Online Communication in the Digital Age

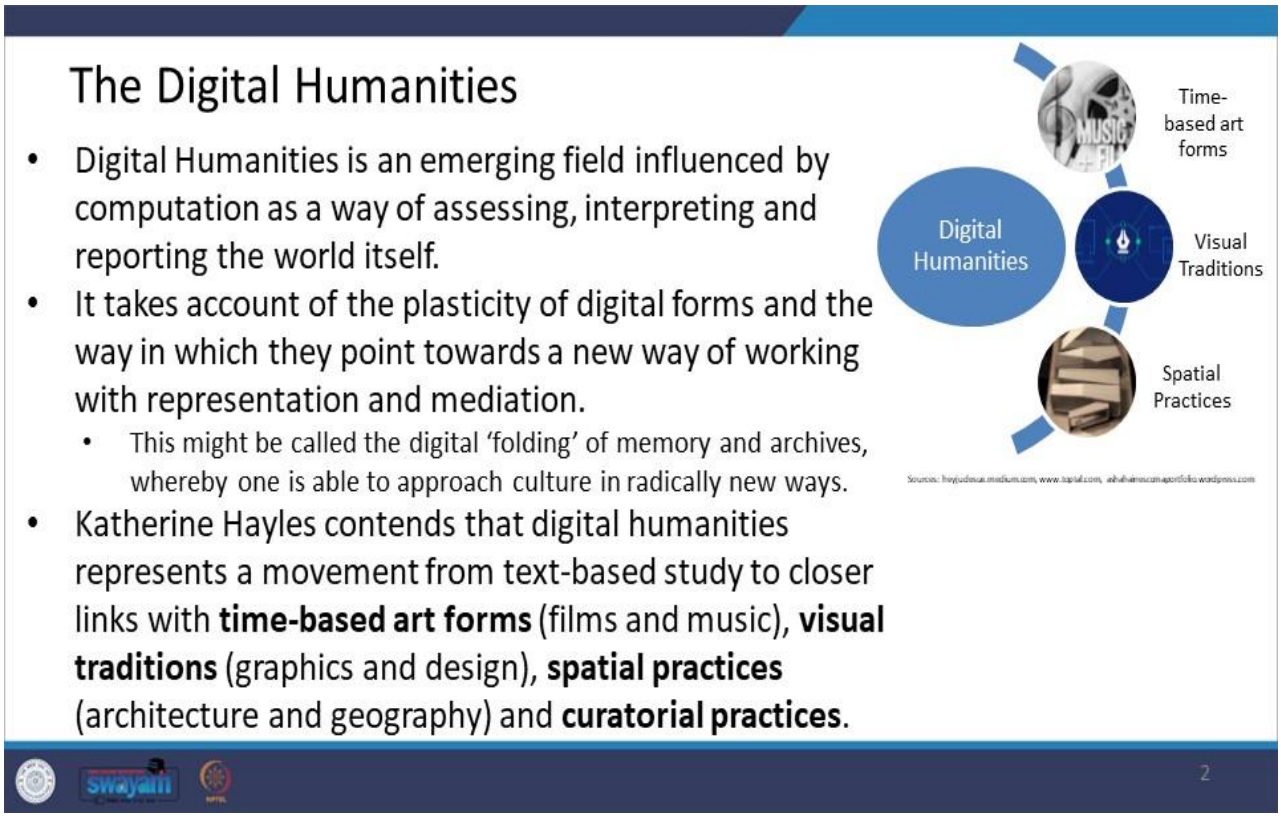
Prof. Rashmi Gaur
Department of Humanities and Social Sciences
Indian Institute of Technology

Lecture – 21

Introduction to the Digital Humanities

Good morning friends and welcome to today's module. In the previous module, we had discussed the digital production of literary texts and digital authorship. We had also talked about the 21st century hypertext, ambient literature, and the new frontiers of selling literature in the digital era. Today, we will begin our discussion with an introduction to digital humanities, its basic challenges and evolution. We will also try to understand the potential of digital humanities for significant change when it comes to practices associated with computational techniques and technological changes. Today, digital technology plays a significant role in the way research is changing.

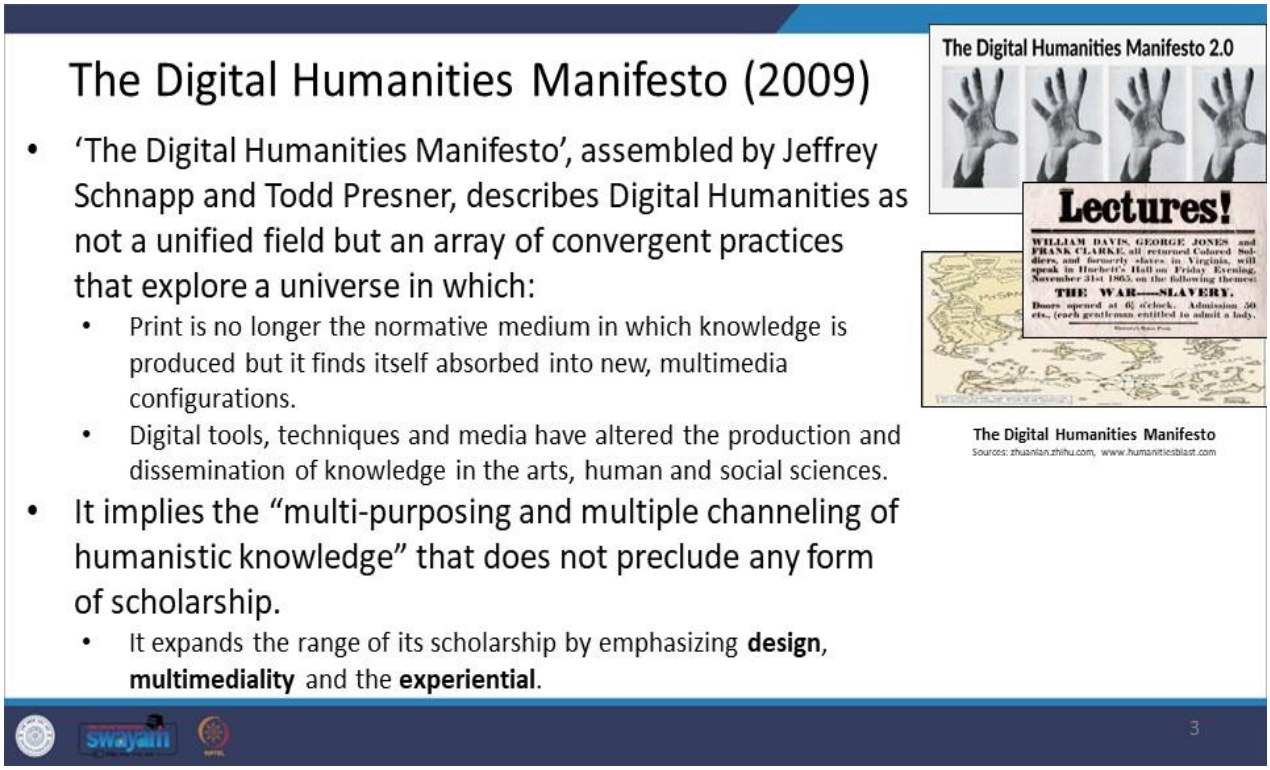
Library catalogues, emails, Google searches, and bibliographic databases have now become crucial for every professional. While some may decry the loss of techniques and skills of older research traditions, others have warmly embraced what has come to be called the digital humanities. It signals how computational techniques have become an intrinsic part of the research process and are linked with many of the questions raised by every branch of humanities today. It demonstrates the contributions of contemporary humanities scholarship to new modes of knowledge formation enabled by networked digital environments.



Digital humanities is an emerging field influenced by computation as a way of assessing, interpreting, and reporting the world itself. It takes account of the plasticity of digital forms and the way in which they point towards a new way of working with representation. This might be called the digital folding of memory and archives whereby one is to able to approach culture in radically new ways.

