On close and distant reading in digital humanities: a survey and future challenges

S. Jänicke\textsuperscript{1}, G. Franzini\textsuperscript{2}, M. F. Cheema\textsuperscript{1} and G. Scheuermann\textsuperscript{1}

\textsuperscript{1}Image and Signal Processing Group, Department of Computer Science, Leipzig University, Germany
\textsuperscript{2}Göttingen Centre for Digital Humanities, University of Göttingen, Germany
Introduction

• This is a survey about close and distant reading, and the different techniques and tools that can be used in each of them.
• It’s going to be fun.
• And it will be over before you even notice it.
Definition of close and distant reading

Close reading:
Definition of close and distant reading

**Close reading:**
Reading to uncovering layers of meaning that lead to deep comprehension.
Definition of close and distant reading

Close reading:
Reading to uncovering layers of meaning that lead to deep comprehension.

Close reading includes the analysis of:
Definition of close and distant reading

Close reading:
Reading to uncovering layers of meaning that lead to deep comprehension.
Close reading includes the analysis of:
• individuals, events, ideas and their development
Definition of close and distant reading

Close reading:
Reading to uncovering layers of meaning that lead to deep comprehension.
Close reading includes the analysis of:
• individuals, events, ideas and their development
• Words and phrases
Definition of close and distant reading

Close reading:
Reading to uncovering layers of meaning that lead to deep comprehension.
Close reading includes the analysis of:
- individuals, events, ideas and their development
- Words and phrases
- Text structure and style
Close reading of the second chapter of Charles Dicken’s “David Copperfield”
CHAPTER 2. I OBSERVE

The first objects that assume a distinct presence before me, as far back, into the blank of my infancy, are my mother with her pretty hair and youthful shape, and Peggoty with no shape at all, and dark that they seemed to darken their whole neighbourhood in haggardness; and cheeks and arms so hard and red that I wondered the birds peck her in preference to apples.

I believe I can remember these two at a little distance apart, dwelling to my sight by stooping down or knealing on the floor, and I go unsteadily from the one to the other. I have an impression on my mind which I cannot distinguish from actual remembrance, of the touch.
Definition of close and distant reading

Distant reading:
Definition of close and distant reading

Distant reading:
Distant reading aims to generate an abstract view by shifting from observing textual contact to observing global features of a single or multiple text(s).
This is a distant reading of “On the Road” by Jack Kerouac in the form of a tree.
LITERARY ORGANISM
A visualization of Part One of On the Road, by Jack Kerouac

BASIC STRUCTURE
Each literary component can be divided into even smaller parts. The notation is in the diagram being words. The diagram is a tree structure starting from the first chapter, paragraph, or sentence.

NOTATION
Each notation can be represented in the diagram as:
3.5 program

COLORS
- Blue: Chapters (Protagonists)
- Orange: Downtown (Protagonists)
- Green: Beat & Jazz Music
- Yellow: Social Events & Interaction
- Red: Travel

WORD COUNT CHART
All notations are in proportion to the largest sentence in Part One, which is 15 words long.
- 0 words
- 10 words
- 20 words
- 30 words
- 40 words
- 50 words
- 60 words
- 70 words
- 80 words
- 90 words
- 100 words
- 110 words
- 120 words
- 130 words
- 140 words
- 150 words
Do we still need close reading?
Do we still need close reading?

Yes.
Do we still need close reading?

Yes.

• Distant reading can’t replace close reading
Do we still need close reading?

Yes.

• Distant reading can’t replace close reading
• It can direct reader to a section that may deserve further investigation
Do we still need close reading?

Yes.

- Distant reading can’t replace close reading
- It can direct reader to a section that may deserve further investigation
- Highlights potentially interesting patterns
Close reading techniques

A visualization that need to retained the structure of the text, in order to allow close reading.
Close reading techniques

A visualization that need to retained the structure of the text, in order to allow close reading.

• Color
Close reading techniques

A visualization that need to retained the structure of the text, in order to allow close reading.
• Color
• Font size
Close reading techniques

A visualization that need to retained the structure of the text, in order to allow close reading.

