DHC Weekly 5/24: Wikipedia Cross-Lingual Image Analysis

Hi DH fans! I have a question for you — when you think about information on the Internet, what benchmarks might come to mind that would reveal a certain topic as having thorough, accessible information online?

I, for one, immediately think about Wikipedia. I'm not alone in this — here at the DHC we've had multiple Wikipedia edit-a-thon's focused on topics like women's reproductive health and *Algorithms of Oppression*, aimed at growing the representation of certain topics and identities on what is now undoubtedly the world's most-used encyclopedia. Another way of thinking about information online, however, has to do with the content of those Wikipedia pages themselves — how long are they? How many sources do they cite? How many sections do they comprise? How densly do they inter-link with other, related pages? And, althought you might not immediately flag it as a representative figure, there's also how many images the page has. Enter the Wikipedia Cross-Lingual Image Analysis tool, a tool that looks at Wikipedia pages across languages and compares how many and which images a given topic has in any of the languages it has a page in.

When you enter the url of a Wikipedia page into the tool, it gives you a list of languages with links out to their respective pages, and a set of thumbnails of the images embedded in each page. To explore the tool, I looked up the wikipedia page of the early 20th century German film star Marlene Dietrich — I was curious whether her page would be more thoroughly built up and image-d out in German or English, the two languages in which she made films. Sure enough, while the majority of languages had two or three images on her page, German and English were among the most densly image-populated — the German page has 12 images and the English page 19. Not entirely surprising, given that the super-star phase of Dietrich's career kicked off once she got to Hollywood. What did surprise me, however, is that the French page had them both beat cold, with 34 pictures uploaded.

Obviously, my example is a bit of a silly one, but one can imagine the use of a tool such as this in research that considers the differing biases and perspectives and therefore differing avaiablity of information on any given subject across cultural and national lines. Give it a shot!

DHC Weekly 5/17: Terms of Service; Didn't Read

Hi DH fans! This week on the blog I want to draw your attention to a resource that is less about research and more about every day safety and security as we navigate the shark-infested waters of "the Internet."

Terms of Service; Didn't Read is a browser extension that assesses and rates the user agreements that many of us automatically agree to without a second thought, for the sake of convenience. If you install TOS;DR, a little button offering a letter grade starts to pop up in the top right corner of websites you visit.

For example, when I open privacy-prioritizing search engine DuckDuckGo, a little green "A". When I open Google, a cautious orange "C". When I click on these letters, I get a little summary of the points in their respective terms of service that led TOS;DR to assign the grade, as well as, of course, a link to the terms of service in full, so I can see for myself.

For DuckDuckGo, here is TOS;DR's assesment of what I'm agreeing to when I use their search engine, copied directly from the drop-down write-up associated with the A grade, which it describes as meaning "the terms of service treat you fairly, respect your rights and follows the best practices":

No Tracking: *DuckDuckGo doesn't save your searches and doesn't send your searches to other sites. No personal information is saved either. No cookies are used by default, but cookies can be saved on your computer for some features (e.g. settings). "We also save searches, but again, not in a personally identifiable way, as we do not store IP addresses or unique User agent strings. We use aggregate, non-personal search data to improve things like misspellings."*

Conditions may change, but your continued acceptance is not inferred from an earlier acceptance flow: *I think this is the correct case: although they mark updates, they do not contact the user. (Which makes sense because they don't have contact details, but still.) Also, the changes take effect immediately.*

There is a date of the last update of the terms

The cookies used by this service do not contain information that would personally identify you.

By contrast, the write-up for Google's C grade, which is described as meaning "The terms of service are okay but some issues need your consideration," is significantly longer, detailing the many issues with Google's privacy policies. Some selected information includes: "Your identity is used in ads that are shown to other users," "this service may collect, use, and share location data," "this service can read your private messages," "Google can use your content for all their existing and future services," but also "you can request access and deletion of personal data," as a positive.

If you were to read the terms of service for these sites in full, you would likely get tired out or bogged down in legalize and pages and pages of text fairly quickly. And yet, this information is absolutely vital to anyone interacting with these sites and services. A tool like TOS;DR goes a long way towards improving our literacy and preparedness when it comes to the use, storage, and exploitation of our personal data and information by big tech companies like Google or Facebook.

DHC Weekly 5/13: twXploder!

MAY 1, 2019
CATEGORIES: DHC WEEKLY, GUIDES AND DEMOS!, SYLVIA KORMAN
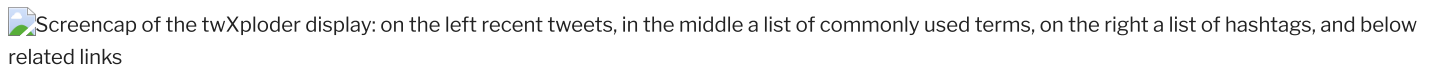**TAGS: GUIDES, TWXPLODER**

Hi DH Superfans!

