

Online Communication in the Digital Age
Prof. Rashmi Gaur
Department of Humanities and Social Sciences
Indian Institute of Technology
Lecture – 22
Digital Humanities and Literary Analyses

Good morning dear friends and welcome to this module. In the previous module, we had introduced the concept of digital humanities. We had discussed how digital humanities or DH shapes the digital models of scholarly discourse. DH has generated the shift in reading practices leading to a democratization and curation of knowledge. Today, we will focus on DH and literary studies. We will look at a range of approaches for the digital quantification of literary style using visualization techniques to understand new ways of conceiving narratives, digital approaches to literary geography and what the continued presence of digital methods might mean for literary studies today.

In short, we will be talking about what constitutes literary data and what is the role of the “literary”, within quotes, in the digital humanities. However, before discussing it, it is important to understand the origin and development of the digital humanities. It will help us in exploring how technology can be leveraged to advance humanistic inquiry. The origin of DH reaches back to the 1930s and 1940s.

The Origin of Digital Humanities

- The advent of Digital Humanities is often attributed to the work of the Italian Jesuit priest, Roberto Busa.
- In collaboration with IBM, he created the ***Index Thomisticus*** project.
 - It was the world's first sizable machine-readable corpus that contained an index verborum of Thomas Aquinas' writings.
- The canonical term 'Digital Humanities' emerged and developed from its competing alternative, that is, 'Humanities Computing'.
 - The latter is restricted to the application of computers in humanities scholarship and had narrower technical goals while the former incorporated a 'humanities of the digital' including the study of digitally created sources, such as art and literature.



Sources: umanisticadigitale.unibo.it,
jajandthedigitalhumanities.wordpress.com,
digitalcommons.unl.edu



Italian Jesuit scholar and philosopher Roberto Busa is known for his pioneering work in this field. Roberto Busa had received certain funding and also technological collaboration with the help of IBM and he had created the Index Thomisticus project. It was the world's first sizable machine readable corpus that contained an index verborum of the writings by and about Thomas Aquinas. In all, it had contained 137 pieces of writings. The canonical term digital humanities emerged and developed from its competing alternative that is, Humanities Computing.

The latter is restricted to the application of computers in humanities scholarship and had narrower technical goals while the digital humanities as a term incorporated a humanities of the digital including the study of digitally created sources such as art and literature. DH is profoundly multidisciplinary. It attracts contributions from scholars and scientists both within and outside the humanities. Digital humanities is profoundly multidisciplinary. It attracts contributions from scholars and scientists both within and outside the humanities as well as the humanistic social sciences.

Digital humanities have taken care to define themselves in an inclusive rather than an exclusive manner. The 1990s were a pivotal period for it. The advent of the world wide web and the availability of powerful personal computers opened up new possibilities for

the digitization and analysis of human data. Projects like Perseus Digital Library or the William Blake archive which we have discussed in the previous modules serve as examples for this.

The Origin of Digital Humanities

- The Text Encoding Initiative (TEI), founded in 1987, developed a standard for encoding texts in a machine-readable format.
- Although historically, the digital humanities have been dominated by text-oriented paradigms, the community is increasingly engaging with multimodal research objects and methods.

The Visual Turn: To engage with large data sets and to represent them visually.

The Spatial Turn: The strategic use of digital mapping is directed towards graphic analysis of location, ownership and distribution within geographic boundaries.

The Sonic Turn: Sound studies have been strongly linked to heritage and conservation, often focusing on the capture of songs, music and sounds of our cultural environment.

'Manuscriptlink', is a DH initiative, that aims to reconstruct "virtual" medieval libraries by collaborating with collections around the world to reaggregate previously lost medieval volumes.

Geographic Information System (GIS) technologies provide ways to map and compare spatial data.

Online forums like **'Sounding Out!'** provide space for publication, posts, discussion and recordings.

Examples for Each Turn

3

The Text Encoding Initiative, TEI founded in 1987 developed a standard for decoding texts in a machine readable format.

Although historically the digital humanities have been dominated by text oriented paradigms, the community is increasingly engaging with multimodal research objects and methods. We can look at three different and specific turns, the visual, the spatial and the sonic turn. The visual turn is to engage with large data sets and to represent them in a visual manner. The spatial turn suggests the strategic use of digital mapping which is directed towards graphic analysis of location, ownership and distribution within geographic boundaries. The sonic turn represents the idea that the studies of sound are strongly linked to heritage and conservation often focusing on capturing of the songs, music and sounds of our cultural environments.

In the figure on the right hand side of the slide, you can look at the examples of each one. An example of the visual aspect is a DH initiative manuscript link that aims to construct virtual medieval libraries by collaborating with collections around the world. The Geographic Information System or GIS technologies provide ways to map and compare special data and online forums like Sounding Out provides space for

publication, posts, discussion and recordings. A particular point which we must notice is that in the digital humanities, the traditional print publication forms are still there. They have not ceased to exist, but they are often complemented and supported by electronic formats.

