

VISUAL TEXT ANALYTICS IN JIGSAW

Nadia Boukhelifa

nb@lri.fr

VISUAL ANALYTICS

8 OCT 2014



BEFORE WE START

1. Import a dataset

InfovisVast-papers.jig

2. Run Computational Analysis

Tools -> compute all *5 min*

OVERVIEW

Introduction to Jigsaw

Visual Text Analytics using Jigsaw

UBIQUITOUS IMAGE



FACEBOOK

BY THE NUMBERS

AN OVERVIEW



Facebook has more than **800 million active users**.
More than **50%** of active users log into Facebook **every day**.



There are more than **2 billion posts** liked and commented on every day and **250 million photos** uploaded every day.



75% of Facebook users are **outside the U.S.**
It is available in **70 languages**.

UBIQUITOUS IMAGE



FACEBOOK

BY THE NUMBERS

AN OVERVIEW

+50%



Facebook has more than **800 million active users**.
More than **50%** of active users log into Facebook **every day**.



There are more than **2 billion posts** liked and commented on every day and **250 million photos** uploaded every day.



75% of Facebook users are **outside the U.S.**
It is available in **70 languages**.

UBIQUITOUS IMAGE



FACEBOOK

BY THE NUMBERS

AN OVERVIEW



Facebook has more than **800 million active users**.
More than **50%** of active users log into Facebook **every day**.



There are more than **2 billion posts** liked and commented on every day and **250 million photos** uploaded every day.

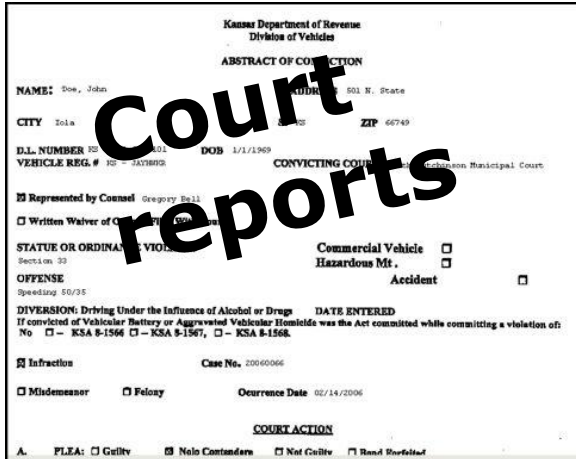


75% of Facebook users are **outside the U.S.**
It is available in **70 languages**.

TEXT EVERYWHERE



News reports



Court reports

Information Week (2008) 7, 118-12

gigsaw: supporting investigative analysis through interactive visualization

Abstract
Investigative analysts who work with collections of text documents connect embedded text and information. Scientists follow this process when they read research papers to learn about related efforts; news reporters perform such analyses when they investigate new stories; law enforcement and intelligence analysts carry out these kinds of investigations when they review case reports. Our common element of all these analytic activities is that they are cognitively very challenging, frequently involving large collections of data and text that tax a person's memory, deduction, reasoning, and general analytic capabilities.

Introduction
Investigative analysts seek to make discoveries and uncover hidden truths from large collections of data and information. Scientists follow this process when they read research papers to learn about related efforts; news reporters perform such analyses when they investigate new stories; law enforcement and intelligence analysts carry out these kinds of investigations when they review case reports. Our common element of all these analytic activities is that they are cognitively very challenging, frequently involving large collections of data and text that tax a person's memory, deduction, reasoning, and general analytic capabilities.

While reading documents and digesting the information therein, analysts gradually form internal mental models of the people, places, and events discussed in the documents. As the number of documents grows larger, however, it becomes increasingly difficult for an investigator to track the connections between data and make sense of it all. The sheer number of entities involved may make it very difficult for a person to learn a

Academic articles

phpBB yourdomain.com

Board index

User Control Panel (0 new messages) | View your posts

It is currently Wed Oct 21, 2009 12:15 pm

View unanswered posts | View unread posts | View topics with new posts

YOUR FIRST POSTION	TOPIC	POSTS	LAST POST
Youngest fo	1	1	by Meow 1 Wed Oct 21, 2009 12:09 pm

Registered users: Meow
Legend: Administrators, Global moderators

STATISTICS
Total posts 1 • Total topics 1 • Total members 1 • Our newest member Meow

Board index

Online forums

Buzzillions Unbiased product reviews from actual buyers.