We've been talking about some text analysis tools lately here on the blog, and this week I'd like to turn to a tool that allows for some light-weight and accessible analysis of the vast and unknowable text-based dataset that is twitter.com!

twXploder is a tool from Knightlab (the folks brought us TimelineJS) that allows you to search a word or phrase and get an overview of what the conversation around your search term looks like on twitter, showing you the most commonly used phrases and hashtags associated with the term, as well as links being shared around it. To get started on twXploder, you have to log in with your twitter account, and then you can pick your language and search for your term!

I searched "humanities." Here's what my results looked like:

Screencap of the twXploder display: on the left recent tweets, in the middle a list of commonly used terms, on the right a list of hashtags, and below related links

As you can see, on the left is a list of recent tweets with my search term in them — I can scroll through these and more will load. In the middle is a list of terms that are also turning up in tweets that contain my searched phrase — with humanities, unsurprisingly, I'm seeing "students," "majors," "grads," etc. I can click on these terms and filter my results to tweets that contain, for example, both "humanities" and "students". I can also click on a hashtag form the list on the right. The links section down below lists links that have been shared in tweets containing "humanities."

twXploder's functions — as you can see — are pretty lightweight and limited, but nevertheless it is a useful tool for capturing easily-digestible snapshots of the conversations ongoing on twitter, and a useful and accessible introductory tool to the work of social media analysis.

DHC Weekly 5/3: Part of Speech Tagger

MAY 5, 2019
CATEGORIES: DHC WEEKLY, GUIDES AND DEMOS!, SYLVIA KORMAN
TAGS: CLAWS, GUIDES

Continuing with the theme of text analysis, this week I want to go hyper-granular and a little old school to talk about a part-of-speech tagging tool, CLAWS.

A part-of-speech tagger is exactly what it sounds like — a tool that will take a piece of text, and auto-tag each word in it with its part of speech. POS tagging predates more robust and fancy tools like Voyant as one of the earliest and most common forms of corpus annotations and, indeed, of computational humanities, as the field was then called. Computational corpus analysis has been being undertaken since at least the 1960's, but the CLAWS POS-tagging software dates its earliest iterations to the 1980's. You can read all about how, grammatically and computationally, CLAWS functions here.

As an example text, I turned to our old friend *The Adventures of Sherlock Holmes* and plugged in the first story, "A Scandal in Bohemia." A (very very small) section of my output looked like this:

```
0000002 080 To            00 PRP
0000002 090 Sherlock      00 NP0
0000002 100 Holmes        00 NP0
0000002 110 she           00 PNP
0000002 120 is            00 VBZ
0000002 130 always        00 AV0
0000002 140 THE           00 AT0
0000002 150 woman         00 NN1
```

Here is the first sentence of "Scandal," word by word, tagged by part of speech according to this tagset. So: preposition, proper noun, proper noun, personal pronoun, -s form of the verb "be", adverb, article, singular noun. This, of course, extends all the way through the end of the story.

So, what use might one find from this sort of POS-tagging? One example is that once you've tagged all your words it becomes much easier to pull them apart from one another and look at them as groups — one could imagine, for example, a line of inquiry along the lines of how active vs. passive the verbs in a text are, or how many proper nouns are in a text as a measure of its specifity of place and person.

DHC Weekly 4/26: Google Ngram…friend or foe?

Last week, we looked at a set of corpora that in part allow one to ask questions about language's changes over time — today I'm going to talk about a similar tool, the Google N-gram Viewer.

The N-gram Viewer has been around for a while, and not without controversy, which I hope to lay out for anyone interested in using it. The N-gram Viewer allows you to search for a word or a phrase across an extremely large corpus of books across 400 years of publishing history — something like 4% of all books published, as Google tells it. The result is rather like the "chart" function in Davies' corpora — a graph of a word's popularity over four centuries.

Theoretically, there should be a lot to learn here about the popularity of certain terms and concepts over time. Like my example of tobacco's rise within early modern texts as it rose in dominance as a crop, the frequency with which a term occurs within printed discourse is a reasonably good indicator of its cultural importance. When you open up the N-gram Viewer, it helpfully supplies the search results for three culturally iconic figures: Frankenstein, Sherlock Holmes, and Albert Einstein, as if to invite you to stage little cross-century popularity races.

There are issues, however, with taking the N-gram Viewer at its word. Unlike Davies' corpora, the N-gram Viewer does not allow you to access the source texts of its results. Therefore, there is no way to see the context in which the searched word or phrase turned up, and no way to assess whether or not the data is useful or applicable. The type of text a word appears in is important, as is having a sense of what kinds of books make up the corpus, and whether their demographics, too, change over time. If, for example, the early English books in the corpus are all philosophical in nature, and the more contemporary books have more scientific texts, as scientific discourses and disciplines have proliferated in the contemporary era, then the results for a word like "courage," unlikely to be used in science writing, may trend distressingly downwards over time but may not be a reflection of a culture that increasingly has no use for the brave so much as an inconsistent set of data skewing results.