- Color
- Font size
- Glyphs
Close reading techniques

A visualization that need to retained the structure of the text, in order to allow close reading.

• Color
• Font size
• Glyphs
• Connections
Close reading techniques - Color
Close reading techniques - Color

• Very common visual attribute to display features of text
Close reading techniques - Color

- Very common visual attribute to display features of text
- Color background

The Natural History of British Shell

This elegant shell occurs very rarely on the coasts of this country, we have observed it sparingly distributed on the sands near Tenby, in Pembrokeshire. Da Costa says, he was informed that it is found near Bangor, among the rocks from Bangor Ferry to Anglesea, in Wales, by which he could only mean that the species is an inhabitant of the Meneh, the arm of Beaumaris bay, communicating with the St. George’s channel, which divides Caernarvonshire from the island of Anglesea. The same writer notes it likewise from Cornwall. Dr. Pultney describes it as a scarf shell, which he had found at Neymouth. Having Da Costa’s specimen of this shell, and also that of his Pectunculus Vetula before us, we should not refrain from observing, that the opinion of Dr. Pultney respecting these shells is incorrect; they are not merely transitions in growth, or varieties of the same kind, but the difference between the two is obvious, and fully authorize us to consider them as distinct species. It should be understood in advancing this remark, that the shells which Da Costa figures and describes, for Pectunculus Vetula is clearly the Linnean Venus Paphia, a shell well known as a native of the West Indies, and never found to our knowledge in any of the European seas, Da Costa was aware after his work had been published, that he had erroneously confounded the variety of Fasciatus, Fig. 8, 9, in our Plate, with the West India shell he had conceived the latter to be the same shell in a more perfect condition, and caused it to be engraved accordingly.
Close reading techniques - Color

- Very common visual attribute to display features of text
- Color background
- Transparency of the background color for importance

The Natural History of British Shell
Close reading techniques - Color

• Very common visual attribute to display features of text.
• Color background.
• Transparency of the background color for importance.
• Font color.

Once upon a midnight dreary, while I pondered weak and weary,
Over many a quaint and curious volume of forgotten lore,
While I nodded, nearly napping, suddenly there came a tapping,
As of some one gently rapping, rapping at my chamber door.
"Tis some visitor," I muttered, "tapping at my chamber door -
Only this, and nothing more."

Ah, distinctly I remember it was in the bleak December,
And each separate dying ember wrought its ghost upon the floor.
Eagerly I wished the morrow; - vainly I had sought to borrow
From my books surcease of sorrow - sorrow for the lost Lenore -
For the rare and radiant maiden whom the angels named Lenore -
Nameless here for evermore.

The Raven by Edgar Allan Poe
Close reading techniques - Color

• Very common visual attribute to display features of text
• Color background
• Transparency of the background color for importance
• Font color
• Colored circles
Close reading techniques – Font size
Close reading techniques – Font size

• Highlight significance or weight
Close reading techniques – Font size

• Highlight significance or weight
• Encodes the number of occurrences of a word among all editions.

Seven English translations of Genesis 1:5 connecting subsequent words displayed with variable font size
Close reading techniques – Font size

• Highlight significance or weight
• Encodes the number of occurrences of a word among all editions.
• Display the number of annotations of each word

Once upon a midnight dreary, while I pondered weak
and weary,
Over many a quaint and curious volume of forgotten lore,
While I nodded, nearly napping, suddenly there came a
tapping,
As of some one gently rapping, rapping
at my chamber door.

"Tis some visitor," I muttered, "tapping at my
chamber door.
Only this, and nothing more."

The Raven by Edgar Allan Poe
Close reading techniques – Glyphs
Close reading techniques – Glyphs

• attaching glyphs to textual entities are to visualize abstract annotation that are hardly expressible with color of font size
Close reading techniques – Glyphs

• attaching glyphs to textual entities are to visualize abstract annotation that are hardly expressible with color of font size.
• Mostly used for poem.
Close reading techniques – Glyphs

- attaching glyphs to textual entities are to visualize abstract annotation that are hardly expressible with color of font size.
- Mostly used for poem.
- Use of glyphs in the form of rectangular and circular maps for the visualization of a poem’s structure.

