Introduction to Digital Humanities
Rocío Ortuño Casanova. Universiteit Antwerpen
Rocio.ortuno@uantwerpen.be

Any questions may be addressed to my e-mail or even better to the discussion section in the Humanities Commons group DAGITAB https://hcommons.org/groups/dagitab/forum/

We are about to sit here and spend a few days talking about Digital Humanities (DH). In this introductory session, we are going to reflect about why we are going to do so and how DH can be useful for your own teaching and research and that of others.

Firstly, I would like to explain that this workshop makes part of a project to be developed along 4 years (2018-2021). It is funded by VLIRUOS and is being developed in partnership between the University of the Philippines Diliman and the University of Antwerp, in Belgium, although we intend to involve the whole UP system (or almost). The project has two parts:

- The first part focuses on the digitization of Philippine historical periodicals which are held at the University of the Philippines Diliman library. This part of the project is led by Chito Angeles, who will be talking about the digitization process and the repository that they are creating for the general public to be able to access it online.
- The second part consists of delivering a series of training sessions on digital humanities focusing specially in text analysis, corpus compilation and distant reading to be able to do MORE things (research and teaching-wise) than conventional, non-digital scholarship allows with those newspapers and other interesting materials, especially related to the Philippine history, society, languages etc.

Here you can find some information about the whole project:

- https://hosting.uantwerpen.be/philperiodicals/

So, the objectives of the summer course are:

Objectives of the summer course

- Introducing different approaches to your topic of study with Digital Humanities
- Encouraging research and publication with a DH component which adds value (not just because it looks pretty)
- Encouraging the Introduction of DH in your lessons (if it adds value to your lessons)

The idea behind the “added value” is important. Digital tools are fashionable, and they can be very useful, but sometimes they are used to achieve things that could just as well be done without them. In those cases, using digital tools does not bring any added value. For instance,
if you would like to find out the topics in a chapter of a book, you do not need to use topic modelling for that. You can do it just reading that chapter. If you would like to find topics in 100 books or more, you may want to use digital tools, as doing it without them would be longer and not so accurate. (That is: Maybe. In some cases.)

Now this has been clarified, let’s do a small quiz about DH to start explaining what this is all about.

**A Mini-Kahoot on Digital Humanities**

https://create.kahoot.it/share/dh-up/d00108bb-8116-43ba-9607-6c001598888a

*Tip for echoing this workshop:* If you are going to use this Kahoot in your classes or workshops on DH, in the link on the slide above you can find the “inners” of the Kahoot. From there, you can clone the quiz, modify it, or you can also choose “play as guest”. Then you will be able to log in with a Google account and to choose if you want your students to play as individuals or as a team.

Choose either one of the options, and you will find the instructions for your students to join the Kahoot with their phones. They just need to enter the page www.kahoot.it, and enter the
PIN indicated on your screen (that should be projected for the students to see it, and the questions and answers). For instance:

![Join at www.kahoot.it or with the Kahoot! app with Game PIN: 368052](image)

Now, after each question, students will have 20 seconds to answer. You can move down to see the results (who has answered correctly and who hasn’t) and the ranking of players by number of points (depending on their number of correct answers and their speed in answering the questions). Before proceeding to the next question, I would recommend stopping and explaining the answer. The explanation is on the slides:

**Question 1 on Kahoot:**

1. Quiz
   - Which one of these disciplines is NOT included under the term ‘Digital Humanities’?

   - Humanist Computing
   - Information technology
   - Computational linguistics
   - Gardening

**Explanation to answer number 1**

**Humanist computing and the history of DH**

- **Humanist computing** was the original term, modelled at the conference held in 1965 at Yale, entitled ‘Computers for the Humanities?’ and coined as trans-discipline in the first academic journal on this topic: Computers and the Humanities. This journal was founded by Joseph Raben in the U.S. a year later, and it suspended its publication in 2004.

Humanist computing or Humanities Computing is how Digital Humanities were called before their current name, but the contents and the ideas behind the name were then same as what we call today Digital Humanities.

