Galileo’s notebook on Jupiter

Alan Turing’s sketch of analysis process for a 25-letter Enigma Cipher Text.

Courtesy of King’s College, Cambridge and the UK National Cataloguing Unit for the Archives of Contemporary Scientists
A Sentence

“Mattie was here last evening, and we sat on the front door stone, and talked about life and love, and whispered our childish fancies about such blissful things -- the evening was gone so soon, and I walked home with Mattie beneath the silent moon, and wished for you, and Heaven.”

Emily Dickinson's Letters to Susan Gilbert

A Single Letter

Early June, 1852
They are cleaning house today, Susie, and I've made a flying retreat to my own little chamber, where with affection, and you, I will spend this my precious hour, most precious of all the hours which dot my flying days, and the one so dear, shall for it I barter everything, and as soon as it is gone, I am sighing for it again.

I cannot believe, dear Susie, that I have stayed without you almost a whole year long; sometimes the time seems short, and the thought of you as warm as if you had gone but yesterday, and again if years and years and hid their silent pathway, the time would seem less long. And now how soon shall I have you, shall hold you in my arms; you will forgive the tears, Susie, they are so glad to come that it is not in my heart to reprove them and send them home. I don't know why it is -- but there's something in your name, now you are taken from me, which fills my heart so full, and my eye, too. It is not that the mention graces me, no, Susie, but I think of each "sunnyside" where we have sat together, and lest there be no more, I guess is what makes the tears come. Mattie was here last evening, and we sat on the front door stone, and talked about life and love, and whispered our childish fancies about such blissful things -- the evening was gone so soon, and I walked home with Mattie beneath the silent moon, and wished for you, and Heaven. You did not come, Darling, but a bit of Heaven did, or so it seemed to us, as we walked side by side and wondered if that great blessedness which may be our's sometime, is granted now, to some. Those unions, my dear Susie, by which two lives are one, this sweet and strange adoption wherein we can but look, and are not yet admitted, how it can fill the heart, and make it gang wildly beating, how it will take us one day, and make us all it's own, and we shall not run away from it, but lie still and be happy!
Early the margin feel afternoons sugaring, weeks intimacy. Mattie,

I your's blows feel Emily book, them and flying without nor a hour –‐

I know an hour ‐‐ touched by the blossoms, it

Susie, I always, there is a daughter express, but for

Oh, Susie, you? and when they are homeless

No, you? and when they are homeless. I am

Darling Emily, I carry in the Marvel," for

Oh, Susie, you? and when they are homeless. I am

Dear Susie, you? and when they are homeless. I am

I know my one parting the sunrise, for

I know my one parting the sunrise, for

Oh, Susie, you? and when they are homeless. I am

Oh, Susie, you? and when they are homeless. I am

Oh, Susie, you? and when they are homeless. I am

Oh, Susie, you? and when they are homeless. I am

Oh, Susie, you? and when they are homeless. I am

Oh, Susie, you? and when they are homeless. I am

Oh, Susie, you? and when they are homeless. I am

Oh, Susie, you? and when they are homeless. I am

Oh, Susie, you? and when they are homeless. I am
Externalization

• External cognition is the interaction between internal and external representations when performing cognitive tasks.
• Computational offloading is the extent to which external representations can reduce the amount of cognitive effort to solve a problem.


Complexity Brings Externalization

Simple: analyze internally
Complex: sketch to aid understanding
Very complex or very large: Cannot practically sketch
Why visualize language?

• To assist in information retrieval
Why visualize language?

• To assist in information retrieval
• To enable linguistic analysis

Why visualize language?