Rainer Maria Rilke
Das I. Sonett

* Da stieg ein Baum. O reine Übersteigung!
O Orpheus singt! O hoher Baum in Ohr!
Und alles schwieg. Doch selbst in der Verschweigung
ging neuer Anfang, Wink und Wandlung vor.
Tiere aus Stille drangen aus dem klaren
gelösten Wald von Lager und Genist;
und da ergab sich, daß sie nicht aus List
und nicht aus Angst in sich so leise waren,
sondern aus Hören. Brüllen, Schrei, Geröhr
schien klein in ihren Herzen. Und wo eben
kaum ein Hütte war, dies zu empfangen,
ein Unterschluß aus dunkelstem Verlangen
mit einem Zugang, dessen Pfosten bebten, —
da schufst du ihnen Tempel im Gehör.*
Close reading techniques - Connection
Close reading techniques - Connection

• Connections aid to illustrate the structure among textual entities.
Close reading techniques - Connection

• Connections aid to illustrate the structure among textual entities.

• Track variation among various text editions.
Close reading techniques - Connection

- Connections aid to illustrate the structure among textual entities.
- Track variation among various text editions.
- Juxtapose the texts of different editions and link related text.

When in the Course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the Powers of the earth, the separate and equal Station to which they are entitled, by the Laws of Nature and of Nature’s God; - a decent respect to the opinions of mankind requires that they should declare the reasons which impel them to the separation.

We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty, and the pursuit of Happiness. That to secure these rights, Governments are instituted among Men, deriving their just powers from the consent of the governed. That whenever any Form of Government becomes destructive of these ends, it is the Right of the People to alter or to abolish it, and to institute new Government; laying its foundation on such principles and organizing its powers in such form, as to them shall seem most likely to effect their Safety and Happiness. Prudence, indeed, will dictate that Governments long established should not be changed for light and transient causes; and accordingly all experience hath shown, that mankind are more disposed to suffer, while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed, but when

United States Declaration of Independence
Close reading techniques - Connection

• Connections aid to illustrate the structure among textual entities.
• Track variation among various text editions.
• Juxtapose the texts of different editions and link related text passages
• Visualize the phonetic and semantic relations within poems.
You’re doing great.
Distant reading techniques
Distant reading techniques

A visualization that display summarized information of the given text.
Distant reading techniques

A visualization that display summarized information of the given text.

• Heat maps
• Tag clouds
• Maps
• Timelines
• Graphs
• Miscellaneous
Distant reading techniques – Heat maps
Distant reading techniques – Heat maps

• Highlight textual patterns
Distant reading techniques – Heat maps

• Highlight textual patterns
• Show relationships between texts
Distant reading techniques – Heat maps

- Highlight textual patterns
- Show relationships between texts
- Can be used to visualize the similarity or the flow of sound in poems
Distant reading techniques – Heat maps

• Highlight textual patterns
• Show relationships between texts
• Can be used to visualize the similarity or the flow of sound in poems
• For example, to highlight the systematic text re-use among the books of the bible.
Distant reading techniques – Heat maps

- Highlight textual patterns
- Show relationships between texts
- Can be used to visualize the similarity or the flow of sound in
- For example, to highlight the systematic text re-use among the books of the bible.