Katherine Hayles contends that digital humanities represents a movement from text-based study to closer links with time-based art forms, that is film and music, visual traditions, that is graphics and designs, special practices, architecture and geography for example, and curatorial practices. It adopts a perspective in which human cognition is enhanced by computer functions and in doing so, the gap between the methodological applications of digital technology within the humanities and the idea of a systematic integration of the human in the digital gets bridged. Jeffrey Schnapp and Todd Presner have authored The Digital Humanities Manifesto where they talk about what exactly and what exactly is not digital humanities. Why is it important today and how does it shape the digital models of scholarly discourse for the newly emergent public spheres and how does it facilitate the production and exchange of networks of knowledge. In 2009, Jeffrey

Schnapp and Todd Presner assembled the digital humanities manifesto with some help from other colleagues too.



As they suggest digital humanities or DH is not a unified field but an array of convergent practices that explore a universe in which print is no longer the normative medium in which knowledge is produced but it finds itself absorbed into new multimedia configurations. As we see digital tools, techniques and media have altered the production and dissemination of knowledge in the arts, human and social sciences and it implies the multi-purposing and multiple channeling of humanistic knowledge that does not preclude any form of scholarship. It expands the range of its scholarship by emphasizing design, multimediality and the experiential. The manifesto raises pertinent questions about the role of humanities within the digital environment. More so as our cultural heritage as a species is fast migrating to digital formats.

Because of its descent from the counterculture that is the cyber culture intertwinglings of the 60s and the 70s, DH values the democratization of culture and scholarship promoting collaboration and creation across domains of expertise.

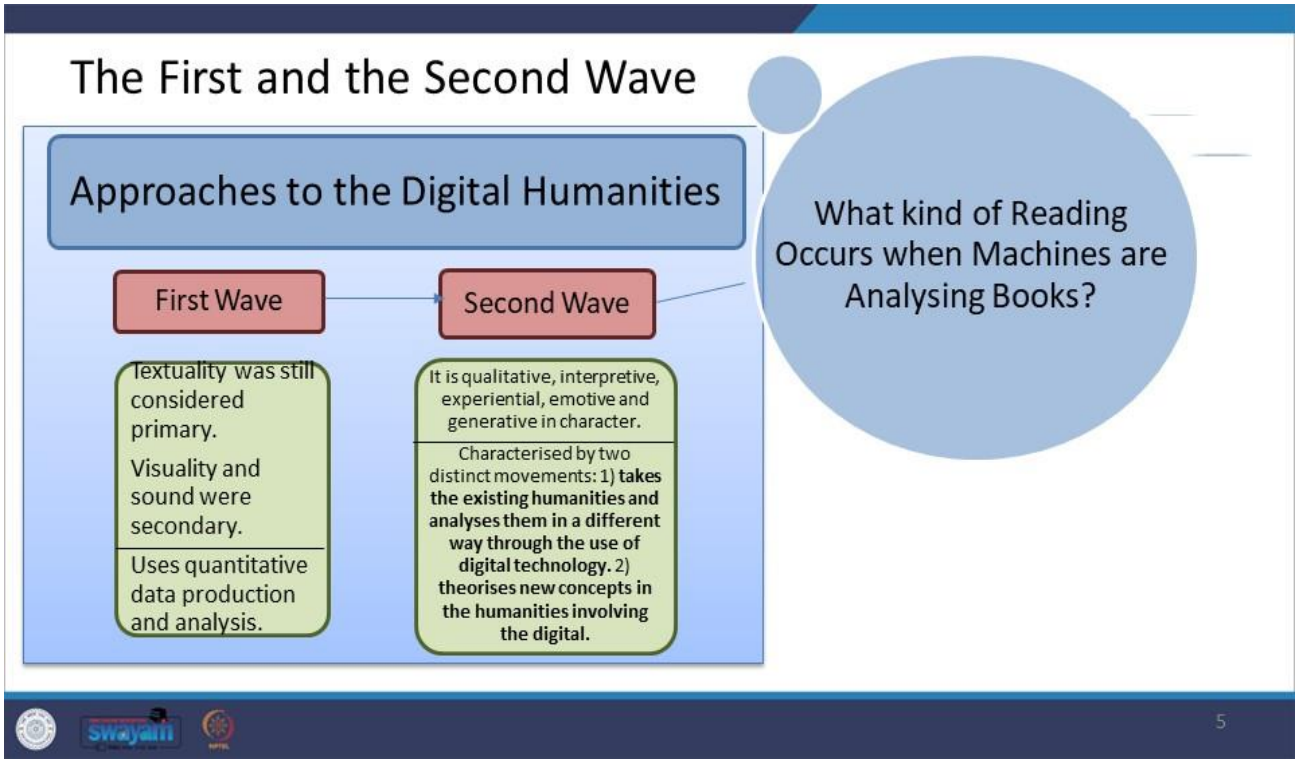
The Digital Humanities Manifesto

- The manifesto advocates a '**neo- or post-print**' model where print becomes embedded within a multiplicity of media practices and forms of knowledge production.
- Knowledge assumes multiple forms.
  - Designing and determining the interface to information, data and knowledge becomes just as central as the crafts of writing, curating and coordinating.
- Digital Humanities deconstructs the very materiality, methods and media of humanistic enquiry.
- It recasts the scholar as 'curator' and curator as scholar.
  - Arguments are curated through objects as well as words, images and sounds.
  - It implies spatialization of both physical and virtual spaces rather than in language alone.

Jeffrey Schnapp and Todd Presner
Sources:elts.ucla.edu, www.morningfuture.com/en

The manifesto advocates a neo or post-print model where print is embedded within a multiplicity of media practices and forms of knowledge production. Knowledge also assumes multiple forms designing and determining the interface to information data and knowledge becomes just as central as the crafts of writing, curating and coordinating. Digital humanities deconstructs the very materiality, methods and media of humanistic inquiry. It recast the scholar as curator and curator as a scholar.

Arguments are curated through objects as well as words, images and sounds and it also implies a specialization of both physical and virtual spaces rather than in language alone. This curation is most explicit in virtual galleries and the transformation and expansion of other scholarly landscapes. DH harnesses the global nature of today's research communities as a post-disciplinary opportunity of our time. It emphasizes models of knowledge production and reproduction that leverage the increasingly distributed social and cultural nature of technologies.



The approaches taken towards understanding digital humanities can be categorized into two waves of activity, the first wave and the second wave.

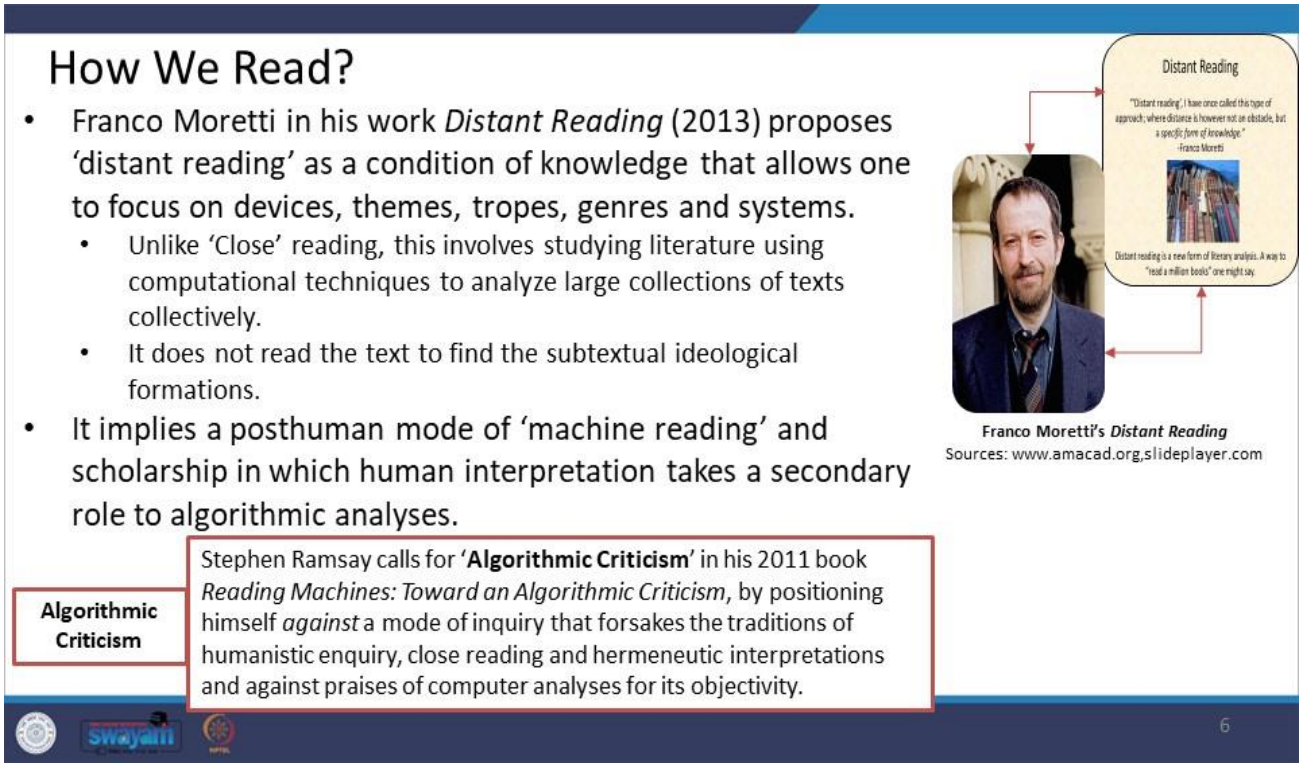
However, we should also keep in mind that such classifications are contested. So, if we look at the classification and approaches to the digital humanities, we find that the first wave is the times when textuality was still considered primary, visuality and sound were considered to be secondary. It uses quantitative data production and analysis. In comparison, we find that the second wave is qualitative, interpretative, experiential, emotive and generative in character. It is characterized by two distinct movements.

