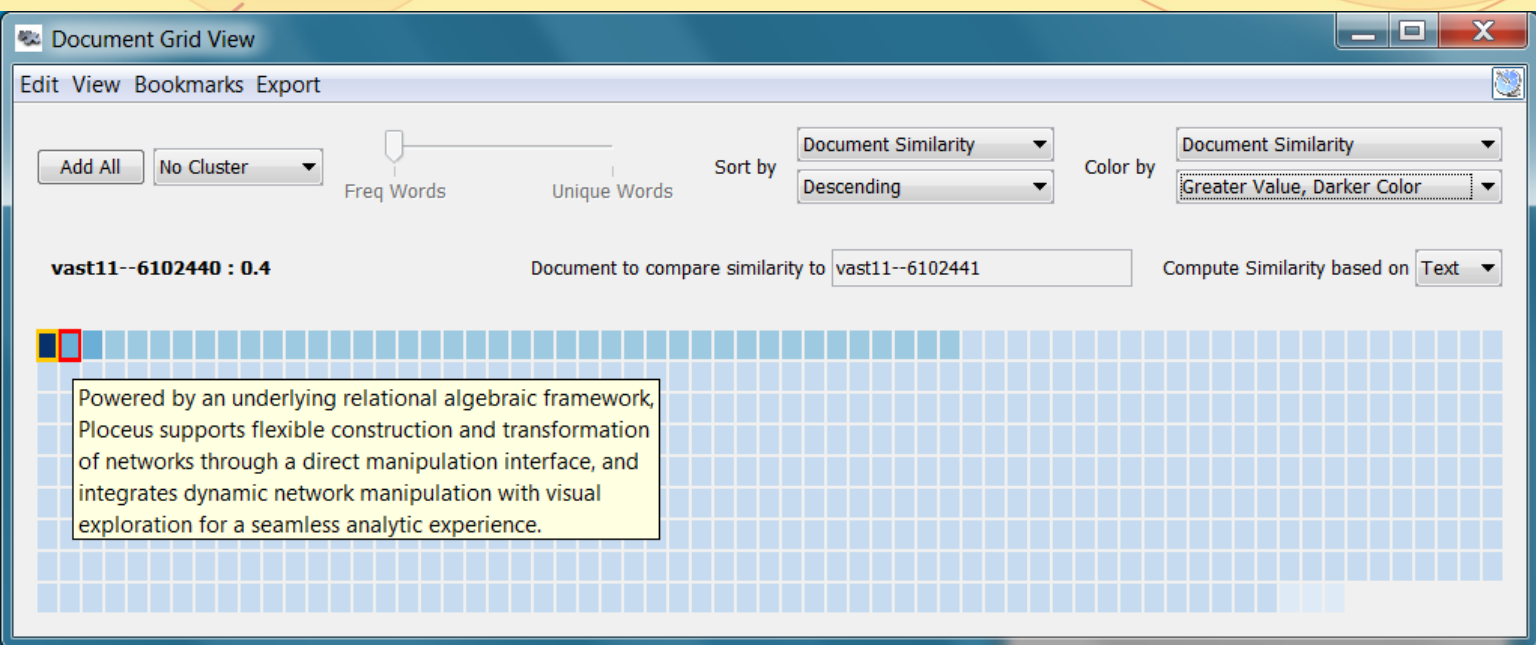
keyword
summaries
(multiple docs)