**Explanation to answer number 2: Information technology**

<table>
<thead>
<tr>
<th>Media and DH</th>
</tr>
</thead>
</table>
| In the past, the digital humanities have only inchoately linked its three underlying concepts (for which ‘digital’ is only shorthand): technology, media, and information. Whether digital humanists privileged one or the other when they thought about their object of inquiry or their tools-- and especially if they split the difference in such hybrid formulae as media technology or information technology -- these concepts fused or split unpredictably, leaving such basic questions as the following unsettled. Are the textual, visual, or other phenomena that the digital humanities treat best understood as technology, media, or information? Is all technology mediational, or vice versa? Is all media informational, or vice versa?

Recently, however, the digital humanities have deepened the conceptual roots of all three cardinal concepts so that they begin to connect at an appropriately foundational level. [...] the trend is toward an integrated logic circuit, as it were, of technology, media, and information. Thus, for example, the digital humanities have recently begun to rationalize the connection between technology and media through such new approaches as platform studies, software studies, and digital forensics.

Alan Liu 'The State of the Digital Humanities'. June 4, 2016. Pp. 5-6. [https://ascholarship.org/uc/item/73h6v6x8](https://ascholarship.org/uc/item/73h6v6x8)

In layman terms: according to Alan Liu, “Digital” just means “technology + media + information”. You can favour one or another of the components according to your object of study and methodology. Although there has been some instability in the nomenclature for certain processes and methodologies, it seems that lately everything is getting more integrated and the tendency is to include information technology under the more inclusive umbrella of Digital Humanities (the idea of a “Big Tent”), or at least to walk towards symbiosis between both.

**Explanation to answer number 3: Computational Linguistics.**

<table>
<thead>
<tr>
<th>Computational Linguistics and DH</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH invites contribution from all Humanities disciplines, including those where language plays a secondary role, such as anthropology, archeology, fine (visual) arts, film studies, and musicology. These are not the most likely disciplines for computational linguists to get involved in, but linguistics and literature (studies) are also Humanities discipline, where language is central, as are history and philosophy, where language is not of central interest, but where archival material in textual form often plays a central role. There are enormous opportunities for contributions from computational linguistics (CL) from all the disciplines where language and text are important.</td>
</tr>
</tbody>
</table>

Nerbonne, John and Sara Tonelli. 'Introduction to the Special Issue on Digital Humanities and Computational Linguistics'. Nerbonne, J. & Tonelli, S., 2017, in : Italian Journal of Computational Linguistics,2, 2, p. 7-10. [https://pdfs.semanticscholar.org/5c5d/a4403fc64f81291e222ac49b2497be66c85.pdf](https://pdfs.semanticscholar.org/5c5d/a4403fc64f81291e222ac49b2497be66c85.pdf)
Linguistics, a discipline that typically falls within the Humanities, scholars have pioneered in the use of digital tools for their quantitative research in the late 20th century. Many methodologies from linguistic research – especially the domain of computational linguistics or natural language processing – with digital tools are being used in Literary studies nowadays and other fields such as History. In this way, we can understand the connection between Computational Linguistics, quite an old discipline actually, and Digital Humanities.

**Explanation to answer number 4: Gardening**

### Gardening

Ehhh... no. Gardening is not a discipline or a component within what we call Digital Humanities.

However, there are several other disciplines and activities that fall within Digital Humanities. You can learn more about this in these three links, for instance:

- [http://computerphilologie.uni-muenchen.de/jg02/unsworth.html](http://computerphilologie.uni-muenchen.de/jg02/unsworth.html)
- [https://cpb-eu2.wpmucdn.com/blogs.ucl.ac.uk/dist/e/218/files/2013/10/Chapter-6_EV.pdf](https://cpb-eu2.wpmucdn.com/blogs.ucl.ac.uk/dist/e/218/files/2013/10/Chapter-6_EV.pdf)

**Question 2 on Kahoot:**

Which of these activities is not a Digital Humanities activity?