It takes the existing humanities and analyzes them in a different way through the use of digital technologies. And secondly, it theorizes new concepts in the humanities involving the digital. So, these waves suggest how the manner of reading has transitioned, has undergone a drastic transition. So, what kind of reading now occurs when machines are analyzing book? As now the reading has become interpretive and emotive in character. The critical question is what kind of reading occurs when machines are analyzing books? It can perhaps never reveal the final detail and meanings that a traditional analysis can provide to us.

It produces a computational analysis of the structure, frequency of lexical choices,

patterns of language use etc. This itself forms the foundation of a qualitative reading of the material. The text itself is treated to a hermeneutic analysis with the advantage of digital tools. This is called text mining which we will discuss later in a module in detail. What we have to remember is that this can uncover patterns, themes and information that a traditional close reading might still miss.

This shift in the reading patterns illustrates a transition from the first wave to the second wave.



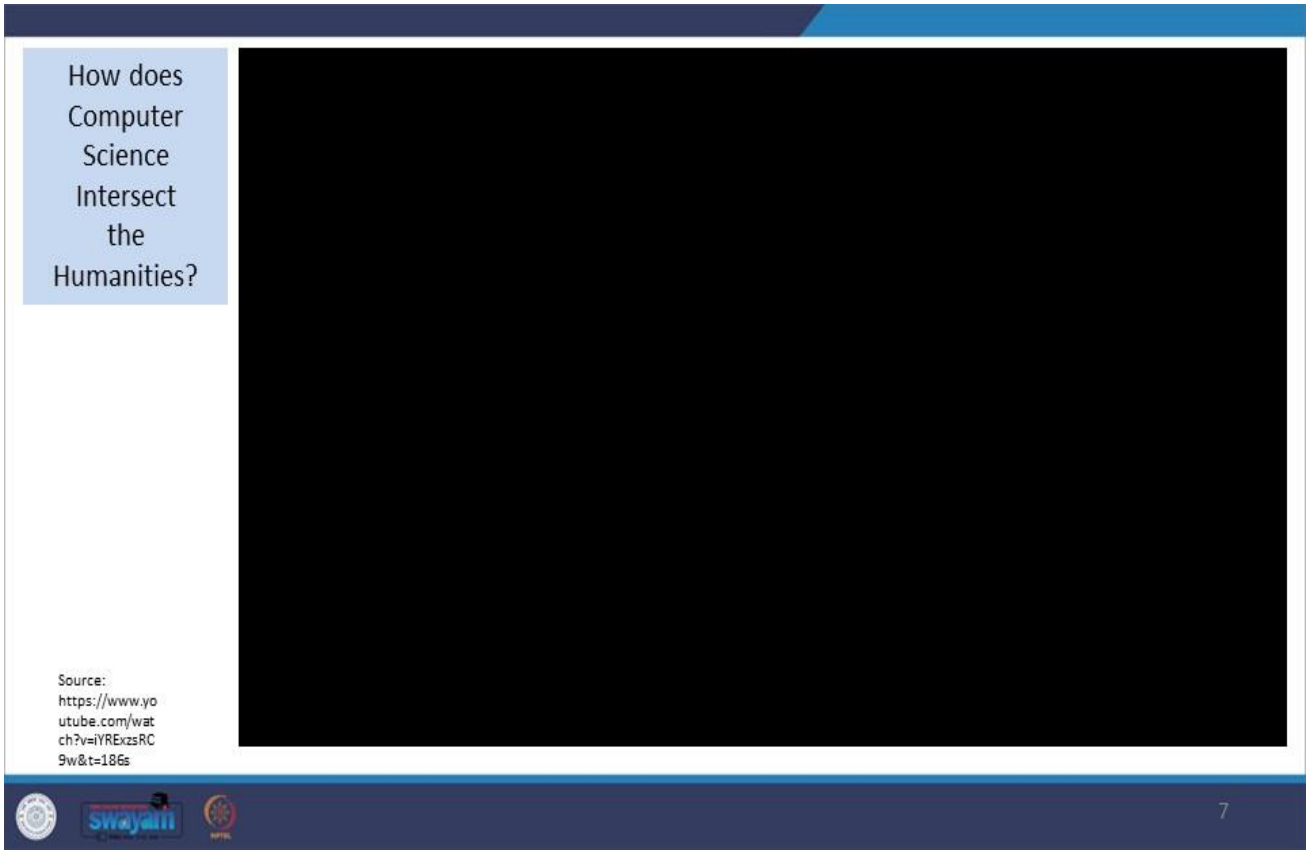
Franco Moretti's *Distant Reading*
Sources: www.amacad.org,slideplayer.com

It would be pertinent to refer Franco Moretti, who has proposed distant reading as a condition of knowledge that allows one to focus on devices, themes, tropes, genres and systems. Unlike close reading, this involves studying literature using computational techniques to analyze large collection of texts collectively. It does not read the text to find the subtextual ideological formations. It implies a posthuman mode of machine reading and scholarship in which the human interpretation takes a secondary role to algorithmic analysis.

If we talk about the argument of Franco Moretti, we should also perhaps discuss the argument provided by Stephen Ramsay. Stephen Ramsay calls for algorithmic criticism in his 2011 book of the same name by positioning himself against a mode of inquiry that foresakes the traditions of humanistic inquiry, close reading and hermeneutic interpretations and against praises of computer analysis for its objectivity. Visual

representations like graphs, charts and maps are often used to gain insights from data. They involve the identification of recurring patterns across a large body of text. This includes patterns in word usage, themes and narrative structures, etc.

In the following video, we can see Jeffrey Schnapp, one of the writers of The Digital Humanities Manifesto and professor at Harvard Graduate School talking about the boundaries of the humanities, the evolution of digital humanities and what exactly encompasses digital humanities as a new form of production of knowledge.