Therefore, we can think of DH as the ways in which digital world is impacting the ways in which we think about the text versus hypertext. This is what Professor Rahul Krishna Gairola of the Department of Postcolonial Studies and Digital Humanities at Murdoch University, Australia explains in his following video.



From Perth, Western Australia, my name is Dr. Rahul Krishna Gairola and I am the Krishna Somers Senior Lecturer in English and Postcolonial Studies at Murdoch University here in Australia. I'm delighted today to be joining you from here all the way to Roorkee to talk a little bit about the meaning of digital humanities.

Now, the digital humanities, it's debatable how long it's been around. Some would say that it goes back as far as into the 1940s when IBM was working with the Nazis to create the punch card system that would lead to the extermination of over a million Jews during World War II. Some people see digital humanities as a more recent phenomenon and what I would say is that the digitization and mechanization of human life has been a defining aspect of our generation since, well before my generation even, the previous generation, since World War II because we saw a lot of mechanization and the invention

of certain weapons and chemicals that we've never seen before. This is also true in terms of the explosion of technology around not just warfare but also technologies of erasing life from Earth. That said, we could roughly say that the digital humanities has arisen in academia as a field in the past 15, 20 years and there are some very important texts that define digital humanities.

Perhaps the most, one of the most important ones is this one, *Defining Digital Humanities*, a Reader, and it makes, it offers us many different definitions for what the digital humanities is and why it's important. A more recent volume is from 2018 and it is *Disrupting the Digital Humanities* and it actually challenges the kinds of ways in which we've thought about digital humanities in the dominant sense. More recently there have been new articulations of the digital humanities, for example, post-colonial digital humanities and queer digital humanities and these are all different ways of thinking about how various lenses of identity force us to rethink the meaning of human agency in the digital milieu and the ways in which technology is shaping human life and even more so now that we've come out of a three-year global pandemic. In terms of my own definition, which is something that students and colleagues and professors all over the world ask me, how would I define digital humanities? Well, if we just look at the two terms, the first term is digital, the second term is the humanities, therefore for me the way in which I would define the digital humanities is the ways in which the digital world is impacting and reshaping the ways in which we think of the traditional humanities and for English majors the ways we even think about the text versus the hypertext and conversely the ways in which the humanities is impacting the digital world and in that sense it's a synergistic and mutual reciprocal relationship wherein the digital is impacting the ways we think about what is human and what is human is indeed impacting the ways in which we think about the digital. Now nowhere is this more evident today than in the current arguments and debates about the use of AI or chat GPT, ongoing anxieties that people have that robots will continue to take human jobs and ongoing arguments against this that there are some jobs that robots and AI simply cannot fulfill because they lack the empathy, compassion, and emotions and indeed EQ, emotional intelligence, that human beings have.

I hope that was helpful for a first definition of digital humanities and in the next short video I will offer a definition of post-colonial digital humanities. Thank you for listening. While the origin of digital humanities can be traced back to the mid-20th century it has significantly evolved over the years and found particular resonance in fields like literary studies where scholars employ computational techniques for distant reading and other innovative approaches to analyze and interpret texts. The study of literature with the aid of computers is undoubtedly controversial. Proponents of DH have explicated the possibilities for broad scale literary analysis which extends beyond the

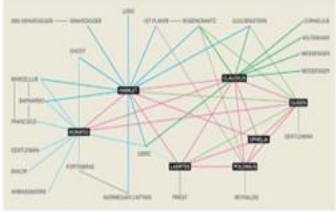
limitations of readings made by the finite human lifespan.

They also attempt to better understand the genres and forms through visualization and spatialization. They also focus on fresh perspectives and methods which might be important for rethinking core theoretical assumptions about literature itself.

Distant Reading

- 'Distant Reading' is a term coined by **Franco Moretti** to refer to the method of literary analysis that involves the analysis of extensive corpora of texts using computational tools and techniques.
 - Identifies patterns, trends and relationships like style, theme, content etc., that might be difficult or impossible to discern through traditional close reading methods.
- It focuses on the **macroscopic** view of literature and this involves the recognition that literary history is not an exhausted, well-mapped field but an 'uncharted expanse'.
- Distant reading often involves examining texts in their historical and geographical context.
 - This can reveal broader cultural and societal trends.

What Is Distant Reading?



To detect hidden aspects in the plot of *Hamlet*, Moretti turns characters into 'nodes' and their verbal exchanges into 'connections'.

Source: www.historyofinformation.com

5

Within DH and literary studies a notable methodology that has gained prominence is known as Distant Reading, that is a computational approach to literary analysis that allows scholars to examine large corpora of texts for patterns and trends. This term was coined by Franco Moretti. It refers to the method of literary analysis that involves the analysis of extensive corpora of texts using computational tools and techniques.

It often involves examining texts in their historical and geographical context and focuses on the macroscopic view of literature which involves the recognition that literary history is not an exhausted well-mapped field but is still an uncharted expanse. This approach is helpful in revealing broader cultural and societal trends. Moretti claims that within the plot and the soliloquies of Hamlet, for example, specific nodes make visible specific regions which further enable experimentation. To further elaborate the importance of distant reading, Moretti gives the example of Victorian fiction. We might know George Eliot from George Meredith but what about the 60,000 other novels of the 19th century that did not make it to the canon of Victorian fiction.

