Interactive Visualization for Computational Linguistics

ESSLII 2009

Interaction and animation
References

- Slides in this section are based on:
  

- and were designed by Petra Isenberg.

Why Interaction?

- Datasets are too large to:
  - display in one view
  - comprehend in entirety

- Interest in only subset of the data

- Interest in different views of the data

- Extract relevant information & transform

- ...
Sense-making Cycle

Card et al., 1999

The problem

The task

Defining the data

Choosing an analytical abstraction

Choosing a visual abstraction

Spatial layout, colours, etc.

Creating a view

Mapping data (values) into view

Interaction (view & value)

Explore, interpret, decide

Raw data

Data tables

Visual structures

View

Card et al., 1999
Interaction Techniques

Based on user intent

- Select – *mark something as interesting*
- Explore – *show me something else*
- Reconfigure – *show me a different arrangement*
- Encode – *show me a different representation*
- Abstract/Elaborate – *more or less detail*
- Filter – *show me something conditionally*
- Connect – *show me related items*

Yi et al., InfoVis 2007

Selection

- Mark something as interesting
- Often combined with other techniques

Collins, 2007
Explore

- Show me something else
- Examine subset of data cases (view-based)
  - E.g. Panning (move viewpoint across representation)
  - E.g. Direct Walk (move viewing focus through clicks)

Reconfigure

- Show a different arrangement
  - Move data items to
    - Enable better comparison
    - Avoid occlusion
    - Correspond to some mental model of the data
Star Glyphs

Parallel Coordinates
Rearrangement of Selected Data Items

2D Parallel Coordinates to 3D Parallel Glyphs
Star Glyphs aligned to coordinate axes

EPS Lens
EdgeLens

Show a different:
- Representation Type
- Visual appearance: Colour, Size, Shape,…

Isenberg and Carpendale, InfoVis, 2007
Encode

- Animation can aid encoding changes

![Diagram](image1.png)

DynaVis - Heer & Robertson, 2007

Abstract/Elaborate

- Show me more or less detail
  - Adjust level of abstraction
  - Detail-on-demand
  - Zooming (as long as representation isn't fundamentally altered)

![Diagram](image2.png)

Warning: Not every technique belongs to just one category.

Isenberg and Carpendale, InfoVis, 2007
Filter

- Show subset of data based on condition
  - e.g., by selecting a data range

- or filtering based on distance from focus

AlphaSlider, Ahlberg & Shneiderman, 1993; DocuBurst, Collins, 2007

Connect

- Show related items
  - e.g. brushing

Single view
Heer & boyd, InfoVis 2005

Multiple view
Collins & Carpendale, InfoVis 2007
“Overview first, zoom and filter, then details-on-demand”
Tools for Visualization
Visualization Websites
Many Eyes

Try Our Featured Visualizations
US government expenses 1962-2004
University of Michigan faculty salaries
Overweight Adults per Country
Privacy Policies

Featured Topic Hubs
Food Safety
Illness statistics, food recalls and alerts, etc.
Transportation
Planes, trains, and automobiles
OECD Factbook 2007
Official statistics

http://www.many-eyes.com
Many Eyes

- 16 visualization types
- Upload plain text, comma or tab delimited
- Discuss and share visualizations with colleagues
- 2 language-specific visualizations (tag cloud and word tree)
- Comparison views, e.g. change treemap
- Interactive, save any state of view in threaded discussion
- www.many-eyes.com
\[ F = 0.5 \times ((\text{NOUNS} + \text{ADJECTIVES} + \text{PREPOSITIONS} + \text{DETERMINERS}) - (\text{PRONOUNS} + \text{VERBS} + \text{ADVERBS} + \text{INTERJECTIONS})) + 100) \]

“If you have a look at those posts, you’ll probably notice that they aren’t really in any way more formal than Scoble’s other writing. The difference is that they tend to be more informational, i.e. have more and more condensed information crammed into to them than most entries.”
Comparison of Corporate Blogs

CorpBlawg, Cornelius Puschmann, 2008; ManyEyes, Wattenberg et al., 2007

Visualizations: Scatterplot of F-score, standard deviation and post frequency in web logs (4)

Can't see the visualization? Download the latest Java plugin here. On Macs: best viewed in Safari.

Created by: Cornelius Puschmann
Created on: Thursday February 08, 11:41 AM

Data file: F-score, standard deviation and post frequency in web logs (4)
Data source: Corporate Blogging Corpus

Not shown: 1 null item

X Axis: F-score
Y Axis: posts

Dot Size: posts

This data set has not yet been rated.
Upload and explore data.
As a preview it's rough around the edges, may your love for data guide you.

Spotlight **YouTube Generation**

![Graph showing viewers ages 18-29 and all video viewers](graph)

According to a recent report from PEW, young adults are almost twice as likely to point to YouTube as a source for online video; 49 percent of viewers ages 18-29 say they watch YouTube, compared to 27 percent of viewers of all ages. This graph shows the percentage of young adults compared to viewers of all ages by different venues of viewing.