- Finding common topics in a set of books with Mallet ✗
- Writing a document with Microsoft Word ✓
- Searching for the most used word in a set of documents ✗
- Mapping Ulysses’ route by marking places on the Odyssey ✗

This question involves a much more difficult and more controversial question in the history of Digital Humanities that is “What is Digital Humanities?”. Although we are not getting in depth
in this debate, which is time-consuming, we can start by proposing a minimal agreement and a definition:

### Consensus

Digital Humanities bring together humanities and computer science; thus we have two aspects:

- Work on *humanities* research question using methods and tools from computer science
- Work on *computer science* methods and tools for tackling research questions in the humanities

Therefore, any of the answers is alright, if using digital tools, except for just writing a document, as Word does not add anything to the writing itself content-wise. That is, you could actually write exactly the same without a computer and the data would be the same.

Regarding the mapping of Ulysses route... check this:

- [http://omeka.wellesley.edu/mappingmythology/](http://omeka.wellesley.edu/mappingmythology/)
- [https://whatisdigitalhumanities.com](https://whatisdigitalhumanities.com)

**Question 3 on Kahoot (multiple correct answers are possible):**
Explanation to answer number 1:

It enables the analysis of large amounts of data: Distant Reading.

The trouble with close reading (...) is that it necessarily depends on an extremely small canon. [...] you invest so much in individual texts only if you think that very few of them really matter. [...] And if you want to look beyond the canon (...) close reading will not do it. [...] We know how to read texts, now let’s learn how not to read them. Distant reading: where distance [...] is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes --or genres and systems. And if, between the very small and the very large, the text itself disappears, well, it is one of those cases when one can justifiably say, less is more.


Distant Reading is one of the main theoretical frameworks for the use of digital tools. It means that we can extract data from texts even without reading them. What Franco Moretti¹ argues in the quote that you have above is that we often tend to characterize a literature, an epoque, or a trend just by reading a few canonical texts. Now, the question is: canonical for whom? Literary history has been relaying on a selection done with a certain bias (aesthetic, political, social, religious, you name it). Against the question of canon there is the possibility nowadays of taking loads of books and extracting information without reading them. This might sound like a pity, but it gives us some other kind of interesting information. There are many ways of extracting information from big amounts of data (or texts). Here you can find some more information on distant reading:

¹ NB: because of the controversial accusations of Moretti in the States, his work is being cited less and less: https://www.stanforddaily.com/2017/11/16/harassment-assault-allegations-against-moretti-span-three-campuses/
http://www.digitalhumanities.org/dhq/vol/8/1/000171/000171.html

https://www.mat.ucsb.edu/~g.legrady/academic/courses/09w259/Moretti_graphs.pdf

- Distant Reading explained in layman’s terms:

An example of this distant reading is a work that Moretti did on *Hamlet*. You can find the graph of interaction of the characters below, and an explanation on it in this link
https://elenadigi.wordpress.com/2015/05/19/distant-reading-vs-close-reading/

---

**Explanation to answer number 2:**

**It makes sharing and re-using data easy**

---

In my opinion and experience, I have found Digital Humanists to be a community of practice in which sharing, and collaborating are highly regarded, unlike many other academic fields. Usually, the humanist’s work is quite solitary and highly theoretical\(^2\): you go to the library or the archive, you get your raw materials/texts/data. You make sense of that data by organizing it and relating it to other texts or materials, and then you publish your conclusions on the whole stuff. The workflow in the Digital Humanities has a similar path, however, there are two important differences:

1. Projects tend to be more ambitious (and often more multidisciplinary). Therefore, more people, with complementary backgrounds and expertise) are needed. They may involve different disciplines, address larger research questions and need more complex team work. The members of the team may be involved only in one component or the whole process: they can be organizing data from texts, for instance, or making sense out of it, or testing that data in different ways.