I am also somewhat skeptical of the implication that sheer instances of a word in print is at all indicative of its cultural capital, especially across 4 centuries, and especially when limited to a corpus of books. Some people and things are written about extensively in newspapers, but less so in books. Some might be written about extensively and in large numbers but by a relatively small, super-engaged constituency.

Google N-gram Viewer is certainly more extensive in its corpus than the corpora I wrote about last week, but with its millions of texts come some tradeoffs, that I suspect makes the tool less than useful for anything other than putting in things you're mildly curious about. Like hey, check it out –

Sylvia's on the rise again!

DHC Weekly 4/19: Corpora Works of Mercy

Last week on the blog, I wrote about Voyant, a text anaysis tool that can be used to discover all sorts of stats about a text or a corpus of texts — what words are used most frequently, in what combinations, in what contexts, and so on. I used The Adventures of Sherlock Holmes as my test corpus, because its consistent tone and easily accessible public domain status makes it an ideal example for the sorts of questions textual analysis tools like Voyant can prompt one to ask of a literary text. But what other sorts of corpora are out there, and what sorts of projects does analyzing them lead to? Today, I want to write about a publically accessible collection of English language corpora amassed by Mark Davies, a professor of linguistics at Brigham Young University.

Davies' corpora are collections of English language texts that range from all of the texts of the Early English Books Online database to a corpus of over 25,000 screenplays from films of the 1930's to the present day. Select of these corpora are fully downloadable, and all of them are queryable along similar lines as Voyant. Terms can be searched through five different engines: list, chart, collocates, compare, and KWIC (keyword in context).

"List" is a simple search function that allows you to find single words, all forms or tenses of a words, and so on. "Chart" allows you to search across time, turning up, for example, the fact that instances of the word "tobacco" in the EEBO skyrocket in the early 1600's as the crop began to globally dominate.
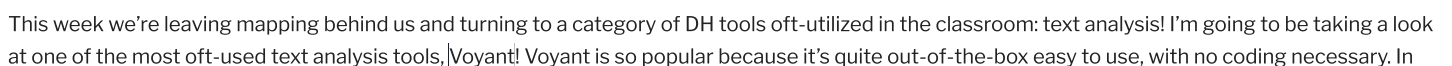
Through the "collocates" function you can see what words occur near other words, and through "compare", you can compare the collocations of two similar terms, revealing, for example, the contextual differences between early modern uses of "small" and "little".

"Keyword in context" I truly do not understand, but here's what the site has to say about how it works:

KWIC (Keyword in Context) display
See the patterns in which a word occurs, by sorting the words to the left and/or right. For example: run (v), put (v), lack, sense (n), goodly, or apt.

Note that you can also limit by section to see how a word is being used in different historical periods, e.g. lust (early 1500s, late 1600s); cunning (early 1500s, late 1600s); stuff (early 1500s, late 1600s).

How to do it:

Select the words that you want to sort with. Select **L** for 1, 2, and 3 words to the **l**eft. Select **R** for 1, 2, and 3 words to the **r**ight. You could also, for example, sort by one word to the left, then one and two words to the right. Click * to clear the entries and start over.

Your guess is as good as mine.

The uses of tools like these are many, and exist across Humanities disciplines. As a reader of, say, a Renaissance or early modern text, for example, I could query the EEBO corpus to understand subtleties of tone and connotation when it comes to unfamiliar words, or words that in a contemporary context no longer mean exactly what they meant in the early modern era. As a linguist or a historian of film or of modern culture, I might be interested in the way language has changed since the 1930's, as reflected in the changes in screenplay vocabulary. 9376 users have registered as "researchers" with the website — to get inspired, or learn more about what sorts of projects people are using these corpora towards, why not check out some of their profiles?

DHC Weekly 4/12: Voyant and Text Analysis

Hello DH fans!

This week we're leaving mapping behind us and turning to a category of DH tools oft-utilized in the classroom: text analysis! I'm going to be taking a look at one of the most oft-used text analysis tools, Voyant! Voyant is so popular because it's quite out-of-the-box easy to use, with no coding necessary. In

practice, I have found this to mean that Voyant is a little idiosyncratic and difficult — but I'm going to try to break down its basics for you all this week!

Voyant allows you to upload a corpus of text (it comes pre-loaded with the complete works of William Shakespeare and those of Jane Austen) and analyze various things about it, mostly related to word frequency and order. To play with, I downloaded a .txt file of The Adventures of Sherlock Holmesfrom Project Gutenberg and uploaded it as my corpus. Right away, with no customization added, Voyant gives me a couple views of my corpus text. First, it plainly gives me some facts and figures about my corpus, namely that it has "108,612 total words and 8,307 unique word forms," a Vocabulary Density of 0.076 and and Average Words Per Sentence of 16.2 The most frequent words in the corpus are said (486); holmes (467); man (291); mr (275); little (269).