Source: PEW

Swivel Business
Have you tried [Swivel Business](www.swivel.com)?

Swivelicious Bloggers
- Correlation does not equal causation!
- Buy Cost Of A Postage Stamp online
- GDP per capita of OECD countries
- Whine and grouse
- water
- juxtaexposed.com - daily photography from...

See more blogs >

12157 Data Sets
- Webster County MS Demographics: Population, Crime Rate
  Contributed by American Polls 2 minutes ago
  See more data >

www.swivel.com
Swivel

- Excel-type charts only
- Discussion forum
- Comma, tab delimited data upload
Programming Libraries
Prefuse is a set of software tools for creating rich interactive data visualizations. The original prefuse toolkit provides a visualization framework for the Java programming language. The prefuse flare toolkit provides visualization and animation tools for ActionScript and the Adobe Flash Player.

Prefuse supports a rich set of features for data modeling, visualization, and interaction. It provides optimized data structures for tables, graphs, and trees, a host of layout and visual encoding techniques, and support for animation, dynamic queries, integrated search, and database connectivity. Prefuse is written in Java, using the Java 2D graphics library, and is easily integrated into Java Swing applications or web applets. Prefuse is licensed under the terms of a BSD license, and can be freely used for both commercial and non-commercial purposes.

The visualization gallery and demonstration video provide numerous examples of the types of applications that can be built with the prefuse toolkit.

To learn more about prefuse, take a look at the user's manual or the frequently asked questions. For users of the alpha version of the toolkit, there is also a porting guide for migrating to the beta version.

Need help? Visit the Help Forum on SourceForge.net (You'll need a SourceForge login to post). Please be sure to include detailed information (e.g., stack traces, source code, etc.) if you need debugging help.

If you are interested in tools for ActionScript and Flash, see the prefuse flare project instead.

announcements

2008.04.02: Our friends at the IBM Visual Communication Lab are using prefuse flare to create visualizations for the Many-Eyes visualization service. Check out their new Comparison Tag Clouds, made with Flare.

2007.10.22: We’re happy to announce the first alpha release of prefuse flare, a new prefuse-based visualization library written in ActionScript! Flare brings the visualization capabilities of prefuse to the web and runs in the Adobe Flash Player.

2007.06.11: The Toronto Star, Canada’s most highly circulated daily, just ran a story on the prefuse-based Docuburst visualization! Congrats to Chris, the author of Docuburst! Check out the prefuse gallery for Docuburst and other great visualization projects.

2007.02.11: A number of new projects have been added to the prefuse gallery. Check them out!

2006.05.10: The prefuse.org website has moved to a new server, with better performance and new features. For example, you can now add comments to pages of the user manual. Apologies to any visitors who have encountered 404 errors by using outdated prefuse.sourceforge.net URLs.

2006.04.10: Prefuse has now surpassed 10,000 downloads! Thanks to everyone who has contributed to the toolkit along the way.

2006.03.03: The prefuse-based Vizster visualization appeared on the CBS crime drama Numbers! Watch the video clip (WMV, 4.7M).

releases

2007.10.21: prefuse-beta 2007.10.21 released. See the release notes for more.

2006.04.02: flare-alpha 2006.04.02 released. See the release notes for more.
prefuse

- Open source Java programming library
- BSD license
- Software architecture follows sense-making cycle
  - Standard data formats supported (I/O)
  - Interaction out-of-the-box
  - Supplied collection of layouts and renderers
- Active user support forums
- Relatively fast prototyping
- Easily link with java NLP code libraries
- Our tool of choice!
prefuse examples by Collins et al.
/**
  * Words.
  *
  * The text() function is used for writing words to the screen.
  * 
  * Created 15 January 2003
  */

size(200, 200);
background(152);

// Load the font. Fonts are located within the
// main Processing directory/folder and they
// must be placed within the data directory
// of your sketch for them to load
PFont fontA = loadFont("Ziggurat-MTF-Black-32.vlw");

// Set the font and its size (in units of pixels)
textFont(fontA, 32);

int x = 30;

// Use fill() to change the value or color of the text
fill(0);
text("zich", x, 60);
fill(51);
text("mi", x, 95);
Processing

- Open source programming language and IDE
- Simplified graphics and interaction (2D & 3D)
- Based on Java, can import Java packages
- Easy to learn
- Many help resources (online and print)
- Very wide programmer base
  - mostly designers, artists, students
Jonathan Harris, www.wefeelfine.org