Home Electronics Sports Home & Garden Shoes Clothing Health & Beauty Baby Toys Pets Office Clean See All

3,019,512 Customer Reviews

Home Electronics Cameras & Photo Digital Cameras Canon Point & Shoot Camera Reviews

Canon G9 Powershot Digital Camera Reviews

Write a Review Save Print

Review Snapshot Description Specs Q&A

4.6

vs. 4.6 category average Based on 172 reviews

Most common customer feedback:

PROS
Bright LCD (134) Quality constructio (118) In low light (7) Art (3) Photography (2)

CONS
None (0) None (0) None (0) None (0)

WHERE TO BUY
\$439.99 - \$549.00 from 7 Stores

Amazon.com \$500.11

Best Buy \$539.95

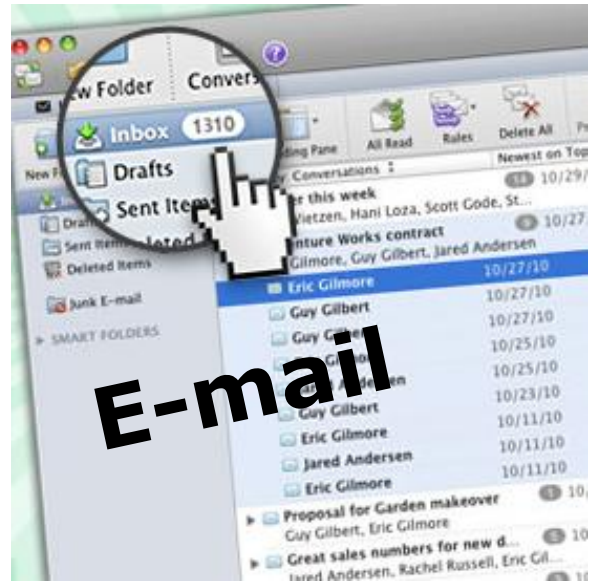
Out of Stock at Best Buy \$599.00

Similar Products
Canon PowerShot SD170 IS Digital Elph Digital Camera

Showing 1 - 2 of 182 Reviews

Show all reviews by people who are:
All Reviewers Casual User Semi-Enthusiast Professional

Product reviews



E-mail

TEXT SEARCH



TEXT SEARCH



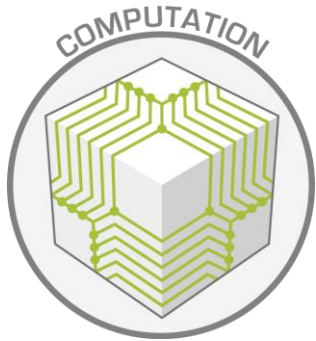
VISUAL TEXT ANALYTICS

- Assist analysts to overcome **information-overload**
 - Information foraging
 - Sense-making



VISUAL TEXT ANALYTICS

- Text mining & Computational Analysis
- Interactive text visualization



enterprise technology infrastructure operations
information systems cards objectives
analysis text mining capitaliz
metrics applications manage
connections techniques bank
solution stakeholder

+



CHALLENGES



- How to *extract & visualize* concepts and relations from large noisy corpora
- How to design new visualization / interaction techniques to *browse* and *refine* content without becoming data mining expert

Jigsaw

The screenshot displays the Jigsaw software interface with several panels:

- Jigsaw Main Panel:** Shows 'Infovis & VAST Papers '95-'11' with 578 documents. A color legend identifies entities like 'author' (1139), 'concept' (78), 'conference' (2), 'journal' (18), 'year' (17), 'document' (231), and 'keyword' (1197).
- Calendar View:** A calendar interface for filtering documents by date, with 'Use Doc Date' and 'Use Date Entities' checked.
- List View:** A table of documents with columns for 'year', 'author', and 'concept'. The author 'Keim, D.A.' is highlighted, with lines connecting to the Document Cluster View.
- Document Cluster View:** A network graph showing clusters of documents based on concepts. Clusters include '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', and 'radial security'.
- Document Grid View:** A grid of documents with a similarity score of 0.26. It lists 10 documents with their titles and a bar chart representing their similarity to the selected document.
- Document View:** A detailed view of a document titled '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'. It includes a summary, source information, and a list of features and ratings.
- Word Tree View:** A hierarchical tree diagram showing the relationships between words in the document, with 'ride' as a central node.



Stako, Görg, and Liu. Jigsaw: Supporting investigative analysis through interactive visualization.