On the subject of word frequency, there's also my cirrus:

A multi-color wordcloud with words from Sherlock Holmes

This is a word cloud, with the most frequently repeated terms from your corpus, arranged with a greater size implying a higher frequency. My Sherlock Holmes corpus yielded unsurprising key terms, like "Holmes," "Mr.," "man," "said," and "think," as well as words that are fairly common but that, even as a devoted, life-long Holmes fan (my college common app essay was about Sherlock Holmes – please do not make fun of me, I was 16 and really ticked off about BBC Sherlock) I don't know if I would have guessed were terms that were so central to Sherlock Holmes. "Door," for example, "room," "window," and "house," all appear in the cirrus, perhaps in an example of the sort of spatialization of plot in detective fiction that Peter Brooks (in his book Reading for the Plot) lays out in a reading of "The Musgrave Ritual". Body parts are another category that crop up here – "hand," "hands," "face," "head" – in a way that is not unlike the stories themselves paying the same sort of attention to the same markers of trade and personality that Holmes himself tracks in order to perform his sleuthing.

Even just from the basic, un-altered word cloud, I have two inferences about *The Adventures of Sherlock Holmes* that are perhaps worth pursuing! But let's take a look at some other Voyant features. There's the trends feature, which in this case is not very useful, since *The Adventures of Sherlock Holmes* is a set of short stories, whose bounds don't line up Voyant's automatically applied segmentations. Now, if my corpus had consisted of individual files of each story, I could use the trends feature to track keywords across stories, but…it didn't, so, that's that on trends.

 line chart with instances of "holmes" "little" "man" "mr" and "said" (the biggest words in the cirrus) plotted against "document segments"

There's also a feature called Contexts, whereby you can select a term and see the sentences in which it apepars in each instance – something that is far more convenient than having to track down all those "faces" to see, for example, whether it's Holmes or Watson's narration who is marking them.

There are many more features you can enable on Voyant – you can see them all listed here – but it is perhaps worth noting that most of them are just different ways to graphically represent word frequency and word collocation (aka, which words appear together).

There is also – *theoretically* – much customization that can go into which words Voyant counts and which it ignores. Voyant automatically filters for words it sees as too ubiquitous to be worth counting, things like "a" or "the" or "I", and, again, theoretically, it's possible to edit that filter-list and tell Voyant that you *do* want "me" to be counted, or that you *don't* want "Mr." to show up in your cirrus because it's…not even a word, so how is it *literally* the 4th highest occuring word in the whole corpus?! Theoretically, one would do such a thing by clicking on the "options" toggle over the cirrus cloud, and then click "Edit List" next to the "Stopwords" category. I say "theoretically" (I have said it a couple times, perhaps you have noticed), because in many many trial runs and experiments with Voyant I have never been able to get a word I added to the stopwords list to actually stop showing up in the cirrus cloud. I have seen others online succesfully use this feature, so maybe it's just inconsistent, or I'm just doing something wrong, but – something to keep in mind and proceed with caution around if you are interested in using Voyant!

DHC Weekly 4/5: Ethical EdTech

Hello DH-er's! This week on the blog I'm profiling a project that recently passed through the DHC — on March 29th, we hosted, along with some folks from Columbia Libraries, an inaugural edit-a-thon event centered on the idea of ethical EdTech.

Ethical EdTech is educational technology that does not cost money, sell students' data, or lock users into proprietary suites and systems. Ethical EdTech is hard to find, however, when institutions have relationships with peddlers of more profit-seeking educational technology (like how many schools use google suites for their institutional communcation — Barnard included). From this conundrum arose the Ethical EdTech wiki, a collaboration between Erin Glass of UCSD and Nathan Schneider of CU Boulder, that now boasts dozens of active contributors.

The Ethical EdTech wiki collects tools that, in the words of the wiki itself:

‣ provide students and educators with greater control over their data and greater understanding of data collection practices
‣ avoid commercialization of the educational experience and the power relations involved in it
‣ expose students to the principles and practices of free/libre/open software

- foster more participatory, critical modes of relating to software
- teach students how to recreate ethical tech practices outside of the classroom

Tools with pages on the wiki range from LibreOffice, an open-source alternative to the Microsoft Office Suite; to the CUNY Academic Commons, a WordPress-based academic platform; to Atom, an open-source plain-text editor commonly used in software development.

At our edit-a-thon, we added pages the wiki, which as it currently stands is small but ever-growing. If you have any interest in open source and/or ethical pedagogical tools, why not make an account?! Check out the contribute page for a list of tools that need pages!

Awkwafina's Wikipedia Page Created at Barnard

APRIL 2, 2019
CATEGORIES: GENERAL UPDATES, SYLVIA KORMAN
TAGS: MEDIA, WIKI, WIKIPEDIA

At a Wikipedia Edit-a-Thon in April of 2013, now 6 years ago (wow!), BLAIS staff partnered with the Asian American Alliance, a student group on campus, to create and edit Wikipedia pages for various Asian Americans. Among the ten articles created on that April day was the page for rapper and breakthrough actress, **Nora Lum**, now known as **Awkwafina**.