2. Once the data has been extracted, organized and prepared for being examined, it can be shared for other researchers to apply different kinds of methods to that data. Even more interestingly, those “tools” created to approach the data (digital tools), let us say, some app, logarithm, some piece of programming, is also usually shared to be applied by different researchers to different data. I have two examples for you:
   a. The first one is Stylo package, about which Mike Kestemont, one of the creators, will talk in the next few days
      https://sites.google.com/site/computationalstylistics/stylo. That is a “tool” to be used in ‘R’ for finding out about a text’s authorship and writing style.
   b. The second one is the Textbox of CLIGS, a research group in Germany working on some shorts of distant reading of literary texts: https://github.com/cligs/textbox In their Github repository (a very popular platform to share and develop chunks of code, materials, information about projects, and all sort of things to share) they have uploaded the texts they are working with in different formats, after having OCRd and “cleaned” those texts. Some of them are also tagged and annotated. You may want to use them for performing some sort of analysis or for having a corpus to compare with your own, or training tools... just by curiosity, you might want to have a look at other Github repositories like Mike’s http://mikekestemont.github.io/ or Enrique’s https://emanjavacas.github.io/ (these two are a bit more fancy: they are repositories within their own websites created with github also).

All this has to do with some sort of set of values that are important to practitioners and a kind of identity imprint for the discipline. These values have been discussed and developed in a book series called Debits in the Digital Humanities. You can see the link to the whole article at the bottom of this coming slide:

Now, the last two answers are quite wrong. Firstly, because no machine does your job. Your job is the Human part in Digital Humanities: you need to make sense out of the data, explain it in context, reach conclusions. Saying that computers are doing the whole research job would be like saying that microscopes and test tubes are doing the whole research job for biologists. The last question is wrong because well: fashion comes and goes and is all about perception. Much ado about nothing. Do not invest in fashion for the long term.

Now, I have an extra question out of the Kahoot for you all. Let’s get local and brainstorm a bit:

Why using DH in the Philippines and/or to work on the Philippines is a good idea

I am talking about this as someone who has been working on the Philippines for a few years, and most of those years not being presently in The Philippines. So, no magical recipes, just my experience.

The first thing that caught my attention when approaching Filipino Literature in Spanish was the fact that there was not much literature about it. People working on postcolonial studies would rarely refer to the Philippines, and it was totally out of the circuit of studies on Literature in Spanish. Wondering about the reasons for this, I realised that people from abroad
(Spanish speakers) had difficulties to reach the texts. I myself could not access the literary texts I was interested in when working from England. That can also happen if you are working from Davao, or from Iloilo, or from Batanes: most of the materials are gathered in a few archives and libraries in Metro Manila. The second reason was that, even for Filipino researchers it was difficult to access and analyse those texts because they could not understand them anymore. Nobody speaks Spanish in the Philippines, right? And there are so many rich literary traditions in other languages that it is not really a concern. So, given these problems, I thought that digitization (as a first step) may contribute to:

### Availability and diffusion of Philippine works

- It can contribute to the diffusion of primary sources (texts) along the archipelago and to preserve them from ambient conditions through digitization.
- It can help to the intercomprehension of historical and cultural heritage in the Philippines in its different languages through automated transcription and translation of materials.
- It could encourage updated research on the Philippines due to the availability of primary sources and the implementation of distant Reading techniques that might facilitate a global vision of the archive, in different languages simultaneously. It might lead to challenge preconceptions and previous limited research provoked by the study of materials in just one language at time.

Regarding the research using digital tools, some ideas that came to my mind were:

### Production and visibility of research

- It should promote visibility of Philippine studies and Philippine researchers abroad. In this way, alternatives to discourses on the Philippines built from abroad would get available
- It would enable the inclusion of Philippine studies and texts in comparative and transnational projects i.e. WeChangEd (Agents of Change – Women Editors and Socio-Cultural Transformation in Europe, 1710-1920), MapModern (MapModern).
- It can facilitate resources to increase research output (IPA)

Can you think of other answers?

**Question 4 on Kahoot (multiple correct answers are possible):**
This tricky question just aims to show you a few projects related to Digital Humanities in different ways and from different fields of knowledge, that may give you a better idea of what this is all about. What kind of questions you can answer using digital tools and the kind of materials that you can analyze. I have prepared a summary of each project, but you can also consult the website and the output papers produced by entering the links at the bottom of the slides.