Jigsaw: Supporting Investigative Analysis through Interactive Visualization

John Stasko, Carsten Görg,
Zhicheng Liu, Kanupriya Singhal

School of Interactive Computing & GVI Center
Georgia Institute of Technology

NAMED ENTITIES

Entities are connected if they appear in the same document



CORDINATED MULTIPLE VIEWS

Jigsaw

List View

year	author	concept
1995	Keim, D.A.	insight
1996	Oelke, D.	text
1997	Schneidewind, J.	pixel
1998	Dayal, U.	distortion
1999	Dao, M.C.	document
2000	Mansmann, F.	geographic
2001	North, S.	hierarchy
2002	Panse, C.	interaction
2003	Sips, M.	parallel coordinates
2004	Bak, P.	case study
2005	Janetzko, H.	document
2006	Rohrdantz, C.	color
2007	Schreck, T.	evaluation
2008	Stoffel, A.	network
2009	Anbuquerre, G.	time series
2010	Ankerst, M.	tree map
2011	Berchtold, S.	animation
	Danon, G.	business
	Deussen, O.	cluster
	Eisemann, M.	financial
	Haug, L.-E.	geospatial
	Heilmann, R.	high-dimensional da...
	Hsu, M.	overview
	Jenny, M.	radial
		security

Document Cluster View

animation,trends,causality transform,quality,studied insight,genes,expression
tables,database,interfaces classification,geographic,statistics
freemaps,coloring,hierarchically dimensions,coordinates,parallel network_graph,social graphs,edge,algorithm
spaces,internet,search
interact,understand,cognition collaborative,uses,framework
diverse,environments,bookit text,features,topic videoexplorer,stories
3d,displays,navigation history,mining,patterns analytics,anomalies,detect querying,series,temporal
state,displayed,explored

Document Grid View

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

179641953 - 0.26 Document to compare similarity to Compute Similarity based on Text

1) controls,need,works
2) package,drivers,passenger
3) look,nice,dealership
4) buy,vakab,5cyl
5) tire,fuel,economy
6) comfortable,sterno,navigation
7) drive,comfort,boy
8) improvement,rear,trunk
9) loving,infinity,books
10) passenger,package,gas

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

Document View

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

Summary: It is a gorgeous car, but after I drove it 850 miles it reminds me of the gorgeous, brassy woman who can't boil water. The BMW 518i 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 from high school this year. The steering is also numb and it becomes very tiring to be driving with the intent to find a good centrefire feed and have that suspensions 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: 10; exterior, design, rating: 10; features: bluetooth, noise, steering, suspension, tires; fuel, economy, rating: 5; fun, to, drive, rating: 5; interior, design, rating: 10; make: BMW; overall, rating: 10; performance, rating: 6; reliability, rating: 7

Calendar View

Begin: 2008, End: 2011

Word Tree View

Keim

176 matches

- and - comfort in driving, listening to Lexicon, and the way the car is put together
- and - 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 door, 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: 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
- 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
- excitement, handles extremely well in tight spots.
- and handling - of this car
- and - quietness are excellent
- quality - is bad
- craftsmanship, ergonomics, handling, and the premium leather seats
- of the - BMW's and Audi's
- 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

STRATEGIES

- Overview, filter & detail



Summarisation, doc metrics

- Build from detail

Doc similarity, recommended related docs

- Hit the keyword

Keyword search

- Find a clue and follow the trail

Document clustering

DEMO

Exploring academic publications: Infovis and VAST 1995-2013

ASSIGNMENT

1. Import a dataset

articles_enriched.jig

2. Extract entities 3 minutes

entities -> identify entities -> Illinois-NER

entities -> clean up entities

3. Run Computational Analysis 10 minutes

Tools -> compute all

ASSIGNMENT

- Don't forget to save your work!
- Send me PDF file with answers
- Deadline: Mon13th Octobre 2014, 23:00

ACKNOWLEDGEMENTS

- John Stasko, Georgia Tech
- Alex Kachkaev, City University London
- Aidan Slingsby, City University London

REFERENCES

- John Stasko, Carsten Görg, and Zhicheng Liu, "**Jigsaw: Supporting Investigative Analysis through Interactive Visualization**", *Information Visualization*, Vol. 7, No. 2, Summer 2008, pp. 118-132.
- Carsten Görg, Zhicheng Liu, Jaeyeon Kihm, Jaegul Choo, Haesun Park, John T. Stasko, "**Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw**", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 10, October 2013, pp. 1646-1663
- <http://www.cc.gatech.edu/gvu/ii/jigsaw/>

URLS

1. <http://mashable.com/2011/10/21/facebook-infographic/>
2. <http://markjowen.wordpress.com/2012/12/19/a-very-brief-history-of-search/>
3. <http://msr-waypoint.com/en-us/um/people/shliu/tasmc/>
4. <http://www.cloudsoftwareprogram.org/superior-user-experience>
5. <https://projects.cs.dal.ca/visualtextanalytics/>