What Alexis Seeley started as a short statement with two music video links has transformed into an extensive biography detailing Awkwafina's acting career, television appearances, and accolades with a reference list of almost 40 sources. Comparing Seeley's first edits with the current page, you can see how Awkwafina's explosive roles in *Crazy Rich Asians* and *Ocean's 8* has permanently put her on the map of Hollywood.

Lum at the premiere for Crazy Rich Asians

Looking through the revisions of Awkwafina's Wikipedia page, you can watch how hundred of versions of the same page slowly evolve over the course of six years. Editors left notes to each other suggesting where the page could use more information and how to format certain sections. I urge you to look through the revision histories of your favorite celebrities or topics to see how information has progressed.

Now there's an educational procrastination activity for you.
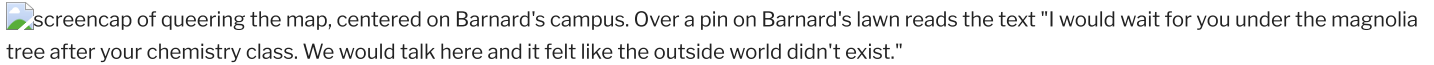
DHC Weekly 3/29: Queering the Map

MARCH 5, 2019
CATEGORIES: DHC WEEKLY, METHODS AND ETHICS OH MY, SYLVIA KORMAN
TAGS: PROJECTS, QUEERING THE MAP

These past weeks I've earmarked many tools, of various complexity and variously high skill-based barriers to entry, that can be used to make map projects. This week, I'd like to close out my unofficial mapping in DH series by looking at an example, one of my favorite public-facing digital humanities projects, Queering the Map.

Queering the Map is a community built map, kind of like what HistoryPin allows one to achieve, where any user can pin queer memories, moments, experiences, and histories.

screencap of queering the map, centered on Barnard's campus. Over a pin on Barnard's lawn reads the text "I would wait for you under the magnolia tree after your chemistry class. We would talk here and it felt like the outside world didn't exist."

The value of a digital humanities project like *Queering the Map* is in the way its form suits its content. The parameters of the project, eg. what "counts" as a pinnable queer moment, are as ephemeral and un-fixed as any other aspect of queer life — if you know, you know, and the opportunity is yours to make that knoweldge however you like. Accordingly, meaning is made on *Queering the Map* in open-ended, unexpected, playful ways; people pin first kisses and bad breakups, self-realizations and hard conversations. These are experiences that in other contexts might be only incidentally geographically linked, one part of a whole array of sensations related to a memory, but on *Queering the Map* they appear purely as geography. With no markers of time period or identity, the pins on *Queering the Map* create a landscape of queer experience, unbeholden to conventional modes of storytelling.
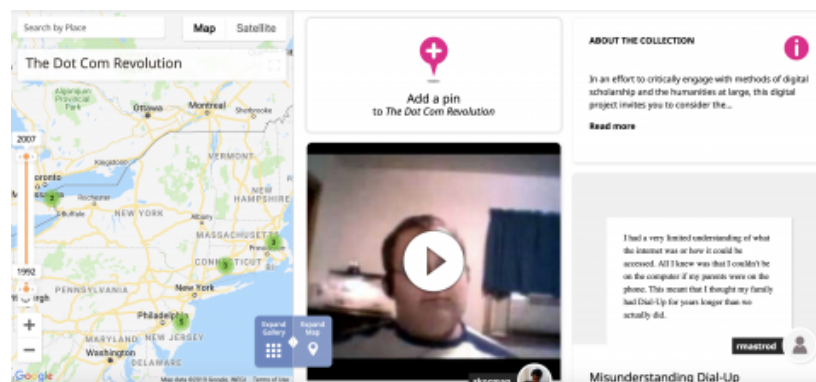
It's a remarkable project – and I hope an inspiring one in terms of how DH projects allow us to discover new ways of seeing, relating to, and questioning the world around us.

DHC Weekly 3/23: Historypin
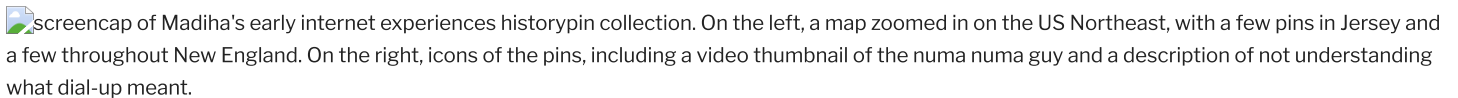
MARCH 6, 2019
CATEGORIES: DHC WEEKLY, GUIDES AND DEMOS!, SYLVIA KORMAN
TAGS: GUIDES, HISTORYPIN

Hello DH-fans! I know it's been map city on the blog lately, but I want to talk this week about one more mapping tool, historypin. Historypin, unlike the other mapping tools I've written about is designed to be created by communities rather than by solo authors. You set up your collection, and then any user can submit a pin, and with it a memory or piece of media.