Some successful examples of DH projects

Explanation to answer number 1:

**Periodizing Samuel Beckett’s Works A Stylochronometric Approach**

- Dirk Van Hulle and Mike Kestemont
- **What?**
  - A quantitative study of Samuel Beckett's writing style, focusing on grammatical function words.
- **How?**
  - "Stylochronometry" (Starnou): a text's writing style is studied as a function of its date of composition.
- **Findings:**
  - The analyses nuance the traditional three part periodization of Beckett's œuvre (as they show the possibility of stylistic relapses, "early", "middle" and "late" periods) disturbing the linearity of most periodizations) as well as different turning points depending on the language of the corpus, suggesting that Beckett's English œuvre is not identical to his French œuvre in terms of patterns of stylistic development.


Explanation to answer number 2:
The Women Writers Project

- **What:**
  - to bring texts by pre-Victorian women writers out of the archive and make them accessible to a wide audience of teachers, students, scholars, and the general reader.

- **How?**
  - Transcriptions are encoded in XML, following the specifications of the Text Encoding Initiative (TEI), with documented TEI extensions to accommodate the needs of our particular corpus and approach.

- **Findings:**
  - Good practices such as “Building Otherwise: Gender, Race, and Difference in the Digital Humanities.”

[https://www.northeastern.edu/](https://www.northeastern.edu/)

Explanation to answer number 3:

Uncovering gender bias in newspaper coverage of Irish politicians using machine learning

- **What?**
  - Text-analytic approach to analysing media content for evidence of gender bias.

- **How?**
  - Irish newspaper content is examined using machine learning and natural language processing techniques. Systematic differences in the coverage of male and female politicians are uncovered.

- **Findings:**
  - Evidence of gender bias in how female politicians were portrayed, the policies they were associated with, and how they were evaluated.
  - A methodology whereby natural language processing and machine learning can be used to identify gender bias in media coverage of politicians.

[https://doi.org/10.1093/llc/fqy005](https://doi.org/10.1093/llc/fqy005)

Explanation to answer number 4:
Paul Robeson’s place in YouTube: A social spatial network analysis of digital heritage

• What?
  • This article presents a spatial component to the technique of analyzing YouTube social networks, presenting specific spatial data, which can be mapped and analyzed.

• How?
  • Utilizing values of connectedness among videos of different topical clusters, as well as audience reactions to videos of a specific topic or place

• Findings:
  • Digital places are embedded with meaning. His social networks reveal an equal level of complexity, meaning, politicizing, and spatial characteristics to his traditional memorial landscapes. Robeson is remembered primarily as a musician and secondarily as an internationalist. Robeson’s activist videos, however, are more significant for exposing new memories when looking at his entire presence on YouTube as a social spatial network.
  • While the focus in this study has been Robeson, replicating this work for any other would open similar avenues of understanding

https://doi.org/10.1093/jlic/fqy027

Not everything is so beautiful, and therefore, it might be useful to have a look at why Digital Humanities is also criticized (heavily) from some sectors of academia. There are some ideas about the criticism that there is around Digital Humanities nowadays in this article: https://mkirschenbaum.files.wordpress.com/2014/04/dhterriblethingskirschenbaum.pdf and in this one https://www.journals.uchicago.edu/doi/abs/10.1086/702594

Beyond this, there are also debates on different aspects of Digital Humanities, which leads us to question number 5, the last question:

Question 5 on Kahoot (multiple correct answers are possible):

![](5-Quiz.png)

Explanation to answer number 1:
This is a question that was the title of one of the papers of that series of debates in the Digital Humanities. The author, Tara McPherson, from Minnesota, addresses what she thinks that is an internal division within the discipline: The Digital side, and the Humanities side. She feels that according to her experience, DH practitioners tend to be more aligned to one of these sides. Se also thinks that those more aligned on the side of the “digital” tend to suspect from those more on the humanities side, and vice versa. She advocates for closing tight the gap “from diaspora to database, from oppression to ontology, from visual studies to visualizations”.