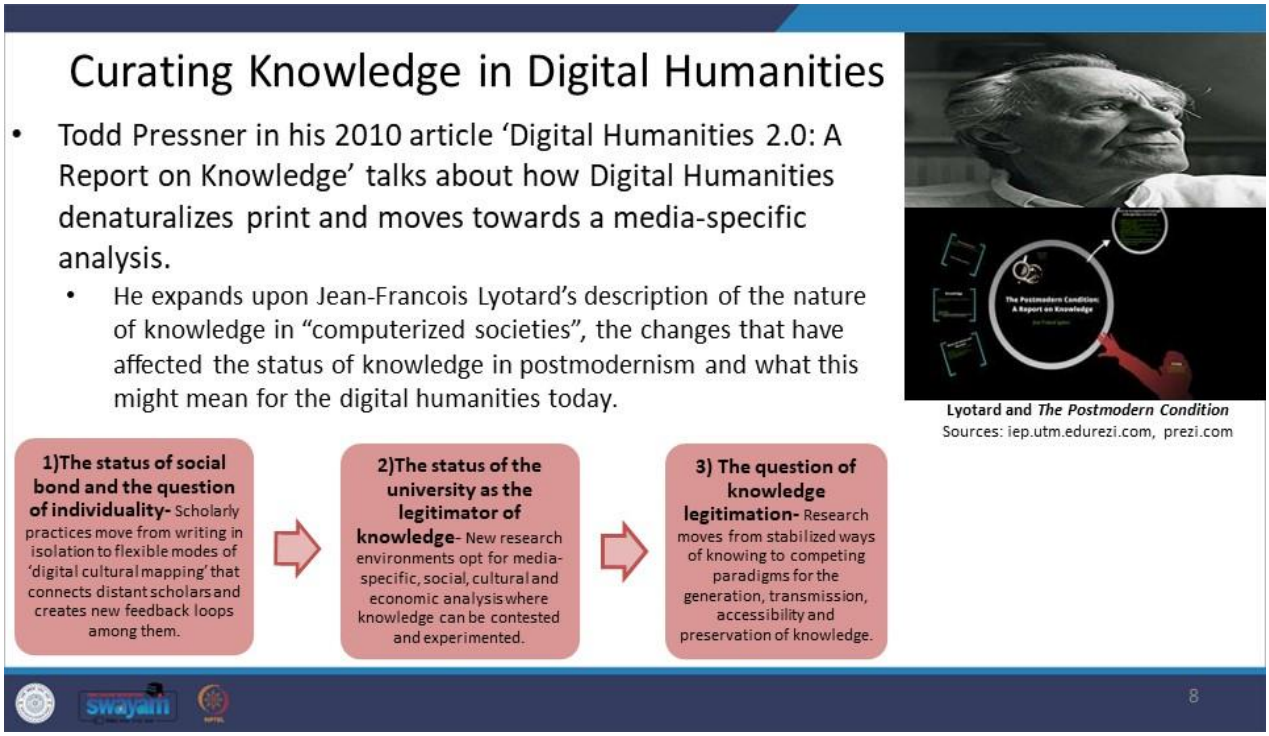


The humanities disciplines have been undergoing a very radical transformation over the course of the last 20 years, but it is a transformation that has deep roots in the history of the evolution of those disciplines that make up the core humanities. Starting really in the immediate post-World War II era when mainframe computing began to have a powerful impact on other areas of research, some early pioneers of what at the time was called computational humanities or computing in the humanities or humanistic computing or humanities informatics, it had many, many different labels, began experimenting with applying the power of mainframe computers and systems to the analysis of cultural objects, in particular literary texts. So starting in that period, we begin to see a kind of conversation emerging between the core humanities disciplines, whether initially certainly text-based but gradually encompassing other areas of culture and the whole

area of informatics and computer science. That conversation goes on in a rather marginal way with respect to its impact on the core humanities disciplines, probably for its first 20 or 30 years, but really comes to a head I think in the 1980s, starting in the 1980s but really coming to a head in the 1990s with the revolution in personal computing and the emergence of the World Wide Web as the defining public space of our time.

And it's in that context that what now has come to be called digital humanities really becomes a transformational area of experimental scholarship within the humanities disciplines. I call it an area of experimentation because it isn't a new department, a new field in a conventional sense, it simply represents an umbrella under which we can group together a whole series of experimental approaches to some of the core questions that have animated the humanities disciplines from the very beginning, questions like what is the meaning of a given cultural object, how do different categories of objects relate to aspects of the social history of a given country or place, what is truth, beauty, meaning in a given social circumstance, how do myths or stories undergo transformations, how do ideas become symbols, objects, how do they get translated into aesthetic objects and vice versa. All these kinds of questions come to animate what has come to be called the digital humanities today. Schnapp explains how computational humanities transitioned into the digital humanities as we know it today and how in the present era digital technologies are widely used in the analysis of cultural artifacts especially literary texts. Lyotard in his most prominent work, The Postmodern Condition has talked about how and also what knowledge is communicated in the light of new technological and social changes in the post industrial age.

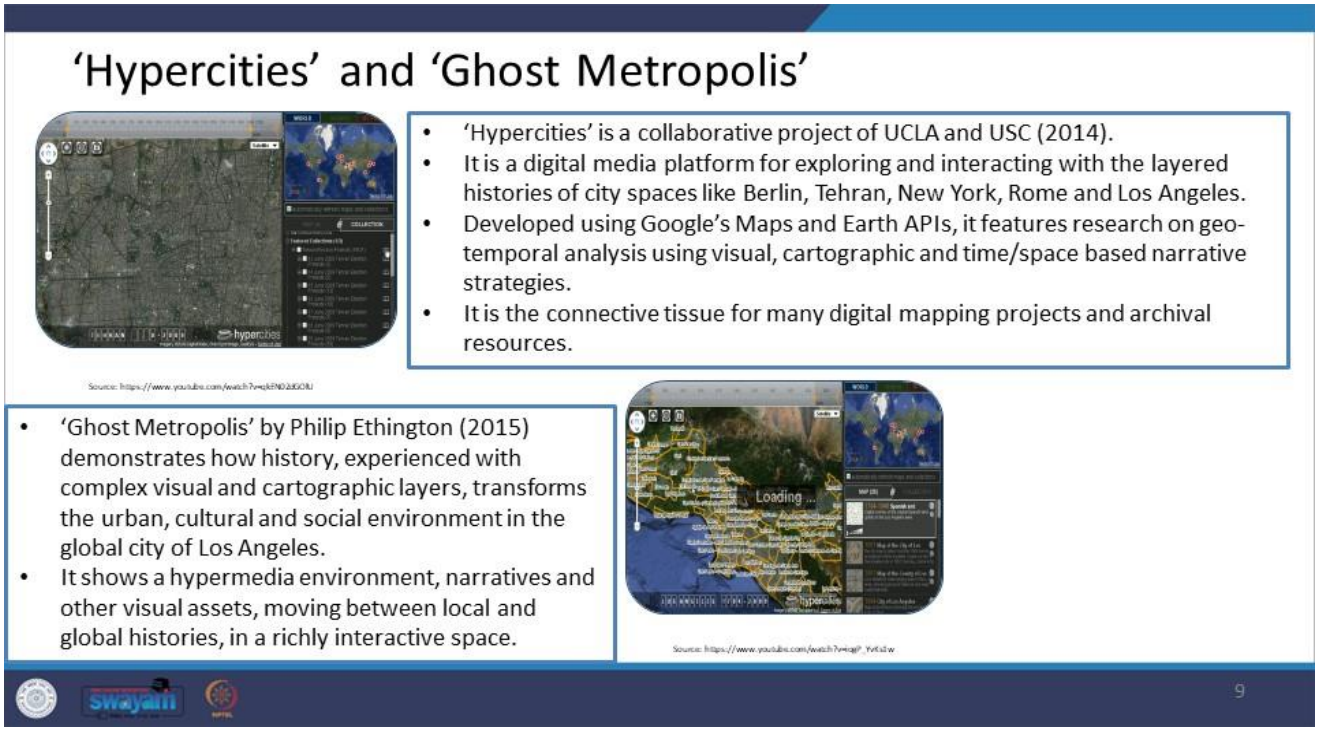
Scientific knowledge has been transformed into bits of information with the rise of cybernetics and informatics. Knowledge has become something to be managed, controlled and sold. Let us look at how precisely this means in the context of digital humanities today.



Todd Pressner in his 2010 article Digital Humanities 2.0: A Report On Knowledge talks about how digital humanities denaturalizes print and moves towards a media specific analysis.