I first encountered historypin via Madiha, our Digital Scholarship Librarian here in the DHC, when she made a collection to gather stories of the early internet from folks around the library — I added an early meme encounter (my first time watching a youtube video, which was in fact the numa numa guy… ah, nostalgia), pinned to the town I grew up in. Other library staffers pinned memories of early chat rooms and Eastern European internet cafés. If you have any early internet memories of your own, feel free to add them!

screencap of Madiha's early internet experiences historypin collection. On the left, a map zoomed in on the US Northeast, with a few pins in Jersey and a few throughout New England. On the right, icons of the pins, including a video thumbnail of the numa numa guy and a description of not understanding what dial-up meant.

Historypin is a wonderful tool for creating collective community memories that are tied to place and time. At its best, it allows for the decentralizing and democratizing of narratives – as an example, check out this collection, "People's History of Pop," which tells the story of over 50 years of British pop, not from the perspective of stars or music journalists, but of fans themselves. People upload pictures of their signed memorabilia, recount stories of meeting their idols, share scans of their adolescent diaries, all tied to towns and cities all over Britain; at over 4,000 pins, the collection is a moving testament to the ardour and community of pop music fans.

Other collections that in my opinion showcase well the power and value of historypin's platform compile stories of lesbian communities in Hungary, of Chinese-Americans in Southern California, or map artworks from the Imperial War Museum in England.

Making your own historypin collection, or adding to any open collection is as easy as creating an account, or logging in with your gmail, Facebook, or Twitter account – why not start by adding a pin to Madiha's early internet collection?

DHC Weekly 3/15: Marisa Parham at the DHC

MARCH 5, 2019
CATEGORIES: DHC WEEKLY, SYLVIA KORMAN
TAGS: EVENTS

Last week in the DHC, we had the profound joy of welcoming Marisa Parham for a discussion of her work, as a professor of English at Amherst College (where she is a prolific producer of DH projects), director of the Immersive Reality Lab for the Humanities, and former director of Five Colleges DH.

DHC Interim faculty director Janet Jakobsen introduces Professor Parham, standing before a screen displaying her website while event attendees look on.

Interim faculty director Janet Jakobsen introduces Professor Parham.

Professor Parham's talk was engaging and inspiring in many ways and directions, but I want to focus this week on responding to an aspect of her talk that I found particularly compelling. Professor Parham talked about her process as one of taking ideas from conceptual into digital form in ways that "put pressure on the digital," rather than the other way around. I found this idea to be eloquently reflected in the projects she showed us, many of which veered far from the expected structures of a web page in which one follows a linear path through hyperlinks to consume information in the order one desires.

Rather, Parham's projects are nonlinear, playful, and resistant to the mastery of their users. A project created with a student, for example, features a 3D tic-tac-toe board of images that each lead to a poem, an array that reshuffles itself when the user navigates back to it, making it all but impossible to find the same piece twice (that is, until the user has clicked a certain number of times, at which point they are bounced out to a more straightforward index). Another project is set up to sync with average reading times, so that text, images, and videos appear and disappear on the screen as your eye travels down the page. This project, which deals with black womanhood, also incorporates a kind of digital code-switch capability, with hyperlinks that don't announce themselves as passages to the next thing, but are more like passwords to further discussion, easy to suss out if you know what the look for and just as easy to miss if you're less culturally literature in the topics at hand.

Essentially, the digital media of Parham's projects do not self-obfuscate or efface; there is no level on which you could forget that what you are reading is not a weirdly backlit piece of paper. They put, to use her words, the pressure on the digital. One of the reasons I find this so interesting is the way in which "putting the pressure on the digital" means putting a kind of pressure on the reader, too, who must renounce some of the complete control that the typical web-browsing experience leads us to expect. *What if I really wanted to find one of those poems again?* I found myself thinking as Parham walked us through the project. And, of course, the answer is, too bad! Why was I ever asuming that a webpage has to give me exactly what I want when I want it, simply because it is a webpage and not, say, a video, a painting, or a piece of choroegraphy.

Marisa Parham, wearing a white dress with colorful embroidery all over it and a grey cardigan, mid-talk.

Professor Parham took us through several exciting digital projects.

The other reason I have felt so drawn to the idea of "putting the pressure on the digital" is for how it challenges one of the ways I myself have been describing the work of the Digital Humanities. When asked by members of the Barnard community who have no experience with DH what it is we do in the center, I often find myself using some variation of the formulation "humanist inquiries plus digital methodologies." This is an easy way of conveying the fact the Digital Humanities come in peace — that the work we do is Humanities work, and we're not some alien invasion of scary coders hell bent on, I don't know, Marie Kondo-ing everyone's bookshelves into oblivion and replacing them with many many laptop carts or something — but Professor Parham's talk really revealed to me how lazy a formulation it is.