• To assist in information retrieval
• To enable linguistic analysis
• To augment analytics on mixed data
More Generally

- Make decisions
  - Support analysis and reasoning
- Explore and discover
  - Look at things a new way
  - Encourage creativity
  - The purpose of computing is insight, not numbers (R. Hamming, 1962)
- Communicate information to another
  - Make a point
  - Tell a story
- Inspire
  - Lead to new thoughts
- Answer a question

Creating a Visualization

1. Understand a system of related information and tasks.
2. Create a mapping from the data (digital representation) to a visual representation.
3. Present this visual representation on the computer screen.
4. Provide methods of interacting with this visual representation that can include methods for varying the presentation and methods for varying the representation.
5. Verify the usefulness of the representation, the way it is presented and/or and its interaction methods.
‘Language’ Data

• Text
  – Content, form, temporality, repetition, semantics, origin
• Speech
  – Content, tonality, prosody, temporality, semantics, origin
• Other forms of language:
  – Sign languages
  – Sets of commonly known symbols (semiotics)
  – Mixed data (e.g., text + social networks, music)
  – Multimodal communications (gesture + speech)

Difficult Data

• Too much data – what to use?
  – Millions of blog posts,
  – Hundreds of thousands of news stories,
  – 183 billion emails,
  – ... per day
  – + all the text corpora!
• Data is noisy:
  – Newswire stories are syndicated (but differ slightly)
  – 70-72% of email is spam
  – Text contains section headings, figure captions, and direct quotes
  – Annotator disagreements
Visualizing Language is Difficult

• Many of the common challenges still exist:
  – Screen real estate / occlusion
  – Choosing appropriate visual variable mappings
  – Colour and perception issues
  – Maintaining “graphical integrity”
  – Interaction and usability

• Specific challenges for language?

Once you have the data ...

Most meaning comes from our minds and common understanding.

“How much is that doggy in the window?”
  – how much: social system of barter and trade (not the size of the dog)
  – “doggy” implies childlike, plaintive, probably cannot do the purchasing on their own
  – “in the window” implies behind a store window, not really inside a window, requires notion of window shopping

(Hearst, 2006)
Language is ambiguous

• Words and phrases can have many meanings, determined by context and world knowledge.
• Interesting language is often figurative:
  – “Tables encourage casual interaction.”
  vs
  – “I encouraged her to take a day off.”

Language is ambiguous

• I saw Pathfinder on Mars with a telescope.
• Pathfinder photographed Mars.
• The Pathfinder photograph mars our perception of a lifeless planet.
• The Pathfinder photograph from Ford has arrived.
• The Pathfinder forded the river without marring its paint job.

(Hearst, 2006)
Labelling

• A general issue in information visualization

• How to label a visualization while:
  – Positioning labels in a meaningful way
  – Making them long enough to be useful
  – Avoiding label overlap
  – Fonts large enough to read; minimal rotation

• If data is textual, labelling is a problem.

Data Processing Decisions

• Many levels of data processing can take place:
  – Word counting
  – Stemming: “reads” and “reading” ⇒ “read”
  – Parsing: “USA invaded Iraq” ⇒ Invaded(USA,IRAQ)
  – Summarization
  – Sentiment analysis: “Electronics shops have terrible customer service” ⇒ negative assessment
  – Word sense disambiguation: “We go to the bank to obtain a loan to purchase a boat” ⇒ bank=financial institution(72%)

• Each step of extra processing introduces uncertainty and takes time
Data Processing is Task Dependent

• What are the information needs?
• What errors may be introduced by processing?

El español es la lengua más hablada del mundo tras el chino mandarín por el número de hablantes que tienen como lengua materna.

El español es la lengua más hablada del mundo tras el chino mandarín por el número de hablantes que tienen como lengua materna.

Spanish (0.90) is the most spoken language (0.39) [about the] world (0.93) following the Chinese (0.87). Mandarin (0.80) for (0.24) the number (0.79) of (0.99) speakers (0.88) who take (0.21) it (0.95) [as a] (0.46) mother tongue (0.35).