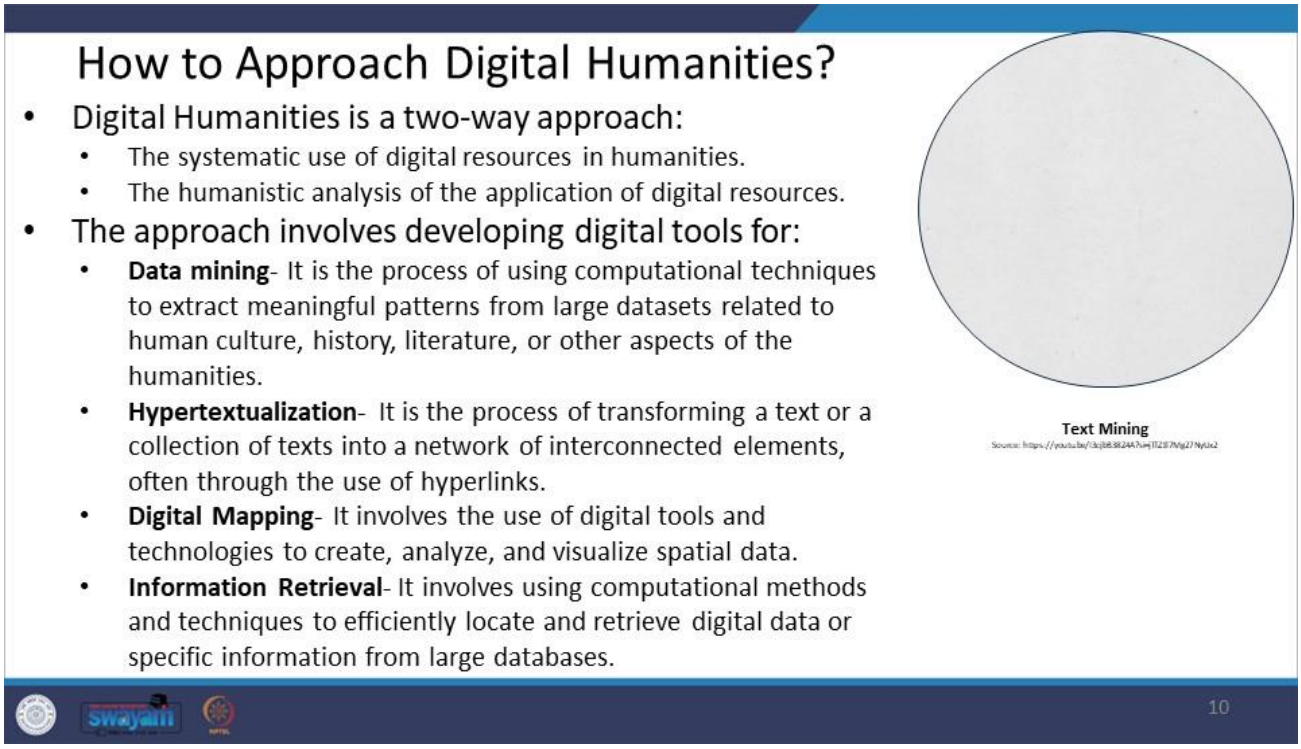
He expands upon Lyotard's description of the nature of knowledge in computerized societies, the changes that have affected the status of knowledge in postmodernism and what this might mean for the digital humanities today. Firstly, he talks about the status of social bond in the question of individuality. Practices move from writing in isolation to flexible modes of digital cultural mapping that connects distant scholars and creates new feedback loop amongst them. Secondly, we should be aware of the status of the university as the legitimator of knowledge. New research environments opt for media specific social, cultural and economic analysis where knowledge can be contested and experimented.

And thirdly, it is the question of the legitimation of knowledge itself. Research has moved from stabilized ways of knowing to competing paradigms for the generation, transmission, accessibility and preservation of knowledge. Presner grounds his argument on the potential of technologies to exercise exclusionary control over information as well as to democratize it by opening up access and use. This becomes the persistent dialectic of any communication technology. Let us now look at some examples of web-based researches associated with digital humanities.



So, we can refer to a collaborative project of UCLA and USC started in 2014 and known as HyperCities. It is a digital media platform for exploring and interacting with the layered histories of city spaces like Berlin, Tehran, New York, Rome, Los Angeles. It is developed by using Google's map and earth's APIs and it features research on geo-temporal analysis using visual, cartographic and time-space based narrative strategies. And it is the connective tissue for many digital mapping projects and archival resources. The second is Ghost Metropolis by Philip Ethington taken up in 2015 which demonstrates how history experienced with complex visual and cartographic layers transforms the urban, cultural and social environment in the global city of Los Angeles.

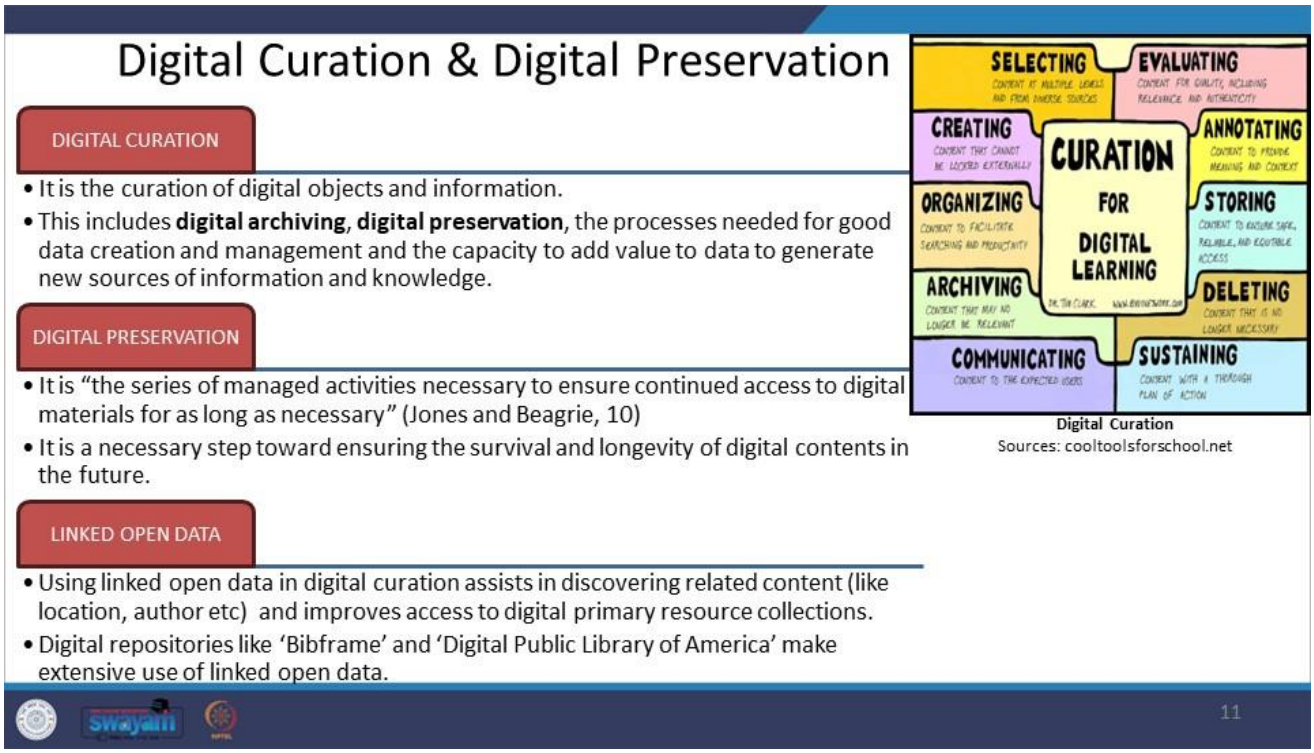
It shows a hypermedia environment, narratives and other visual assets moving between local and global histories in a richly interactive space. What all these projects have in common is an approach to knowledge production that underscores the distributed nature of digital scholarship, their hypermedia approach to argumentation and their open-ended participatory approach to interacting with media objects. This shows that digital humanities needs the development of a set of approaches to be engaged in the challenging combination of tasks such as creation, computation, curation, analysis etc. Let us look at some of them. So, how do we approach digital humanities?



Digital humanities is a two-way approach.

The systematic use of digital resources in humanities and secondly the humanistic analysis of the application of digital resources. This approach involves developing digital tools for data mining, hypertextualization, digital mapping and information retrieval. Data mining is the process of using computational techniques to extract meaningful patterns from large data sets related to human culture, history and in fact, any other aspect of the humanities. Hypertextualization is the process of transforming text or a collection of texts into a network of interconnected elements often through the use of hyperlinks. Digital mapping involves the use of digital tools and technologies to create, analyze and visualize spatial data.