(texts share patterns of consecutive similar sentences)
Distant reading techniques – Tag clouds
Distant reading techniques – Tag clouds

• Intuitive visualizations for number of occurrences of words by using variable font size.
Distant reading techniques – Tag clouds

• Intuitive visualizations for number of occurrences of words by using variable font size.
• Can be limited to displaying only characteristic tags.
Distant reading techniques – Tag clouds

• Intuitive visualizations for number of occurrences of words by using variable font size.
• Can be limited to displaying only characteristic tags.
• Can be summarize to major tags for topics.
Distant reading techniques – Tag clouds

• Intuitive visualizations for number of occurrences of words by using variable font size.
• Can be limited to displaying only characteristic tags.
• Can be summarize to major tags for topics.
• Tag coloring is used to express additional information.
Distant reading techniques - Maps
Distant reading techniques - Maps

• Used to display geographic information contained in a text
Distant reading techniques - Maps

- Used to display geographic information contained in a text
- Some use thematic or density maps, but circles are more frequent

Combination of map, timeline and tag cloud for exploring commodity trading.
Distant reading techniques - Maps

• Used to display geographic information contained in a text
• Some use thematic or density maps, but circles are more frequent
• But you can use more than just circles, to encode different types of places.

- circle - cities, square - regions, triangle - countries, diamond - others.
Distant reading techniques - Timeline
Distant reading techniques - Timeline

• This is used to visualize text corpus that has various types of temporal information.
Distant reading techniques - Timeline

- This is used to visualize text corpus that has various types of temporal information.
- One approach is the straightforward use of the text’s metadata.
Distant reading techniques - Timeline

• Temporal analysis of a word’s usage in ancient Greek texts
Distant reading techniques - Timeline

- The exploration of events in news articles
Distant reading techniques - Timeline

• Sometimes, the temporal information about events reported in a text needs to be extracted in order to visualize.
Distant reading techniques – Graphs
Distant reading techniques – Graphs

• Used to visualize certain structural features of a texts
Distant reading techniques – Graphs

• Used to visualize certain structural features of a texts
• Relationships between the texts (represented by nodes) in the form of a network
Distant reading techniques – Graphs

• Used to visualize certain structural features of a text.
• Relationships between the texts (represented by nodes) in the form of a network.
• Proximity is used to show similarity based on similar paragraphs or style.
Distant reading techniques – Graphs

• Used to visualize certain structural features of a text.
• Relationships between the texts (represented by nodes) in the form of a network.
• Proximity is used to show similarity based on similar paragraphs or style.
• Phrase nets connect textual entities that appear in the certain form.
Phrase nets for ‘[word] of [word]’ within the Bible’s New Testament.
Distant reading techniques – Graphs

- Social networks are graphs visualizing the relationships between people.
Distant reading techniques – Graphs

• Social networks are graphs visualizing the relationships between people.
• Used to illustrate the relationships between characters in literary texts.
Distant reading techniques – Graphs

• Social networks are graphs visualizing the relationships between people.
• Used to illustrate the relationships between characters in literary texts.
• Size of a node can be used to encode the frequency of a character name in the text.
Distant reading techniques – Graphs

• Social networks are graphs visualizing the relationships between people.
• Used to illustrate the relationships between characters in literary texts.
• Size of a node can be used to encode the frequency of a character name in the text.
• the proximity of the nodes and the thickness of an edge can serve to reflect the strength of a relationship.
Excerpt from the social network in Mikhail Bulgakov’s Master and Margarita
The visualization of Thomas Jefferson’s social relationships, the nodes placed on a vertical axis are connected with arcs.
Distant reading techniques – Miscellaneous
Distant reading techniques – Miscellaneous

• Dust-and-magnet for Thematic analysis of historical newspaper archives metaphor
You got this.
Combining close and distant reading
Combining close and distant reading

Top down -
Combining close and distant reading

Top down -

1. A distant view on the text is shown, and the user can manipulate the visualization by filtering or by zooming
Combining close and distant reading

Top down -
1. A distant view on the text is shown, and the user can manipulate the visualization by filtering or by zooming
2. Highlights certain patterns, can direct the reader
Combining close and distant reading

Top down -
1. A distant view on the text is shown, and the user can manipulate the visualization by filtering or by zooming
2. Highlights certain patterns, can direct the reader
3. Can retrieve details-on-demand by clicking on a potentially interesting data item.
Combining close and distant reading

Top down -

1. A distant view on the text is shown, and the user can manipulate the visualization by filtering or by zooming
2. Highlights certain patterns, can direct the reader
3. Can retrieve details-on-demand by clicking on a potentially interesting data item.