Humanitist inquiries, digital methodologies is a neat turn of phrase for a field of work that is itself profoundly un-neat. Professor Parham described the way her research focuses on aspects of black culture that have always been in some way digital, even before the platforms themselves existed (her example is Black Twitter arguably pre-existing Twitter itself), which challenged me to think of digital not as something being added to humanities. Rather,

the inquiries themselves can be — and have always been — digital, and the methodologies shouldn't take digital-ness for granted but, again, should always be putting the pressure back on the digital. The playful or even mischevious qualities of Professor Parham's work are not pre-existing qualities implemented by digital tools, but qualities inseperable from the digital medium of her work and its possibilities (arguably they are themselves digital qualities, but this is perhaps a topic for another post).

A view from outside the DHC, with the flier for Professor Parham's lecture up on the glass wall, and the attendees visible through it.

By putting pressure on the digital, we leave behind the world of the toolkit, challenging our conceptual thought to be fully integrated with digital practices, and questioning our assumptions about both. What a tremendous resource scholars like Professor Parham are in this process!

If you joined us on Wednesday for Professor Parham's talk, thank you so much for your time and your engagement! If you missed it, be sure not to miss the DHC's next event, a reading group for Philip N. Howard's *Pax Technica: How the Internet of Things May Set Us Free or Lock Us Up.*

DHC Weekly 3/8: Timemapper

MARCH 5, 2019
CATEGORIES: DHC WEEKLY, GUIDES AND DEMOS!, SYLVIA KORMAN
TAGS: GUIDES, TIMEMAPPER

Hi DH fans! Do you remember a few weeks ago, when I looked at ArcGIS for mapping? And then a few weeks before that, when I looked at TimelineJS for making timelines? Well, what if you could do both…at the same time?

Thanks to TimeMapper, from Open Knowledge Labs, you can! TimeMapper is a tool that is quite like TimelineJS, in that it uses a google sheets spreadsheet to build a timeline. The difference is, this spreadsheet has slots for geographical coordinates, so as you add entries to your timeline, you're also adding points on a map, which are then displayed side-by-side.

As an example of the sort of project that could benefit from a tool like this, I'm going to return to the topic of opera seasons (I'm sure none of you got enough last time!). If you're interested in critically examining a canon like the one found in mainstream opera performance, it should be pretty obvious why a timeline/map combo might be useful as a way of quickly and elegantly visualizing when and where these operas are from.

For my spreadsheet, I decided to use the date of premier for each work and the coordinates of the city where the work was premiered. Sometimes this latter category can be a little surprising — did you know that Puccini's *La Fanciulla del West* was premiered in New York City, and that Verdi's *Aida* premiered in Cairo? You can have a look at the spreadsheet I built here — it's quite bare bones, as you will see, but the options are there to add images and all sorts of media (I couldn't help myself, I added some media from the operas in question).

From there, it's as simple as publishing the spreadsheet, plugging the link into the site, and check it out!

This is what it looks like — click the link above to actually explore/interact!

The result is a timeline and a map, side by side in an easy to digest form, with the option to add media of all sorts. For my project, it's an immediately visual representation of the Eurocentricity and the 19th-century-centricity of the operatic season!

DHC Weekly 2/22: Algorithms of Oppression at the DHC

FEBRUARY 5, 2019
CATEGORIES: DHC WEEKLY, METHODS AND ETHICS OH MY, NEWS AND REVIEWS, SYLVIA KORMAN
TAGS: AI, ALGORITHMS OF OPPRESSION, EVENTS

This Thursday, we welcomed faculty members, library staff, and students into the DHC for a lunchtime discussion of Safiya Umoja Noble's book Algorithms of Oppression. We were so thrilled by the turnout, and by everyone's eager and insightful comments.

reading group attendees around the DHC tables

A large part of our discussion of Dr. Noble's book centered on questioning roles and dynamics with regards to the creation and consumption of algorithms that too often go un-scrutinized. Professor Kim Hall challenged us to question the commonly held narrative of the lone-wolf-tech-genius, asking in whose interest it is for us to continue to imagine a single brain in a room creating algorithms, and not the teams of developers who actually do this work?

professor Kim Hall speaks, while other attendees listen

Just as the popular conception of the creators of algorithms such as Google's often does not line up with the reality, so too are our own identities relative to companies like Google not necessarily what we believe them to be. We want to think of ourselves as customers of a product like Google search, a position that implies a certain amount of power and agency. As Dr. Noble points out, however, Google search is first and foremost an advertising platform, albeit one that presents itself as though it were an agenda-less public service. As such, the "product" in question is not the information to be found via Google search, but the views to be garnered for the advertisers; it is the advertisers who are Google's customers, and the public (or its clicks and views) are not consumers but themselves the product being offered.