Information retrieval involves using computational methods and techniques to efficiently locate and retrieve digital data or a specific information from large databases. These are just some of the approaches that would allow researchers to access vast amount of digital data in relation to cultural, historic, linguistic and literary aspects. With the advent of technology, researchers have realigned the methods required for the digital environment for managing primary resources and research based on them. Some of the main practices that combine traditional humanities with computational methods are digital archiving, digital curation and digital preservation. So, what exactly is the digital curation?



Digital Curation
Sources: cooltoolsforschool.net

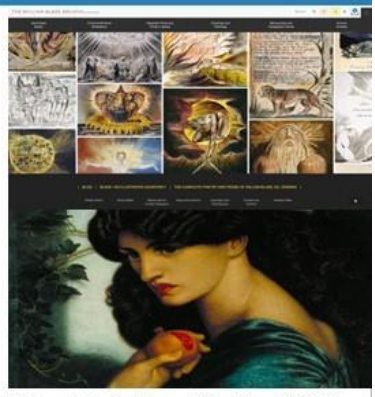
Digital curation is the curation of digital objects and information.

This includes digital archiving and digital preservation. The processes needed for good data creation and management and the capacity to add value to data to generate new sources of information and knowledge. Digital preservation is the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. Linked open data in digital curation assist in discovering related content and improves access to digital primary resource collection. Digital repositories like BibFrame and Digital Public Library of America make extensive use of linked open data.

In short, digital curation refers to the actions needed to maintain digital research data and other digital materials on a permanent basis. As we discussed, digital curation is often accompanied by digital archiving. They serve as repositories of primary source materials that can be analyzed, interpreted and used as a basis for research and scholarly inquiries.

The Role of Archives in Digital Humanities

- Digital Humanities provide access to digitized and curated collections of cultural, historical and literary materials.
- The role of digital archivists is to reinforce the semantic connection between the digital object and context.
- Trevor Owens in his article "Defining data for humanists: text, artifact, information or evidence?" (2011) recognizes that the meta-data in digital archives represent three dimensions of humanistic knowledge.
  - The data itself represents a digital object that may or may not be the digital reproduction of the physical original.
  - The data itself supports a new narrative.
  - The data itself may be useful in data visualizations for additional interpretations with applicable technologies.

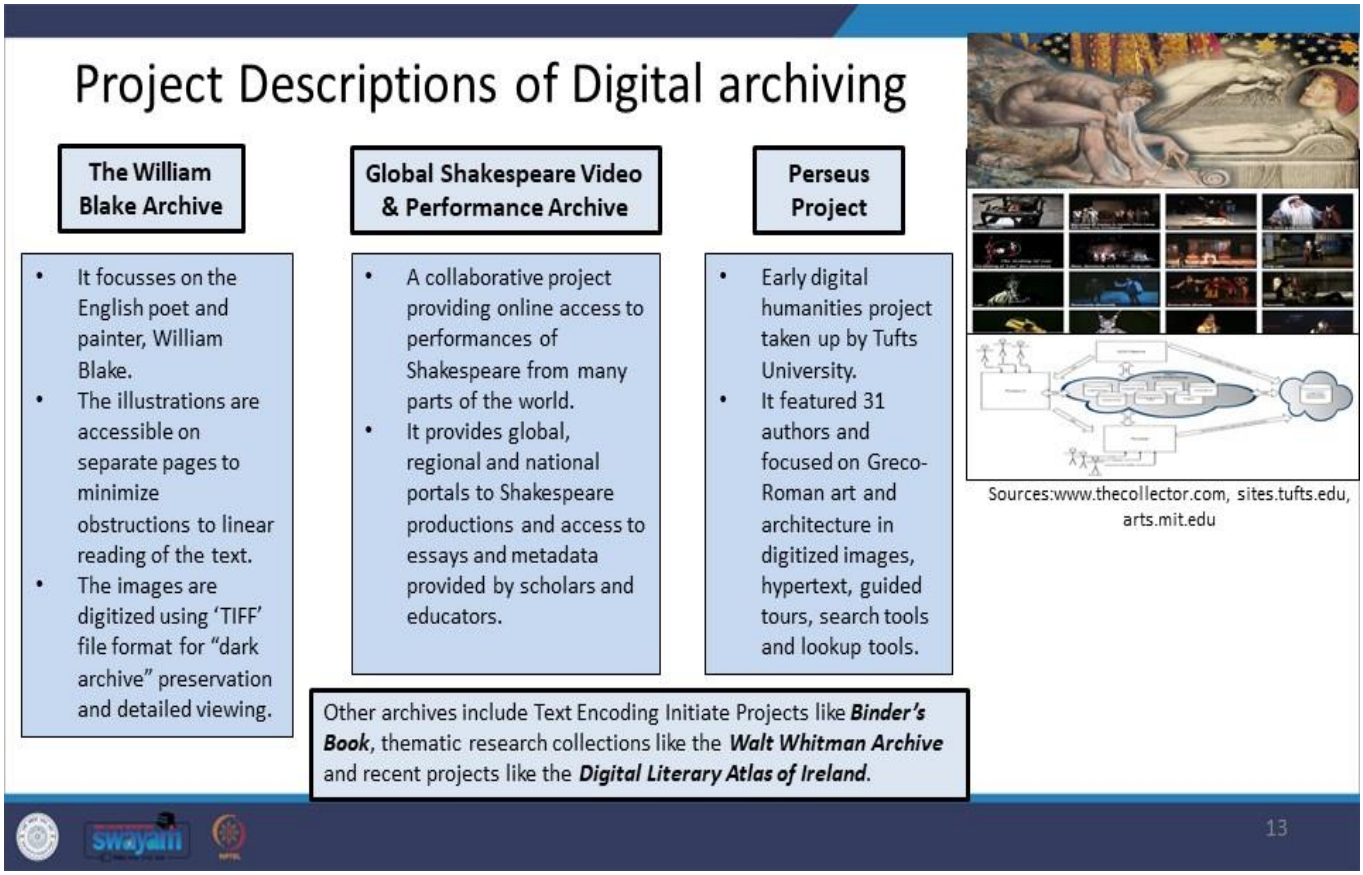


William Blake Archive and The Rossetti Archive
Sources:blakearchive.org, www.neh.gov

So, digital humanities provide access to digitized and curated collections of cultural, historical and literary materials. The role of digital archivist is to reinforce the semantic connection between the digital object and the context.

Trevor Owens in his article published in 2011 recognizes that the meta data in digital archives represents three dimensions of humanist technology. Firstly, the data itself represents a digital object that may or may not be the digital reproduction of the physical original. Secondly, the data itself supports a new narrative and thirdly, the data itself may be useful in data visualization for additional interpretations with applicable technologies. Digital archiving is equivalent to presenting an abstract world in the form of a digital object comprised of a combination of media, text and data. This provides a meaningful relationship between a digital item, say a document, image or video and the information surrounding it providing context that helps users understand its significance and relevance.

Let us look at some examples of projects that illustrate multiple scenarios for using digital archives.

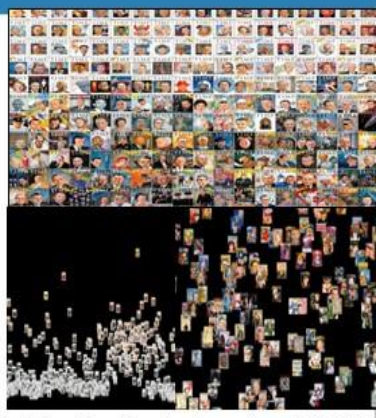


We can also refer to Global Shakespeare Video and Performance Archive which provides global, regional and national portals to Shakespeare productions and access to essays and meta data provided by scholars and educators. We can also refer to Perseus Project, an early digital humanities project taken up by Tufts University. It featured 31 authors and focused on Greco-Roman art and architecture in digitized images, hypertext, guided tools, search tools and look up tools. Other archives include text Encoding initiate projects like Binder's book, thematic research collections like the Walt Whitman Archive and recent Projects Like The Digital Literary Atlas Of Ireland.

Archivists and digital curators collaborate on the preservation of content and access to meta data which helps researchers. Author and theorist of digital culture Lev Manovich came up with the term cultural analytics. It refers to the computational analysis of patterns and trends in contemporary digital culture. The term extends its analysis to all large scale cultural data and visualizations. It incorporates all kinds of media as opposed to only texts and historical culture.