And this is where distant reading proves its functionality. It expands the sample size of texts through macro analysis. Moretti has used the phrase Slaughterhouse Of Literature as a metaphor to describe this process of analysis.

Slaughter House of Literature

- **Franco Moretti's 'Slaughter House of Literature'** is a metaphor used to describe the process of systematically analyzing a large corpus of literature for the purpose of extracting broader patterns and insights using computational methods.
- Moretti developed a coding scheme to describe the role of "clues" in *detective fiction*. He read a sample of 20 stories, taking notes on the presence or absence of each aspect of clues in order to sort the stories into a tree and to map the genesis of detective fiction.

- First bifurcation- Conan Doyle's rivals use no clues at all and all these writers are completely forgotten.
- Second bifurcation- Readers must have "discovered" clues, and writers must have understood that these curious little details were popular. However some writers (like Boothby and Hume) placed clues with no function and necessity.
- Third bifurcation- Clues are present, they have a function, but are not visible in the course of the story (we lose rivals like Pirkis, Bodken and Groller).
- Fourth bifurcation- Clues become decodable by the reader.

6

So, Franco Moretti's slaughterhouse of literature is basically a metaphor which is used to describe the process of systematically analyzing a large corpus of literature for the purpose of extracting broader patterns and insights using computational methods. He has developed a coding scheme to describe the role of clues in detective fiction.

He read a sample of 20 stories taking notes on the presence or absence of each aspect of clues in order to sort the stories into a tree and to map the genesis of detective fiction. He illustrates it further by analyzing the different bifurcations which are visible. The first bifurcation occurs when the rivals of Arthur Conan Doyle use no clues at all and all these writers are completely forgotten. The second bifurcation occurs when readers must have discovered clues and writers must have understood that these curious details were popular. However, some writers like Boothby and Hume, however, some writers like Boothby and Hume place clues with no function and necessity.


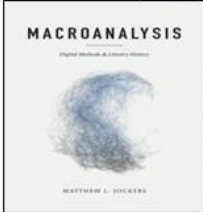
The third bifurcation occurs when clues are present. They have a function but are not

visible in the course of the story and we lose rivals like Pirkis, Bodken and Groller. The fourth bifurcation occurs when clues are decodable by the reader also. It is very interesting to note that based on the use of clues in these stories, most of the detective novels of Doyle do not make it up to the fourth bifurcation. Distant reading allows us to understand that Doyle proceeds by a method of trial and error and that he stumbled upon the clues while he was working at something different.




Doyle uses these clues as a support for Sherlock Holmes' omniscience. Moretti's distant reading has had a significant impact in shaping how scholars approach literature from a computational and quantitative perspective. But when it comes to stylometric analysis or the study of authorship attribution, it is Matthew Jockers methodology of macro analysis that is widely used.

Matthew Jockers' Macroanalysis

- Matthew L. Jockers' idea of microanalytic approach helps us to see and understand the operations of a larger "literary economy" and "the degree to which literature and the individual author....respond to or react against literary and cultural trends" (Jockers 28).
- Macroanalysis offers specific insights into:
 - the historical place of texts and authors, literary patterns and lexicons employed over time, the cultural and societal forces that impact literary style and the waxing and waning of literary themes.
- It provides a practical method for approaching questions on whether literature is evolutionary, whether there are stylistic patterns inherent to particular genres or whether literary trends correlate with historical events.



Matthew Jockers' *Macroanalysis* (2013)
Sources: usm.maine.edu, cdrh.unl.edu

7

Jocker's idea of micro analytic approach helps us to see and understand the options of a larger literary economy and the degree to which literature in the individual author respond to or react against literary and cultural traits. Macro analysis offers a specific insights into the historical place of texts and authors, literary patterns and lexicons employed over time, the cultural and societal forces that impact literary style in the waxing and waning of literary schemes.

It also provides a practical method for approaching questions on whether literature is

evolutionary, whether there are stylistic patterns inherent to particular genres or whether literary trends correlate with historical events. At its most basic, macro analytic approach is a method of gathering bits of information that may have escaped our attention because of the sheer multitude of texts out there. Let us look at an example of macro analysis of literary theme in the American writer Herman Melville's famous 1851 novel *Moby Dick* and how Jockers regards it as an outlier by examining a corpus of almost 3500 19th century novels through computational methods. Two most important things we have to look out for here are topic modelling and word cloud, which we have already discussed in the previous module.

**Macroanalysis
of Themes
in
Moby Dick
(1851)**

Topic models provide ways to analyze large volumes of unlabeled text. Using contextual clues, it can connect words with similar meanings and distinguish between uses of words with multiple meanings. The topics in a given collection of texts are not provided as a set of “themes” but rather as a set of words which are “ranked/weighted” according to their probabilities of appearing together in a given topic.

- Jockers visualizes topic-word distributions as ‘word clouds’ to derive the keywords of **Melville’s *Moby Dick***, that are central to the topic.

