CMPSC382
Lecture 38: Visualizing Text

Prof. John Wenskovitch

12/04/2015

(adapted from Harvard CS171, Hendrik Strobelt)
Last Time

• Maps
  – Projections, Pros and Cons
  – Choropleth Maps
  – Contour and Stream Maps
  – Maps without Accurate Geography
<table>
<thead>
<tr>
<th>Proto-writing</th>
<th>Cuneiform</th>
<th>Phoenician</th>
<th>Latin letters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chauvet cave</td>
<td>Sumerian</td>
<td>Phoenician</td>
<td>Latin letters</td>
</tr>
<tr>
<td>proto-writing</td>
<td>cuneiform</td>
<td>abjad</td>
<td>letters</td>
</tr>
<tr>
<td>~20,000 years ago</td>
<td>~5,000 years ago</td>
<td>~3,000 years ago</td>
<td>~2,500 years ago</td>
</tr>
</tbody>
</table>

- **Chauvet cave proto-writing**: ~20,000 years ago
- **Sumerian cuneiform logographic**: ~5,000 years ago
- **Phoenician abjad predecessor of alphabet**: ~3,000 years ago
- **Latin letters**: ~2,500 years ago

Abstraction

- Chauvet cave painting
- Sumerian cuneiform
- Phoenician script
- Latin alphabet
Features of Text

• Abstract
• General for mental concepts
• Different across population groups
• Semi-structured
  – Content: grammar, words, sentences, paragraphs
  – Appearance: typography, calligraphy
• Legibility is key!
Visualization for Full Text

- Hypertext links are graph navigation
Chapter 7. Home sweet /home

This chapter is about configuring your environment. Now that we know how to use an editor, we can change all kinds of files to make ourselves feel better at home. After completing this chapter, you will know more about:

- Organizing your environment
- Common shell script files
- Shell configuration
- Configuring the prompt
- Configuring the graphical environment
- Sound and video applications
- Display and window managers
- How the X client-server system works
- Language and font settings
- Installing new software
- Updating existing packages

7.1. General good housekeeping

7.1.1. Introduction

As we mentioned before, it is easy enough to make a mess of the system. We can't put enough stress on the importance of keeping the place tidy. When you learn this from the start, it will become a good habit that will save you time when programming on a Linux or UNIX system or when confronted with system management tasks. Here are some ways of making life easier on yourself:

- Make a /bin directory for your program files and scripts.
- Organize non-executable files in appropriate directories, and make as many directories as you like. Examples include separate directories for images, documents, projects, downloaded files, spreadsheets, personal files, and so on.
Word Clouds
Word Tree

A

you

thou

i

if

B

Click on “blind”

if love be blind,

love cannot hit the mark.

it best agrees with right.

C

cannot forget the precious treasure of hi

is his love, and best betits the dark.

bow boy's butt - shaft and is he a man t
Language Analysis

- Literature fingerprints:

  "Fingerprints of books of Mark Twain and Jack London. Different measures for authorship attribution are tested. If a measure is able to discriminate between the two authors, the visualizations of the books that are written by the same author will equal each other more than the visualizations of books written by different authors. It can easily be seen that this is not true for every measure (e.g. Hapax Dislegomena*). Furthermore, it is interesting to observe that the book *Huckleberry Finn* sticks out in a number of measures as if it is not written by Mark Twain."

*method to measure the vocabulary richness
Time-Evolving Document Collections
Figure 6: Translation lattice for the German sentence, “Hallo, ich bin gerade auf einer Konferenz im Nationalpark in Banff.” The statistically-identified best path (along the bottom) was incorrect and has been repaired. Photo nodes provide an alternative representation for words not in the translation vocabulary. Mouse over expands the node and reveals four photos, while other nodes move away to avoid occlusion.
Text + Maps

What's in a Surname?

America is a nation of Smiths, Johnsons, and Sullivans—but also of Gamibes and Nguyenos. Zoom in on the map below to see what surnames proliferate in your part of the country.

![Map of surnames in the United States](image-url)
Any Questions?