Cultural Analytics

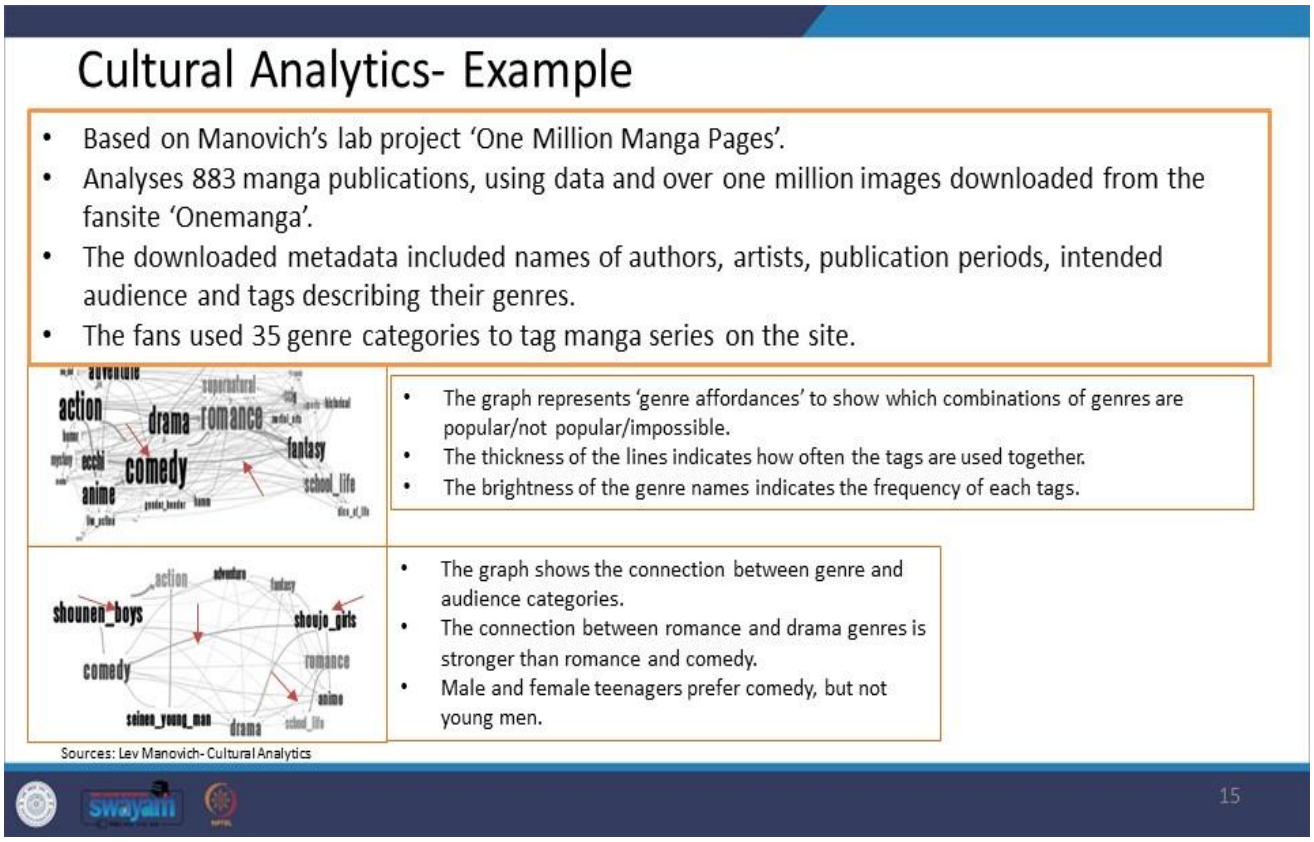
- Lev Manovich, in his book *Cultural Analytics* (2020), defines cultural analytics as the use of computational and design methods for exploration and analysis of contemporary culture at scale.
- The use of numerical representation, data analysis and visualization methods offers a new language for describing cultural artifacts, experiences and dynamics.
- The intention of cultural analytics is to augment our human abilities by providing new interfaces and techniques for observing massive cultural datasets and flows.
  - This includes **critical examination** of these methods and their assumptions.



Analyzing Time Magazine Covers from 1923-2009 using Cultural Analytics
Source: lab.culturalanalytics.info/

Lev Manovich in his 2020 book has defined cultural analytics as the use of computational and design methods for exploration and analysis of contemporary culture at a scale. The use of numerical representation, data analysis and visualization methods offers a new language for describing cultural artifacts, experiences and dynamics. The intention of cultural analytics is to augment our human abilities by providing new interfaces and techniques for observing massive cultural data sets and flows. This includes critical examination of these methods and their assumptions. The pictures given on the right hand side of this slide show a visualization analysis of Time magazine covers from 1923 to 2009.

It reveals a number of historical patterns like the shift in medium, color, hue, brightness and content etc. It also reveals an important meta pattern about the cultural history and presents that the new communication strategies emerge slowly over a number of years or perhaps even decades. Let us look at Manovich's own example of using quantitative analysis in the cultural category of genres of Japanese manga series.

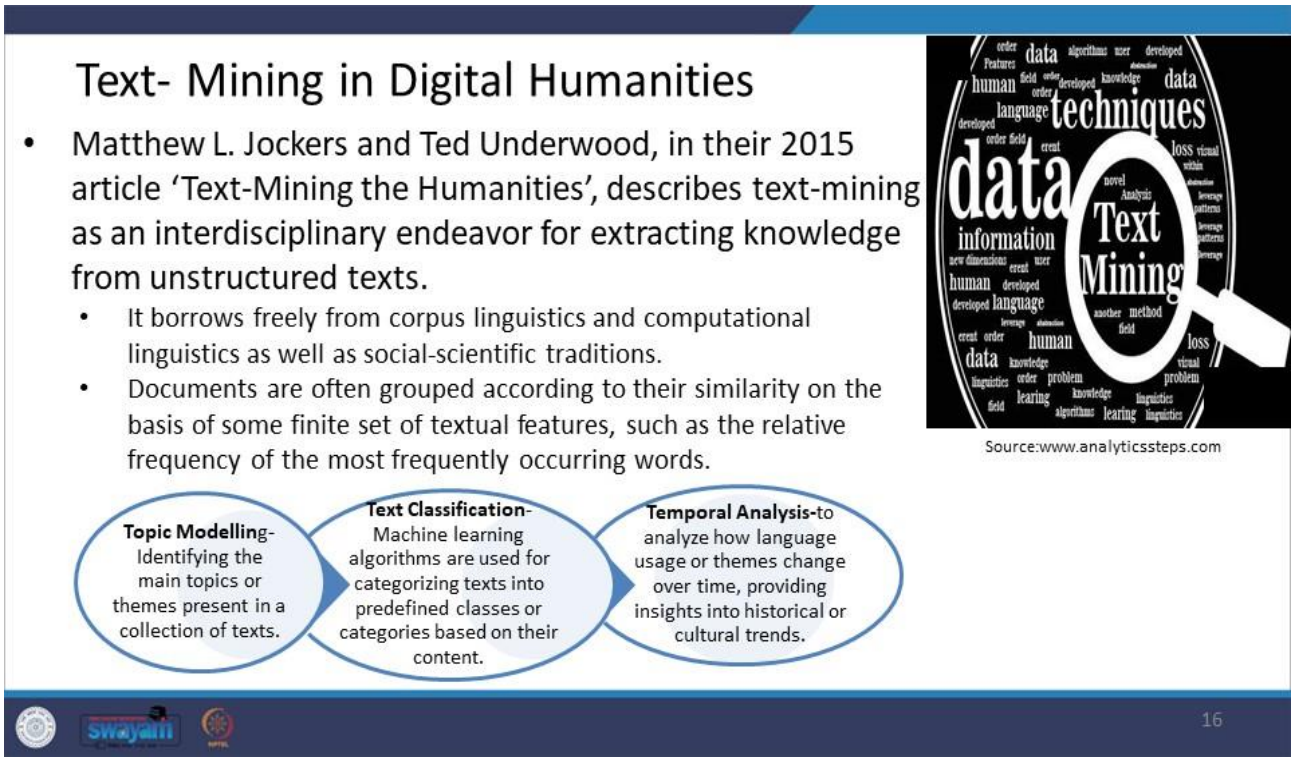


The lab project of Manovich, One Million Manga Pages, analyzes 883 manga publications using data and over 1 million images downloaded from the fansite, Onemanga. The downloaded metadata included names of authors, artists, publication periods, intended audience and tags describing their genres.