Figure 8.4. Top ten topics in *Moby Dick*

Top-ten topics in *Moby Dick*

- The keywords are labelled as the top ten topics in *Moby Dick*.
 - The presence of topic ‘Seas and Whaling’, a theme dealing with islands and sea voyages, is prominent in *Moby Dick*. This is followed by topics like boats and ships that deal with themes like captains and crews.
- The darker gray bars represent a corpus of 3500 19th century themes of novels that are barely noticeable.
- This proves how *Moby Dick* was an outlier in terms of the themes that dominate this 19th century corpus.

Figure 8.5. Top-ten topics in corpus

Top-ten topics from the overall Corpus
Sources: Matthew Jockers, <https://www.youtube.com/watch?v=yID7ymnaaQw>

8

Topic models provide ways to analyze large volumes of unlabeled text.

Using contextual clues, it can connect words with similar meanings and distinguish between uses of words with multiple meanings. The topics in a given collection of texts are not provided as a set of themes, but rather as a set of words which are ranked or weighed according to their probabilities of appearing together in a given topic. Jockers visualizes topic word distribution as word clouds to derive the keywords of *Moby Dick* that are central to the topic. The keywords are labeled as the top 10 topics in *Moby Dick*. For example, the presence of topic ‘seas and whaling’, a theme dealing with islands and sea voyages is prominent in *Moby Dick*.

This is followed by topics like ‘boats and ships’ that deal with themes like captains and crews. The darker gray bars in the diagram represent a corpus of 3500 19th century themes of novels that are barely noticeable. This proves how Moby Dick was an outlier in terms of the themes that dominate this 19th century corpus. If you look at the top 10 topics from the overall corpus given on the right hand side of the diagram on the slide with the exception of themes associated with ‘morning and night’, the themes that dominate the overall corpus are completely foreign to Moby Dick. Making it an outlier in terms of themes that dominate the 19th century corpus.

This also explains the reason why the novel was not very well received at the time of its publication. We will look at one more aspect from Jockers macro analysis which is literary influence. Whether consciously influenced by a predecessor or not, every book is in some sense connected to those before it through a shared pool of stylistic and thematic material.

Macroanalysis of Literary ‘Influence’

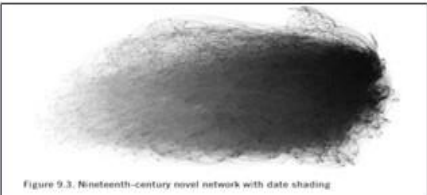


Figure 9.3. Nineteenth-century novel network with date shading

- The book nodes are coloured according to the publication years of the books.
- The lighter gray nodes indicate works from the earlier parts of the century; the darker nodes are later.
- Books are being pulled together based on the similarity of their computed stylistic and thematic distances from each other.
- Incidentally, they line up in a chronological alignment that reveals that thematic and stylistic change occur over time.
- With a few exceptions, books and authors in this network cluster into communities with their chronological peers.

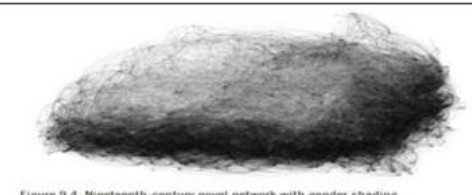


Figure 9.4. Nineteenth-century novel network with gender shading

- The network layout is reshaded according to author gender in the 19th century novels.
- Male nodes and edges are coloured lighter gray; female nodes are black.
- Works by female authors are more stylistically and thematically similar to each other and they cluster together in the South; Males are drawn together in the North.
- Three ‘outlier’ communities are also visible.
- The outlier community at the lower-right corner belongs to books authored by Margaret Oliphant who has a unique stylistic signal which is unusual both within her gender and to the network as a whole.

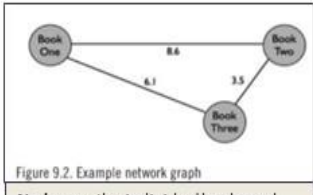


Figure 9.2. Example network graph

Nodes are the individual books and **edges** are the distance between them calculated with the Euclidean metric. Nodes with smaller distances are more similar and more closely connected.

Sources: Matthew Jockers

Before we look further into the macro analysis of literary influence, we must also understand the difference which is there between nodes and edges. Nodes are the

individual books and edges are the distance between them circulated with the Euclidean metric.

Nodes with the smaller distance are more similar and more closely connected. So, let us look at the first diagram. The book nodes are colored according to the publication year of the books. The lighter gray nodes indicate work from the earlier parts of the century and the darker nodes represent the later. Books are being pulled together based on the similarity of their compound stylistic and thematic distances from each other.

Interestingly, they also line up in a chronological alignment that reveals the thematic and stylistic change which occurs over the time. If we look at the second figure, we find that the network layout is reshaded according to author gender in the 19th century novels. Male nodes and edges are colored lighter gray whereas female nodes are black. Works by female authors are more stylistically and thematically similar to each other and they cluster together in the south.