1) Heat map visualizes to show similarities between texts
Combining close and distant reading

Top down -

1. A distant view on the text is shown, and the user can manipulate the visualization by filtering or by zooming
2. Highlights certain patterns, can direct the reader
3. Can retrieve details-on-demand by clicking on a potentially interesting data item.
   1) Heat map visualizes to show similarities between texts
   2) Clicking on a grid cell opens a close reading view showing the corresponding two texts juxtaposed with connections between related text passages.
Combining close and distant reading

Bottom-up
Combining close and distant reading

Bottom-up
• Rarely applied.
Combining close and distant reading

Bottom-up
- Rarely applied.
- Major focus is the source text
Combining close and distant reading

Bottom-up
- Rarely applied.
- Major focus is the source text
1. The user selects a desired text passage in a text
Combining close and distant reading

Bottom-up
- Rarely applied.
- Major focus is the source text

1. The user selects a desired text passage in a text
2. Distant reading visualizations are processed based on that selection.
Combining close and distant reading

Bottom-up

• Rarely applied.
• Major focus is the source text

1. The user selects a desired text passage in a text
2. Distant reading visualizations are processed based on that selection.

1) The user selects a certain phrase in the text
Combining close and distant reading

Bottom-up

- Rarely applied.
- Major focus is the source text

1. The user selects a desired text passage in a text
2. Distant reading visualizations are processed based on that selection.

1) The user selects a certain phrase in the text
2) That phrase is searched within the text corpus and the phrase’s distribution is shown in the form of a heat map.
Combining close and distant reading

Top-down & bottom-up -
Combining close and distant reading

Top-down & bottom-up -

• Allow for switching between close and distant reading while taking into account manipulations of the preceding view.
Combining close and distant reading

Top-down & bottom-up -

• Allow for switching between close and distant reading while taking into account manipulations of the preceding view.
• Textual entities can be selected in both the graph (distant reading) and the text itself (close reading), triggering mutual updates.
Future challenges –
Future challenges – Novel techniques for close reading
Future challenges –
Novel techniques for close reading

• In most cases only simple visualization techniques, like color coding, are provided.
Future challenges – Novel techniques for close reading

• In most cases only simple visualization techniques, like color coding, are provided.
• Displaying information about textual features and structure alongside the source text.
Future challenges -
Visualizing transposition in parallel texts
Future challenges -
Visualizing transposition in parallel texts

• Observing similarities and differences between different editions of a text and detecting transpositions.
Future challenges -
Visualizing transposition in parallel texts

• Observing similarities and differences between editions of a text and detecting transpositions.
• May occur on different text hierarchy levels.
Future challenges - Visualizing transposition in parallel texts

- Observing similarities and differences between editions of a text and detecting transpositions.
- May occur on different text hierarchy levels.
- A methods exist for the first two hierarchy levels (words, sentences)
Future challenges - Visualizing transposition in parallel texts

- Observing similarities and differences between editions of a text and detecting transpositions.
- May occur on different text hierarchy levels.
- A methods exist for the first two hierarchy levels (words, sentences)
- there’s a need for techniques to visualizing transpositions on all hierarchy levels by combining means of close and distant reading.
Future challenges –
Geospatial uncertainty
Future challenges – Geospatial uncertainty

Many visualizations deal with extracting place names in texts to illustrate geographical knowledge.
Future challenges – Geospatial uncertainty

Many visualizations deal with extracting place names in texts to illustrate geographical knowledge.

• fictional place names which are hard to represent.
Future challenges – Geospatial uncertainty

Many visualizations deal with extracting place names in texts to illustrate geographical knowledge.

• fictional place names which are hard to represent.
• Place names can themselves carry uncertainty: “Sparta” and “Atlantis”
Future challenges – Geospatial uncertainty

Many visualizations deal with extracting place names in texts to illustrate geographical knowledge.

• fictional place names which are hard to represent.
• Place names can themselves carry uncertainty: “Sparta” and “Atlantis”
• Expressions like “in London” and “close to London”
That’s it.