From decolonial studies there is an even more complex debate that involves the centres of production of digital tools and the centres of training in Digital methodologies (normally in the so-called North countries). How Digital humanities are expensive, and therefore, not so democratic as the intend to be. And how Digital Humanities being created and developed and taught from the North, deal with northern concerns and give little space for other realities to bring their own questions and answers into it.

**Explanation to answer number 2:**
Why DH do always sound promising but never seem to reach their peak?

- The Eternal September of Digital Humanities
- What’s next: the Radical Unrealized Potential of Digital Humanities
  [http://dhdebates.gc.cuny.edu/debates/text/54](http://dhdebates.gc.cuny.edu/debates/text/54)

The Eternal September of Digital Humanities refers to the fact that there are always new practitioners who challenge the marked paths, and continuous new beginnings in the discipline. But (and these are my words) there is also a problem of repetition: people are creating tools for doing things that previous tools already did, wondering about questions that others already answered. Being not such a new discipline, a wider job of creating a state of the art and not thinking that you are a pioneer might be useful to start with it.

**Explanation to answer number 3:**

Are Digital Humanities a way to democratize knowledge?

- Hastac: Democratizing knowledge
  [https://www.hastac.org/initiatives/hastac-scholars/scholars-forums/democratizing-knowledge](https://www.hastac.org/initiatives/hastac-scholars/scholars-forums/democratizing-knowledge)
- Globalization, Open Access, and the Democratization of Knowledge
  [https://er.educause.edu/articles/2017/7/globalization-open-access-and-the-democratization-of-knowledge](https://er.educause.edu/articles/2017/7/globalization-open-access-and-the-democratization-of-knowledge)

“in the context of the digital humanities, democratizing knowledge often refers to making scholarship public, to opening access to university resources and research through, for example, the creation and preservation of digital archives and journals.”


Indeed, the digital allows ways of breaking academic barriers and even of being able to transfer knowledge from academia to a wider public. It also has to do with the “value of sharing” that we were discussing in question number 2 and the “values” of DH.

**Explanation to answer number 4:**
Ehhh... the slide is quite self-explanatory, I think. Although I would think that digitization is green. But not so much DH research.

These are some of the topics being discussed around DH, but not the only ones. Other topics can be found in these open access books and articles:

**Debates in the DH (other topics)**

- What are and what are not Digital Humanities: definition and limitations
  

- What does Digital Humanities bring to the table
  

- Others: the series “Debates in the Digital Humanities”
  
  http://dhdebates.gc.cuny.edu/

Now, if those were all the questions we intended to answer in this introduction, it might be useful to come down to the local again and, from the mindset of the presentation, wondering about Digital Humanities in Asia. We have some initiatives. The focus is moving from the US/Canada, Australia and Europe, which were probably the main three foci of DH work. Regarding to Asia, I found some initiatives:
As I said, Digital Humanities are expensive, and the initiatives that I found are taking place in richer countries. So, besides Singapore,

Where is Southeast Asia in Digital Humanities on Southeast Asia?

Well, there are also some initiatives in the Philippines, actually. Although Mr Chito Angeles will be talking about this more in depth, here is a teaser of some of the projects going on, that you can search on their websites if you are interested in them (just search them on Google):
And finally, some bibliography:

**Bibliography and resources**

- [http://computerphilologie.uni-muenchen.de/jg02/unsworth.html](http://computerphilologie.uni-muenchen.de/jg02/unsworth.html)
- [https://mkirschenbaum.files.wordpress.com/2014/04/dhterriblethingmkirschenbaum.pdf](https://mkirschenbaum.files.wordpress.com/2014/04/dhterriblethingmkirschenbaum.pdf)
- [http://www.digitalhumanities.org/companion/](http://www.digitalhumanities.org/companion/)
- [https://blogs.ucl.ac.uk/definingdh/sample-chapter/](https://blogs.ucl.ac.uk/definingdh/sample-chapter/)

Maraming salamat po.