Males are drawn together in the north. Three outlier communities are also visible. The outlier community at the lower right corner belongs to books authored by Margaret Oliphant who has a unique stylistic signal which is unusual both within her gender and to the network as a whole. A clear boundary thus can be seen which divides the network into male and female regions. At the same time, we find that there are male authors who have been placed firmly within the female dominated regions of the graph and vice versa. Thus, computer visualizations help us to verify the claims made by digital and quantitative scholarship.

David McClure, the technical director of the Stanford Literary Lab delineates how the distribution of words inside a text can have a strong thematic focus using kernel density as thematic metric.

Computational Visualization

- David McClure, in his work, '(Mental) Maps of Texts' (2014), delineates how words are distributed inside texts.
 - The distribution of some words can have a strong semantic focus and they tend to hang together with other words that orbit around a shared topic.
- McClure explains that the type of clustering where words tend to gravitate around a certain point can be captured in a kernel density estimate metric.
 - It shows how frequencies grow and decay around a specific point, rather than being an absolute measure.

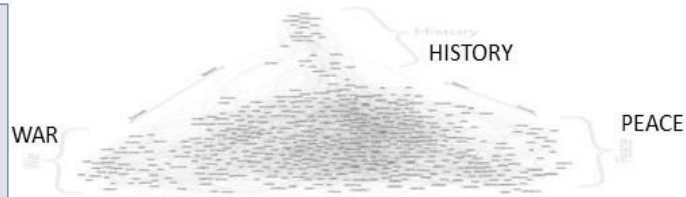


Figure 2.8 Network visualization of kernel density estimates of all terms in *War and Peace*. Image by David McClure (2014) and released under a Creative Commons Attribution License.

- McClure notes how words are distributed in Leo Tolstoy's *War and Peace* (1869).
- Algorithm manages to cluster portions of the narrative together in terms that make sense for the plot, based on linguistic similitude.
- "Napoleonic" as the man winning battles appears along the gradient between 'battles' and 'history'; "Bonaparte" representative of the Russian imagination of the man appears along the gradient between 'war' and 'peace'.
- This plotting is a type of deformation that forms a structural-thematic outline of the novel.

Sources: Martin Paul Eve

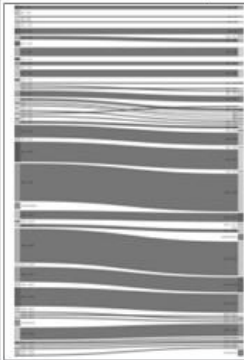
In his 2014 work, (Mental) Maps of Texts, David McClure delineates how words are distributed inside texts. The distribution of certain words can have a strong semantic focus and they tend to hang together with other words that orbit around a shared topic. He also explains that the type of clustering where words tend to gravitate around a central point can be captured in a kernel density estimate metric. He notes how words are distributed in clusters over 'War and Peace'.

Algorithm manages to cluster portions of the narrative together in terms that make sense for the plot based on linguistic similitude. Napoleon is the man winning battles appears along the gradient between battles and history. Bonaparte, representative of the Russian imagination of the man appears along the gradient between war and peace. This plotting is a type of deformation that forms the structural thematic outline of the novel.

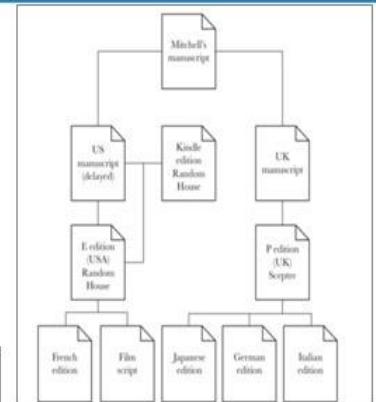
This plotting has been termed as deformation. Why? Because it deforms the original text into a now reshaped one. Another area where we can use computational visualization is in the version variance of contemporary fiction. Variance in different versions goes unremarked until academic work catches on. Let us look at David Mitchell's novel *Cloud Atlas* to understand how visualization plays its role to understand the textual variance. *Cloud Atlas* has two versions, the UK's Sceptre edition and the USA Random House edition.

Textual Variations and Visualization

- David Mitchell's novel *Cloud Atlas* (2004) has two different versions that entered general circulation alongside translations deriving from different editions (UK and USA).
- Most of the changes primarily occur in the section, 'An Orison of Sonmi~451', which is an interview with a death-penalty convict in a standard question-answer format.



- Visualization will help to understand how the novel has been reordered and how the text has been moved around, cut and edited in the two versions.
- This type of diagram, consisting of pathways, is called a **Sankey** diagram.
 - The column on the left represents the UK Sceptre edition and that on the right is the US Random House version.
 - Joined lines between the editions represent contiguous blocks of consistent, shared material.
 - The whitespace represents instances where there is no correlative block in one or the other of the texts.



Sources: Martin Paul Eve

When we compare the textual variations of these two versions, we find that most of the changes primarily occur in the section titled An Origin of Sonmi 451, which is an interview with a death penalty convict in a standard question answer format. Visualization helps us to understand how the novel has been reordered and how the text has been moved around, cut and edited in the two versions. This type of diagram consisting of pathways is called a Sankey diagram. The column on the left represents the UK Sceptre edition and that on the right is the US Random House version. Joint lines between the editions represent contiguous blocks of consistent and shared material.