Tasks

• Find lowest confidence Spanish words across models
• Find best overall translation by probability
• Find best overall translation by quality
• Discover areas of disagreement between models
• Investigate differences in order between languages
Supporters of Martin, who has been jailed without trial for more than two years, are calling on Prime Minister Stephen Harper to ask Mexican president Felipe Calderon to release Martin text is not preattentive under a section of the Mexican constitution that allows the government to expel undesirables from the country. Martin's supporters believe she has no chance of a fair trial in Mexico. Neither does Waage.
Visual Considerations

**Word counts**

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>god</td>
<td>3370</td>
</tr>
<tr>
<td>we</td>
<td>1971</td>
</tr>
<tr>
<td>You</td>
<td>1544</td>
</tr>
<tr>
<td>lord</td>
<td>955</td>
</tr>
<tr>
<td>hath</td>
<td>951</td>
</tr>
<tr>
<td>your</td>
<td>845</td>
</tr>
<tr>
<td>i</td>
<td>826</td>
</tr>
<tr>
<td>do</td>
<td>670</td>
</tr>
<tr>
<td>make</td>
<td>549</td>
</tr>
<tr>
<td>day</td>
<td>539</td>
</tr>
<tr>
<td>sura</td>
<td>538</td>
</tr>
<tr>
<td>give</td>
<td>502</td>
</tr>
<tr>
<td>us</td>
<td>481</td>
</tr>
<tr>
<td>verily</td>
<td>479</td>
</tr>
<tr>
<td>me</td>
<td>460</td>
</tr>
<tr>
<td>send</td>
<td>437</td>
</tr>
<tr>
<td>people</td>
<td>432</td>
</tr>
<tr>
<td>earth</td>
<td>420</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>lord</td>
<td>7872</td>
</tr>
<tr>
<td>god</td>
<td>4690</td>
</tr>
<tr>
<td>me</td>
<td>4096</td>
</tr>
<tr>
<td>son</td>
<td>3464</td>
</tr>
<tr>
<td>do</td>
<td>2979</td>
</tr>
<tr>
<td>you</td>
<td>2614</td>
</tr>
<tr>
<td>man</td>
<td>2613</td>
</tr>
<tr>
<td>king</td>
<td>2600</td>
</tr>
<tr>
<td>israel</td>
<td>2565</td>
</tr>
<tr>
<td>make</td>
<td>2491</td>
</tr>
<tr>
<td>hath</td>
<td>2264</td>
</tr>
<tr>
<td>house</td>
<td>2160</td>
</tr>
<tr>
<td>people</td>
<td>2139</td>
</tr>
<tr>
<td>child</td>
<td>2003</td>
</tr>
<tr>
<td>give</td>
<td>1872</td>
</tr>
<tr>
<td>we</td>
<td>1844</td>
</tr>
</tbody>
</table>

Simple techniques are often insufficient
Gettysburg address
Martin Luther King's famous "I have a dream" speech
http://www.textarc.org/
Term Set 1: law legal attorney lawsuit
Term Set 2: network lan
http://in-spire.pnl.gov/videos/
Document Contrast Diagram

- a visual summary of content of 2 documents
- illustrates
  - shared words,
  - words unique to one document or the other,
  - word frequency,
  - relative size of the two documents,
  - distribution of emotional tone,
  - related words based on co-occurrence,
  - most common word in each document segment.

Document Contrast Diagrams


Comparing Lexical Relationships

VisLink, Collins & Carpendale, InfoVis 2007
### Themes in Streaming News

<table>
<thead>
<tr>
<th>Month</th>
<th>Themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>May</td>
<td>Nationalization of property begins, Cuba and Soviet relations resume</td>
</tr>
<tr>
<td>Jun</td>
<td>Castro confiscates American refineries, Eisenhower breaks relations</td>
</tr>
<tr>
<td>Jul</td>
<td>Bay of Pigs</td>
</tr>
<tr>
<td>Aug</td>
<td>Yankee(63)</td>
</tr>
<tr>
<td>Sep</td>
<td>Reform(48), Response(62)</td>
</tr>
<tr>
<td>Oct</td>
<td>Oil(44)</td>
</tr>
<tr>
<td>Nov</td>
<td>Import(29), Cooperatives(16)</td>
</tr>
<tr>
<td>Dec</td>
<td>Games(7)</td>
</tr>
<tr>
<td>Jan 1961</td>
<td></td>
</tr>
</tbody>
</table>