The fans used 35 genre categories to tag manga series on the site. The first graph represents genre affordances to show which combination of genres are popular, not popular or impossible. The thickness of the lines indicates how often the tags are used together. The brightness of the genre names indicates the frequency of each tags. The second graph shows the connection between genre and audience category.

The connection between romance and drama is stronger than romance and comedy. Male and female teenagers as per this graph prefer comedy but not young men. This analysis also shows that visual style in manga also constructs gender differences. The connecting lines show young men and boys prefer action while teenage girls prefer anime, romance and drama. Another important practice in digital humanities is text mining and visualization.

Subtle and nuanced details often revealed through close reading are now encoded by new and sophisticated tools for computerized textual research through text mining.

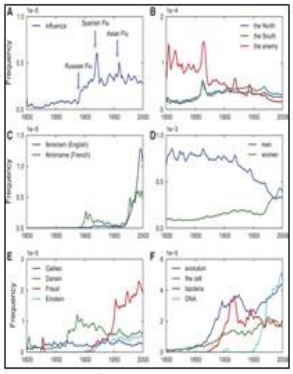


Matthew Jockers and Ted Underwood in their 2015 article describe text mining as an interdisciplinary endeavor for extracting knowledge from unstructured texts. It borrows freely from corpus linguistics and computational linguistics as well as social scientific traditions. Documents are often grouped according to their similarity on the basis of some finite set of textual features such as the relative frequency of the most frequently occurring words. So, we find that topic modeling leads to text classification which leads to temporal analysis.

By topic modeling, we refer to the identification of the main topics or themes which are present in a collection of texts. Machine learning algorithms are used for categorizing texts into predefined classes or categories based on their content when we talk about text classification. Temporal analysis is done to analyze how language usage or themes change over time providing insights into historical or cultural trends. Let us take the research approach of culturomics. An example of text mining to analyze a massive corpora of digitized text to gain insight into cultural trends, historical events and linguistic patterns.

Culturomics

- Culturomics* is a project undertaken by Jean-Baptiste Michel along with his colleagues, in conjunction with Google to study "culture" as it gets expressed in the Google Books Corpus.
  - It applies high-throughput data collection and analysis to the study of human culture in books, manuscripts, newspapers, magazines etc.



Source: www.science.org

- Trajectories of "the North", "the South", and finally "the enemy" reflect how polarization of the states preceded the descent into the Civil War (Fig B). In the battle of the sexes, the "women" are gaining ground on the "men" (Fig C). "Féminisme" made early inroads in France, but the US proved to be a more fertile environment in the long run (Fig. D). "Galileo", "Darwin", and "Einstein" may be well-known scientists, but "Freud" is more deeply ingrained in our collective subconscious (Fig E). Interest in "evolution" was waning when "DNA" came along (Fig F).

*Refer Jean Baptiste Michel- Quantitative Analysis of Culture Using Millions of Digitized Books

Culturomics is a project undertaken by Jean-Baptiste Michel along with his colleagues in conjunction with Google to study culture as it gets expressed in the Google Books corpus. It applies high-throughput data collection and analysis to the study of human culture in books, manuscripts, newspapers and magazines etc. The graphs given on the bottom left are an example of culturomics which suggest different aspects of data analysis. The table clearly shows how culturomics is guided by cultural phenomena such as feminism and also linguistic changes such as changes in lexicon and grammar. These results highlight how culturomics methods might complement existing historical approaches.

More examples like detecting censorship, the period of fame are discussed in Michel's article. Let us conclude today's module by talking about one more but an important practice in digital humanities that is text analysis and visualization.

Text Analysis and Visualization

- The spectrum of text visualizations are often used to present the results of text mining analyses, making patterns and relationships more apparent.
  - They include a variety of stylistic and positional transformations and abstract representations of textual attributes.
- Text analysis identifies patterns of word usage by particular authors, which is a field called stylistics.
  - Computers use algorithms not only to find patterns but also to count patterns and compare counts.
  - One can get a sense of author's writing style by counting function words, which do not convey a lot of semantic content, but which are important syntactically.
- A word cloud is a visual representation of text data in which words are displayed in varying sizes and colors.

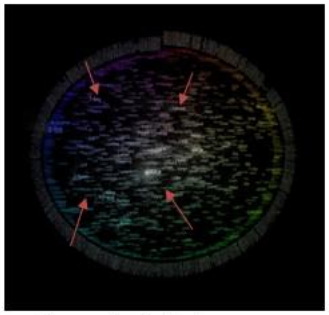
***Wordle* word cloud visualizations of vocabulary from toy commercials for boys and girls**
Sources: criticalmediaproject.org, crystalsmith.ca

Text visualizations often present the results of text mining analysis, making patterns and relationships more apparent. Text analysis identifies patterns of word usage by particular authors which is a field called stylistics. Computers use algorithms not only to find patterns but also to count patterns and compare counts.

A word cloud is a visual representation of text data in which words are displayed in varying sizes and colors. The size of each word in the pictures given corresponds to its frequency or importance within a given body of text. The frequency of words like fashion, love, party will be higher in toy commercials for girls rather than boys. This reinforces other semantic aspects like gender, age, culture, etc.

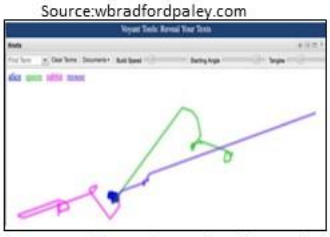
Let us look at two examples of text visualizations. One is Bradley Paley's TextArc and the other is the example of Knots interface of void tools.

Text Analysis and Visualization



Source:wbradfordpaley.com

- In Bradley Paley's **'TextArc'**, words from a text are displayed twice.
- The image shows words in *Alice in Wonderland* as text distributed across the perimeter and then again by plotting each content word within the circle, as if each occurrence in the perimeter pulled the terms toward it gravitationally.
- As a result, the location of the word conveys information about its distribution in the document.
- For example, "king" and "queen" occur more in the last third of *Alice in Wonderland*.



Source: Stefan Sinclair and Geoffrey Rockwell

- 'Knots' is an interface used by the web-based suited 'Voyant Tools' for text visualizations.
- It represents lexical repetition by introducing a kink in a line every time a selected term occurs.
- The more "knotted" a line, the greater the repetition.
- The image shows the 'knots' visualization of *Alice in Wonderland.*

19

In the first example, we find that Paley's TextArc shows words in Alice in Wonderland as text distributed across the perimeter and then again by plotting each content word within the circle as if each occurrence in the perimeter pull the terms towards it gravitationally. As a result, the location of the word conveys information about its distribution in the document. For example, king and queen occur more in the last third of Alice in Wonderland. In comparison, Knots represents lexical repetition by introducing a kink in a line every time a selected term occurs.

The more knotted a line, the greater the repetition. The image shows the Knots visualization of the same text, Alice in Wonderland. Alice which is the most frequently used word forms a knot near the middle while mouse shoots off to the right. Visualizations make use of a visual grammar just as language requires a linguistic grammar. We need to be able to resolve what we see before attempting to analyze and understand it. Considering how digital humanities deconstruct the authority of traditional textual analysis, it also becomes important to understand that it poses some significant challenges.

In conclusion, let us say that it remains a persistent fallacy to ascribe to the computer a capacity to reach beyond human particularities and into the realm of objectivity.

So, digital humanities which works at the intersection of technology and the humanities leverages computational tools and methodologies to analyze, interpret and engage with cultural, historical and literary materials in innovative ways. It harnesses the power of digital technology and through it scholars in this area are able to uncover new insights, facilitate collaborative research and democratize access to cultural heritage. This rapidly evolving field continues to shape the way we explore and understand the complexities of human culture and knowledge in the digital age. In the next module, we will further explore the realm of digital humanities by analyzing digital quantifications and visualizations of literary styles. Thank you.