The white space represents instances where there is no correlative block in one or the other of the texts. For example, it is given in the UK edition, how did you respond and in the USA the text reads, how did you respond to such blasphemous hubris. These two pieces of text do the same thing in terms of narrative progression and so are marked as functionally equivalent. Digital humanities practices have often become involved in ideas of mapping with respect to literary and historical texts. Digital mapping approaches like the Geographic Information Systems or GIS and allows for deeper understanding of the chronologies in literary texts.

Digital Mapping

- The model of literary analysis that drew particular attention to the spatial practices in literature and reading maps of literary cartography came to be called as geocriticism.
- Digital Humanities use 'Geographic Information Systems' (GIS) to map out quantitative data and to visualize and interpret geographical research.
- David Cooper and Ian N Gregory in their article, 'Mapping the English Lake District: a Literary GIS', subdivide the practice of literary cartography into two principal categories:
 - **Writerly Mapping**- It refers to the ways in which an author explicitly explores the relationship between cartography and textuality.
 - **Readerly Mapping**- It denotes the ways in which an individual may recalibrate this relationship between textual and cartographic representations of geographical space through the reading process.



Sources: www.mynextmove.org, www.cityofstuart.us

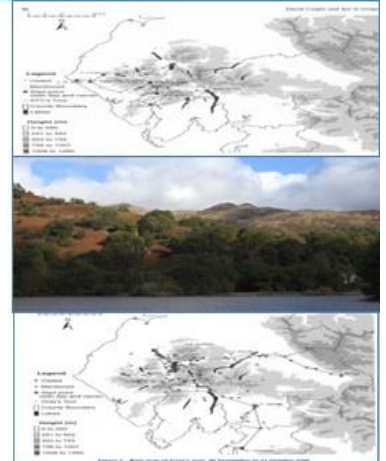


The model of literary analysis that drew particular attention to the spatial practices in literature and reading maps of literary cartography came to be called as geocriticism. Digital humanities use GIS to map out quantitative data and to visualize and interpret geographical research. Cooper and Gregory in their article, Mapping the English Lake District: A Literary GIS, subdivide the practice of literary cartography into two principal categories, writerly mapping and readerly mapping. Writerly mapping refers to the ways in which an author explicitly explores the relationship between cartography and textuality. Readerly mapping denotes the ways in which an individual may recalibrate this relationship between textual and cartographic representations of geographical spaces through the reading process.

GIS clearly intersects with readerly mapping where the individual reader can move between the cartographic and textual representations of space that have been brought together by the writer. Such phenomenological approaches to space emphasize how subjective embeddedness in space contributes to its construction.

'Mapping the Lakes' Project

- The 'Mapping the Lakes' project, funded by the British Academy and hosted at the University of Lancaster, maps the Lake District of the English poets, Thomas Gray and Samuel Taylor Coleridge.
 - It maps out two textual accounts of journeys through Lake District- Thomas Gray's tour of the region in the autumn of 1769; Coleridge's tour of the area in August 1802.
- The entire project aimed to:
 - Construct a spatial narrative by inviting the user to move through a series of experimental and exploratory cartographies.
 - To use GIS maps to document the representation of place across multiple texts and map out the emotional responses to landscape.
- Gray's 1769 tour documents a type of spatial mobilities of a tourist while Coleridge demonstrates a 'socio-spatial insiderness' that informs his walking tour.



Sources: https://youtu.be/UW07p4MSx_8?si=ZPQGi-RDOQCMoKsv



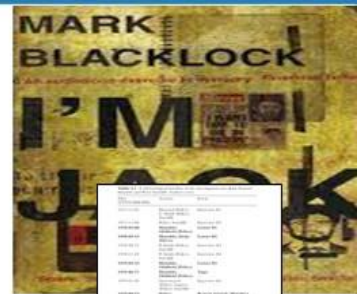
Mapping the Lakes Project maps the Lake District of the English Poets Thomas Gray and Coleridge and highlights many of the advantages of working with the GIS. This project has been funded by the British Academy and hosted at the University of Lancaster. It attempts to map the Lake District of the English Poets particularly with reference to Thomas Gray and S. T. Coleridge.

It maps out two textual accounts of journeys through the Lake District. Gray's tour of the region in the autumn of 1769 and Coleridge's tour of the area in August 1802. The entire project aimed to construct a spatial narrative by inviting the user to move through a series of experimental and exploratory cartographies. To use GIS maps to document the representation of place across multiple texts and map out the emotional response to landscape. Gray's 1769 tour documents a type of spatial mobilities of a tourist whereas Coleridge demonstrates a socio-spatial insiderness that informs his walking tour.

Such an inference was taken by assembling the geographical data and then recombining with the already known biographical accounts of Gray and Coleridge. Let us look at Mark Blacklock's 2015 novel *I am Jack* that uses cartography and geographic coordinate systems in the same way associated with digital humanities.