Neumann et al., KeyStrokes, EuroVis 2007

Steinweber & Koller, similardiversity.net, 2008

linguistic examples made with Processing
InfoVis Toolkit

ivtk.sourceforge.net
Jean-Daniel Fekete
InfoVis Toolkit

- Graphics toolkit for Java
- Software architecture follows sense-making cycle
  - Standard data formats supported (I/O)
  - Supplied collection of layouts and renderers
- Matrix and parallel coordinates visualizations maybe especially useful for NLP
- Fast, small memory footprint
Henry et al., 20 years of 4 HCI Conferences, Int. Journal of HCI, 2008

examples made with the InfoVis Toolkit
Visualization Software
Can do more than you think!
... But be careful, new features can be misleading!
www.juiceanalytics.com has great tips
SpotFire

- Coordinated views
- Multi-dimensional data
- Customized for business intelligence, but applicable to quantitative research
- spotfire.tibco.com
Tableau

- Easily compose visualizations with “VizQL” language
- Drag and drop data columns into a library of visualizations
- Create “dashboards” of always-up-to-date data graphics
- tableausoftware.com
Customized software for syntax tree drawing

Open source initiative lead by Donald Derrick and Daniel Archambault

Lots of alternative packages, most of them not very good

http://www.ece.ubc.ca/~donaldd/treeform.htm
Emerging Research
Collaborative Visualization

Isenberg and Carpendale, InfoVis 2007
Flare is a collection of ActionScript 3 classes for building a wide variety of interactive visualizations. For example, flare can be used to build basic charts, complex animations, network diagrams, treemaps, and more. Flare is written in the ActionScript 3 programming language and can be used to build visualizations that run on the web in the Adobe Flash Player. Flare applications can be built using the free Adobe Flex SDK or Adobe’s Flex Builder IDE. Flare is based on prefuse, a full-featured visualization toolkit written in Java. Flare is open source software licensed under the terms of the BSD license, and can be freely used for both commercial and non-commercial purposes.

Take a look at our initial flare demo reel to see some of the visualizations that flare makes it easy to build.

To get up and running with flare, take a look at the Flare Tutorial and the API documentation.

Need help? Visit the Flare Help Forum on SourceForge.net (You’ll need a SourceForge login to post). Please be sure to include detailed information (e.g., stack traces, source code, etc) if you need debugging help.

Flare is just getting up and running, so please excuse any rough edges you may encounter, and look for more changes in the near future!

announcements

2008.04.02: Our friends at the IBM Visual Communication Lab are using Flare to create visualizations for the Many-Eyes visualization service. Check out their new Comparison Tag Clouds, made with Flare!

2008.02.19: A new flare release is now available.

2007.10.22: The first alpha release for prefuse flare and the launch of flare.prefuse.org!

releases

Data Sets: parliamentary text

Uploaded by: gerald
Data Source: Brown corpus
Description:

THE STV DISSECTED
A reminder first of all of the way in which the STV would work in parliamentary constituencies.
Each constituency would return from three to nine members, five being generally regarded as the optimum number.
Every elector would have one vote.
It would not be cast by marking the familiar X opposite the name of a single candidate.
Instead it would be an expression of preferences, indicated by putting the figure 1 opposite the name of
the first-preferred candidate, the figure 2 opposite the name of the second-preferred and so on.
The preferences could include as many or as few of the candidates as the voter wished.
A candidate would be elected not by securing a plurality of votes but by securing a quota of votes.
A quota is established by dividing the total number of valid votes by the number of members to be
elected plus one, and rounding up or adding one to the quotient.
In the framework of the STV this apparently strange formula is rational.
In a typical United Kingdom five-member constituency 250,000 votes might be cast.
The quota would then be 250,000 divided by five-plus-one, i.e.,
41,667.
No more than five candidates could be credited with more than 41,667 votes each.
...
Choose visualization type for parliamentary text

- Tag Cloud:
  How are you using your words? This enhanced tag cloud will show you the words popularity in the given set of text.
  Learn more

- Wordle:
  Wordle is a toy for generating "word clouds" from text that you provide. The clouds give greater prominence to words that appear more frequently in the source text.
  Learn more

- Word Tree:
  See a branching view of how a word or phrase is used in a text. Navigate the text by zooming and clicking.
Tag clouds – alpha order, size by TF
Transferring data from manyeyes.alphaworks.ibm.com...
Bigram tag cloud
Wordle – like tag clouds but arranged using energy minimization
Applet manyeyes.vis.container.ManyEyesContainer started

Word tree – kind of like a suffix graph
Phrase net: * and *
Phrase net: * of the *
Now we add the statistics from all the texts
Bar graph – first person pronoun ratio
Bar graph - 2nd person pronoun ratio
Bar graph – hifalutin token ratio
Bar graph – average sentence length (tokens)
Bubble chart - 2nd person pronoun ratio
Scatter plot – sent len x hifalutin x adverbs
Matrix chart – almost like parallel coordinates
Tree map – genre x 1\textsuperscript{st} person x 2\textsuperscript{nd} person