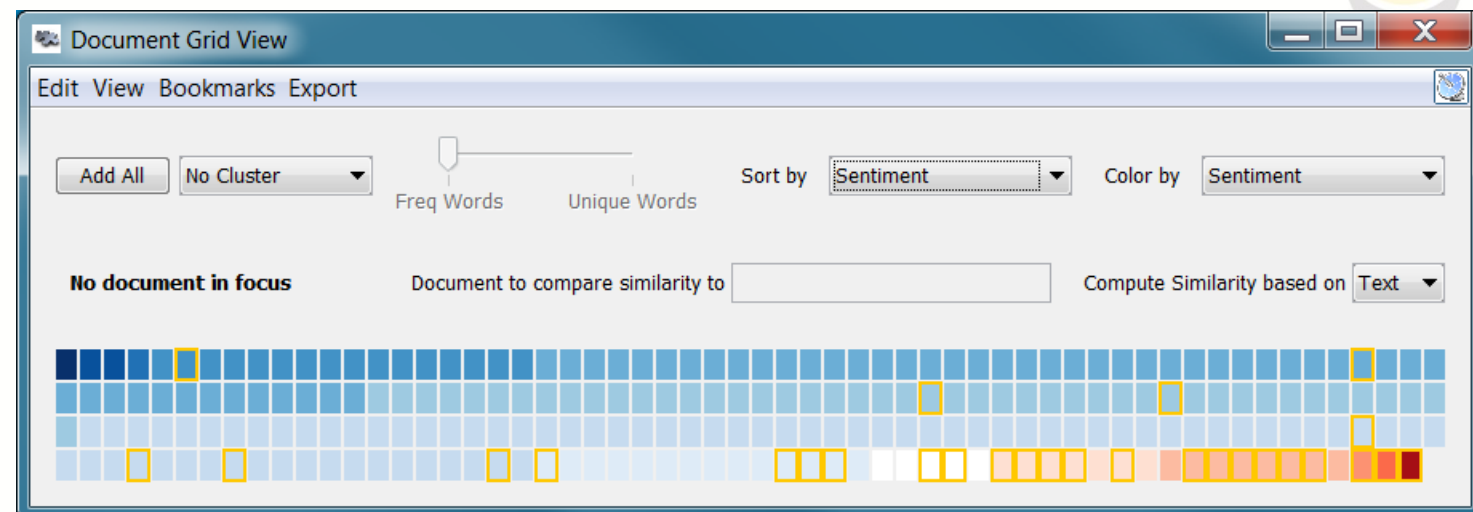
Document Clustering



Document Similarity & Document Metrics

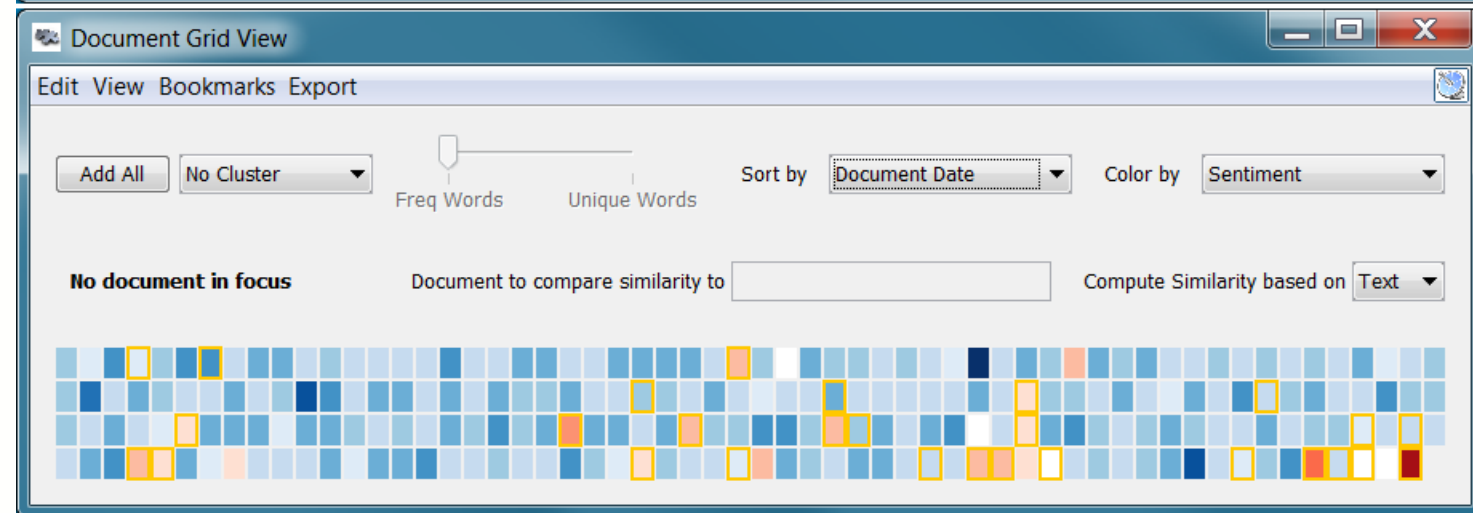


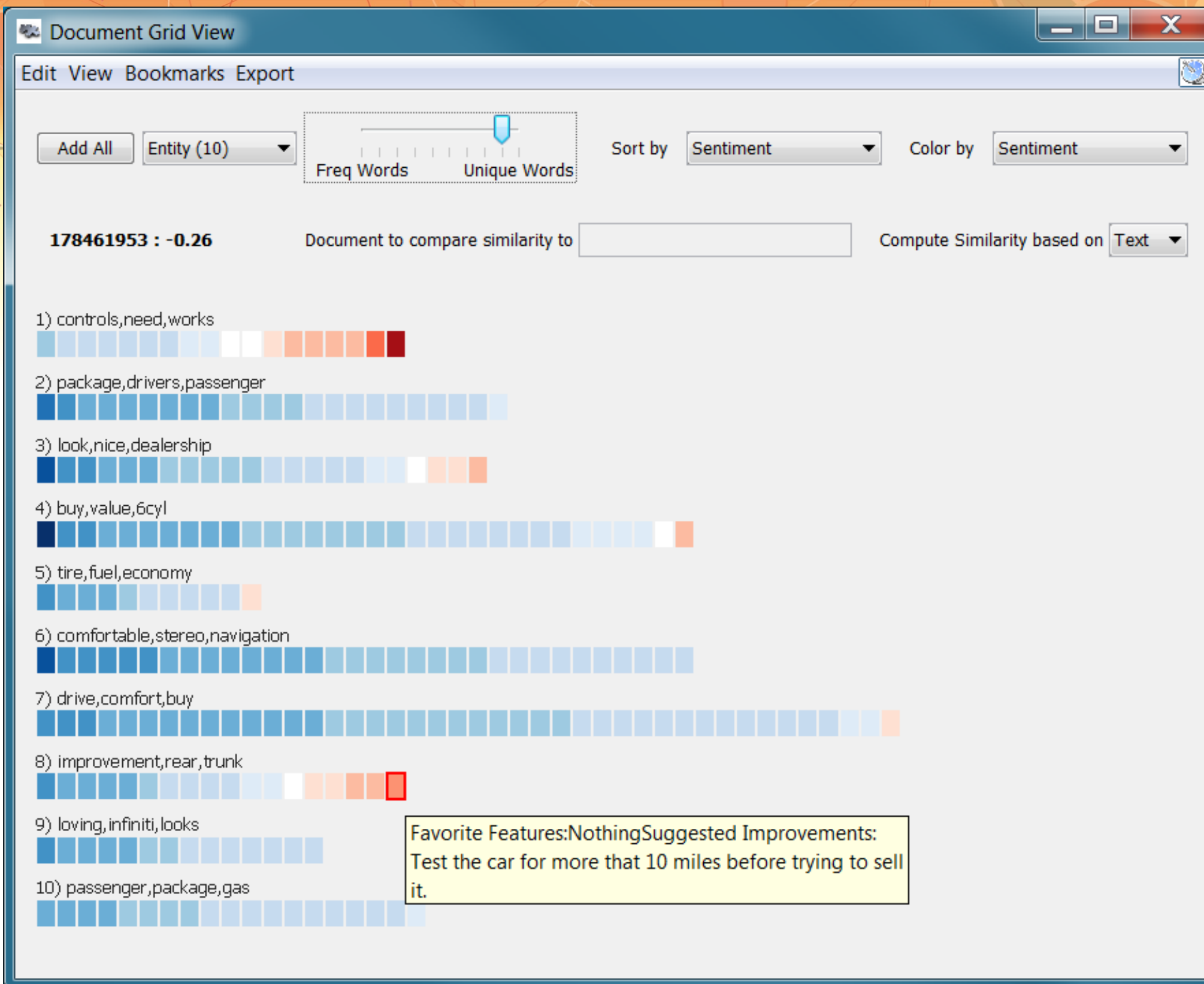
Documents ordered and colored by sentiment



Documents ordered and colored by similarity to selected reference document

Documents ordered by date and colored by sentiment





Videos available
in IEEE DL

Lessons learned

1. Make different computational analysis results available throughout the system in a variety of different contexts and views
2. Flexibly allow analysis output also to be used as input
3. Integrate different, independent computational analysis measures through interactive visualization to extend functionality and power
4. Provide computational support for both analysis directions: narrowing down as well as widening the scope of an investigation
5. Expose algorithm parameters in an interactive user accessible way

Contributions

1. Methods for fluidly integrating computational text analysis and visualization approaches
2. Illustrating of the utility through example usage scenarios
3. Lessons learned toward the design and construction of visual analytics systems for document exploration

Acknowledgments



IIS-091788
CCF-0808863



2009-ST-061-CI0001

Get your own copy (incl. InfoVis/VAST paper dataset)

www.jigsaw-analytics.net

Jigsaw Download Page

Visual Analytics for Exploring and Understanding Document Collections

Search

Home About Version History

Download

For more information on the system, please see the [Jigsaw Home Page](#).

Please register to download Jigsaw. Your information will not be disclosed to any third parties. We will only use your contact data for updates on new versions of the system and for potential questions about system utility and feedback.

Current version: **0.52** (created August 2013)

Name*

Affiliation*

Email*

How did you hear about Jigsaw?*

I agree with the license*

Notes:

1. If you have a slow Internet connection and cannot unzip Jigsaw after downloading it, please check the downloaded file size (should be around 130 MB) and try again. Sometimes the download times out on slow connections.
2. JavaScript has to be enabled for the download. [Here are instructions](#) on how to enable JavaScript in your web browser.

Problems? Questions? Feedback? Send email to info@jigsaw-analytics.net



Plug-in architecture for your visualizations and algorithms