Mark Blacklock's *I'm Jack* (2015)

- Mark Blacklock's *I'm Jack* (2015) is primarily a novel about impersonation.
 - The protagonist John Samuel Humble, dubbed as 'Wearside Jack' sent three letters to the West Yorkshire police and the Daily Mirror impersonated as Peter Sutcliffe or the 'Yorkshire Ripper'.
- Digital mapping helps to unearth the spatial metadata from Humble's letters.
 - The various geometries, points of the address and time signatures of the events in the novel prove why Humble's hoax was so powerful.

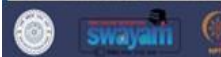
A screenshot of a data visualization tool, likely a tabulation software, showing a table of metadata for the novel 'I'm Jack'. The table has multiple columns and rows, with some cells containing text and others containing numerical values or symbols. The interface includes a search bar and various filters.

The plotting of temporal Metadata of *I'm Jack* made through the datafication method of 'tabulation'



- Metadata coordination of the letters of 'Wearside Jack'.
- The shaded polygon represents the area in which Peter Sutcliffe was operating.
- Letters were sent by Humble to the two circles within the area of Sutcliffe's crimes from a place called Sunderland.
- But the letters were sent from a triangulated portion some 100 miles of North. It is this reverse triangulation that threw detectives off the case of the Yorkshire Ripper, allowing him to roam free.

Sources: Martin Paul Eve, www.bookdealers.co.za



'I'm Jack' is primarily a novel about impersonation. The protagonist Humble, dubbed as Wearside Jack, sent three letters to the West Yorkshire police and the Daily Mirror impersonated as Peter Sutcliffe or the Yorkshire Ripper. Digital mapping helps to unearth the spatial metadata from Humble's letter.

Various geometries, points of the address and time signatures of the events in the novel prove why Humble's hoax was so powerful. So, we can look at metadata coordination of the letters of Wearside Jack. The shaded polygon represents the area in which Peter Sutcliffe was operating. Letters were sent by Humble to the two circles within the area of Sutcliffe's crime from a place called Sunderland. But the letters were sent from a triangulated portion some 100 miles of north.

It is this reverse triangulation that threw detectives off the case of the Yorkshire Ripper allowing him to roam free. The intersection of literature, geography, cartography and digital humanities have become increasingly complex. In addition to understanding these concepts, it also becomes important to understand the social forces with which a map is bound with and this entails what we call as critical cartography. In this slide, let us look at three terms. The first is literary geography, the second is literary cartography and the third is critical cartography.

Literary Geography or Literary Cartography?

- **Literary Geography** demonstrates that spatial dimensions of fictional accounts set along a scale of localizations can actually be a key to understanding the whole plot.
 - It relates only to the reader.
- **Literary Cartography** is a sub-discipline to literary geography where spatial elements of fictional texts are translated into cartographic symbols.
 - It concerns the writer as map-maker.
- **Critical Cartography** involves a paradigm shift from 'trusting' the map to 'mistrusting' it.
 - It shifts the focus from the map itself to the social forces with which it is bound up.
 - It aligns with the 'spatiality of the postmodern' where globalization and forces of capitalism come into force.



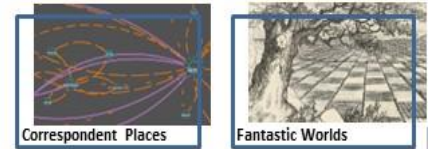
Source: travelbetweenthepages.com

Literary geography demonstrates that spatial dimensions of fictional accounts set along a scale of localizations can actually be a key to understanding the whole plot and it relates only to the reader. Literary cartography is a sub-discipline to literary geography where special elements of fictional texts are translated into cartographic symbols and it concerns the writer only as a map maker. Critical cartography involves a paradigm shift from trusting the map to mistrusting it. It shifts the focus from the map itself to the social forces with which it is bound up. It also aligns with the spatiality of the postmodern where globalization and forces of capitalism come into force.

The geopolitical organization and disruptions in the post-war era called attention to this political essence of geography, but how are we able to map literary text especially those representing fictional realms? Let us analyze this through the project chronotopic cartographies. This project is based on the famous Russian literary critic Mikhail Bakhtin's idea of chronotope.

Chronotopic Cartographies

- **Chronotopic Cartographies** (2017-2020) is a project by writer Sally Bushel that investigates how we can use digital tools to analyze, map and visualize the spaces of literary texts.
- It draws on the literary theorist Mikhail Bakhtin's concept of the chronotope- a way of describing how time and place are linked and represented in different spatial forms of literature.
 - The project focuses on five spatial types- Correspondent Places, Indefinite Spaces, Nested Worlds, Fantastic Worlds, Spaces of Exile
 - It is a custom XML schema designed to capture the qualitative linguistic and narratological aspects of literary spaces and temporalities.



Spatial Types
Source:Sally Bushel



Chronotopic cartographies is a project by Sally Bushel. It was conducted between 2017 to 2020. It investigated how we can use digital tools to analyze map and visualizes the spaces of literary texts.