---

### Streaming News, Thematic

- **Myanmar cyclone kills 10,000 people**
- **Secrets of horror cellars revealed**
  - Ban discussing UN help for Zimbabwe
  - China officials receive envoy from Dalai Lama, hear grievances
  - Israel says 'significant progress' in talks with Palestinians on... |
  - Bolivia debates military autonomy vote
  - Golden Brown: move to seat behind ceremonial president, prime minister to resign
- **Some Roses for Big Brown**
  - Terra, Stars move on after 407th scoring victory
- **Microsoft's Failed Yahoo Bid**: Pressure on Ballmer (Update 4)
  - "Culture clash, huge tax bill and other hurdles"
  - "What's next, we're still in talks" |
  - "Microsoft's bid for Yahoo" |
  - "How did they fail?" |
  - "Yahoo shares: fight to regain ground after open-src move" |
  - "China: Hard, foot and mouth spreading"
Wiki: web based collaboration

History Flow Wikipedia Visualization
• make broad trends in revision histories immediately visible
• preserve details for closer examination.

Wiki: web based collaboration
Enriching communication

The Shape of Conversation

- **A Visual Representation** depicting a conversation between 2 parties

- **A Visual History** of the conversation is developed like a painting on a canvas

With Edward Tse, Holly Simon & Allan Dunning
http://www.youtube.com/watch?v=k5qKkJRmlQg

The Word Tree, an Interactive Visual Concordance

Martin Wattenberg & Fernanda B. Viégas
TVCG 2008

Keywords in context
Word Tree

Search term ‘if love’
In
Romeo and Juliet

If love be rough, be rough with love
If love be blind, love cannot hit the mark
If love be blind, it best agrees with night
Word Tree

- a visual search tool for unstructured text, such as a book, article, speech or poem.
- pick a word or phrase
- See all different contexts
- contexts set in tree-like branching structure
- reveal recurrent themes and phrases.
Literary Analysis: Repetition

Martin Luther King's famous "I have a dream" speech, using the search term "I." Font sizes show frequency of use, so you can see that among King's many uses of "I," the most frequent context is the phrase "I have a dream."
With the right phrase, a word tree can reveal the heart of a data set. Below is a word tree for a collection of personals ads from men who reveal they are married:
Generative Study

- Understand visualization context:
  - How people work *without information visualization* or with pre-existing visualizations
  - How information work is situated in existing workplace practices and environment
  - How teams work together
  - Domain-specific nuances of information use

- Goal is to *describe meaning* not make statistical inference

10 Day Ethnographic Observation

- Observational study:
  - contextual interviews with research team
  - participatory observation
  - artifact analysis
Qualitative Analysis

- Open Coding
  - 6 interviews X 60-90 minutes
  - transcribe
  - code for interesting statements
  - actively work against bias (don’t only seek answers to predetermined questions)
  - refine code set and review

Visualization `in the Wild`

- Some visualizations were in common use
  - Large binders of parse trees and data tables
    - Mostly in printed form
- + Individual *ad hoc* design and usage
- Periodic analysis tasks mixed with programming
- Non interactive
Requirements

• Integrate various data tables and visualizations into an interactive multi-view platform
• Support annotation and collaboration
• Retain ability to print and annotate by hand

Sketches
Domain-Expert Feedback

Collins, 2008

Knight, 2007

Interim Prototypes

Voeckler, 2008
Interactive Prototype

Summary

• Visualization for:
  – Document content and text collections
  – Literary analysis
  – Linguistic analysis
  – Mixed-data visualization
  – Non-textual data

• Challenges include:
  – amount of data
  – ambiguity of language
  – data processing decisions and constraints (lack of linguistic sophistication)
  – designing for the task
  – perception of text