It draws on the literary theorist Bakhtin's concept of the chronotope which is a way of describing how time and place are linked and represented in different spatial forms of literature. The project focuses on five spatial types, correspondent places, indefinite spaces, nested worlds, fantastic worlds and spaces of exile. It is a custom XML schema designed to capture the qualitative linguistic and narratological aspects of literary spaces and temporalities. Let us take the example of the special type. Correspondent spaces are visualized in Virginia Woolf's *To The Lighthouse* also.

Bushel analyzes three realms, the complete map or syuzhet which is the narrative order of the story. Second is the topoi which suggests the recurring themes and finally chronotopes.

Chronotopic Cartography in *To The Lighthouse**

- Visualisation of Virginia Woolf's *To the Lighthouse* (1927) belongs to the spatial type of 'Correspondent Places'.
 - This group of texts contains worlds that can be referenced geographically and historically to real world spaces (Bakhtin defines these as "concrete").



*Refer Sally Bushel- Chronotopic Cartography

- The *syuzhet* map registers the underlying spatiality of the three parts of the novel- 'The Window', 'Time Passes' and 'The Lighthouse'.
- The 'Window' and 'The Lighthouse' – are connected by the horizontal line of time, made visible as a corridor.
- The middle part's *topoi* are joined by jumps, interrupts and intra-textual connections that imply the prominence of time over space.
- This all speaks for the telling of the tale – the narrative abruptly shifts amongst multiple perspectives.
- Overall, the narrative form is *cyclical*. This shows that different spaces hold different values depending on which subjective consciousness the narrator is depicting.



To the Lighthouse- Virginia Woolf
Source:bookanalysis.com

So, visualization of Woolf's novel belongs to the spatial type of correspondent places. This group of texts contains worlds that can be referenced geographically and historically to real world spaces and Bakhtin has defined these as concrete. The *syuzhet* registers the underlying spatiality of the three parts of the novel, the window, time passes and the lighthouse.

The window and the lighthouse are connected by the horizontal line of time which is made visible as a corridor to us. The middle parts *topoi* are joined by jumps, interruptions and intra textual connections that imply the prominence of time over space. This all speaks for the telling of the tale. The narrative abruptly shifts amongst multiple perspectives.


Overall the narrative form can be termed as a cyclical one. This shows that different spaces hold different values depending on which subjective consciousness the narrator is depicting at the moment. These spaces and chronotopes correspond to the book's narrative method of depicting the merging and flowing and creating of consciousness. One of the most basic things that we do with computational techniques is to analyze various linguistic patterns and other idiosyncrasies in a body of text. A set of a



stylometric techniques can help us to understand how authors style change over time and how to determine the likely authorship of a text when the authorship is uncertain.

Stylometry

- Computational stylometry is the use of computers to measure the stylistic properties of texts.
 - It is used for genre classification, diachronic linguistics, literary analysis and authorship attribution.
- Stylometry is often employed to identify or confirm the authorship of texts, especially in cases where authorship is uncertain or disputed.
- Sentiment analysis techniques are used in conjunction with stylometry to study the emotional tone and mood conveyed in texts.
- This is particularly useful when dealing with large datasets where manual categorization would be impractical.



Source: serhack.me

IIT ROORKEE | NPTEL ONLINE CERTIFICATION COURSE

18

Computational stylometry is the use of computers to measure the stylistic properties of texts.

It is used for genre classification, diachronic linguistics, literary analysis and authorship attribution. Stylometry is often employed to identify or confirm the authorship of texts especially in cases where authorship is uncertain or disputed. Sentiment analysis techniques are used in conjunction with a stylometry to study the emotional tone and mood conveyed in texts. This is particularly used when dealing with large data sets where manual categorization would be impractical. Most of the topics that we have discussed today whether macro analysis or distant reading correspond to one or the other aspects of stylometry.

Stylometry risks on many assumptions and formal aesthetics which cannot be covered over the course of this module today. However, we will discuss the concept of the stylometric approaches further, look at its applications and examples in the coming module.

Conclusion

- Digital approaches to literary studies often reintroduce empiricism to the heart of literary scholarship.
 - It brings us back from the brink of elaborate theorization and argumentation to a focus on the text itself.
- Such methods have been united under the term of 'descriptive criticism' which includes, on the one hand, surface reading and on the other hand, distant reading and computational analysis.
- Computational approaches afford us alienation mechanisms that can allow us to understand literary works in greater detail. These methods can show us the longer-scale literary histories within which individual works are situated and their 'invisible' functions.



We can say that the digital approaches to literary studies often reintroduce empiricism to the heart of literary scholarship. It brings us back from the brink of elaborate theorization and argumentation to a focus on the text itself. Such methods have been united under the term of descriptive criticism which includes on the one hand surface reading and on the other hand distant reading and computational analysis.

Computational approaches afford us alienation mechanisms that can allow us to understand literary works in greater detail. These methods can show us the longer scale literary histories within which individual works are situated and their invisible functions. In the next module, we will continue our discussion of stylometric analysis. We will also look into critical digital humanities as an interdisciplinary approach within the field of digital humanities, critiquing the assumptions, values and power structures embedded in digital technologies and their applications also. Thank you. .