Finally, we discussed a recent *Wall Street Journal* article, "A Crucial Step for Averting AI Disasters," which details the realization many tech companies are coming to that their algorithms are biased, and that those biases can and will pull in bad press and public embarrassment. This realization, according to the article, has been leading tech companies to create more diverse teams of developers, who are less likely to replicate the blind spots and biases of a white and male driven field.

Bad algorithms are bad for business — but what about what's good for the people? We closed out our discussion by attempting, in small groups, to imagine our ideal search engines, and got back answers — particularly from the Computer Science majors in the room — centered on lateral rather than hierarchical arrays of information, data privacy and informed consent, and democratized community content moderation.

attendees sit around the dhc tables and discuss

In two weeks, on March 7th, from 12-1:30 pm we will be hosting a follow-up Wikipedia editing workshop — I hope to see you all there!

DHC Weekly 2/15: JSTOR Text Analyzer

Hello DH-ers! Do you remember a few weeks ago, when I introduced you to a new tool from JSTOR? In that post, I mentioned that JSTOR's other tool in beta, the Text Analyzer, hadn't been working for me. Well, after many emails exchanged with an infinitely patient project manager at JSTOR (thank you Michael!!), the issue has been sorted, and I am so excited to tell you all about the Text Analyzer!

The JSTOR Text Analyzer is a tool that allows you to use your own documents to search JSTOR for articles and books — just upload or drag and drop a document onto their portal, and an algorithm will scan it for key terms and conduct a search for you. Then, you can add or remove terms to correct any imprecision on the part of the algorithm. Obviously, this is a pretty cool party trick, but algorithms are far from infallible, so I've been excited to play around with the tool and see how effective it is at actually gauging what's relevant.

screencap of the Text Analyzer interface -- it tells you to upload your document to search

As a first pass, I tried uploading the text of a post I wrote for the DHC site a couple months ago, on Shakespeare and video game glitch memes as embodiments of Bakhtin's grotesque. Here's what the Text Analyzer tagged as the keywords of that essay:

the text analyzer results, explicated in the following paragraph

the text analyzer results, explicated in the following paragraph

As prioritized terms we have "Video Games", "Virtual avatars", "Animation", "Clowns", and "Carnival," with more of a kind of general spread of things having to do with games and computers under "Topics". Shakespeare, which I would call the jumping-off point for the whole piece, doesn't show up until you get down to "People", and even there the algorithm can't really tell the difference between "Shakespeare was mentioned in this document because it was written in response to a paper on Shakespeare" and "James Bond was mentioned in this document because one of the glitches discussed came from Goldeneye video game." So from here I might tweak — take out "Clowns" from my prioritized terms, perhaps, and select "Gestures", weight "Virtual avatars" more strongly than "Video Games".

It's worth noting, however, that the goal here is not necessarily the neat trick of putting in your work and getting out perfectly tailored search terms (although that's pretty cool) — the purpose of a tool like this is to find writing that could potentially be in productive conversation with the work you already have, something that it absolutely does, even before making some refining changes.

The Text Analyzer also says that you can upload an image, something that piqued my interest; I was curious if having to OCR text from an image would add an additional layer of error. I tried with the flier image for the DHC's Algorithms of Oppression reading group event (sidebar: if you RSVP'd to this please come pick up your book!), and got pretty much a slam dunk of the keywords "Library collections", "Search engines", "Digital humanities", "Oppression", and "Algorithms".
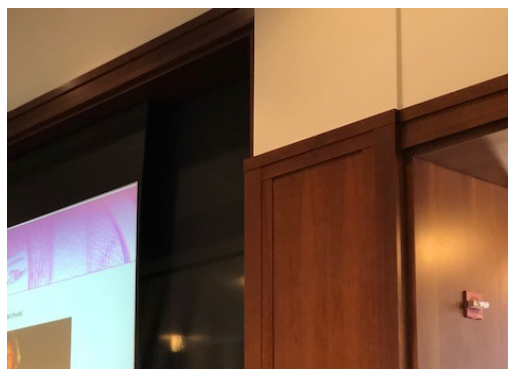
Where I had the least amount of luck was when I uploaded a full 21 page research paper — a piece of writing I'd personally identify as pretty solidly about *Antony and Cleopatra* and performance, but for which the Text Analyzer's output terms were "Ancient Rome", "Roman and Byzantine Egypt", "Art happenings", Mystery plays", and "Arias". I think the correct conclusion to draw here is that the Text Analyzer is most practical given shorter and more to the point copy — less your manuscript than your precis. But either way, it's a fun tool!

DHC Weekly 2/8/17: NYCDH week

FEBRUARY 5, 2019
CATEGORIES: DHC WEEKLY, NEWS AND REVIEWS, SYLVIA KORMAN
TAGS: EVENTS, NYCDH

Hello, DH fans! As I imagine some of you know (since you are, of course, DH fans) this week has been NYCDH Week — that's New York City Digital Humanities Week. NYCDH Week is a week of workshops, panels, and talks happening all over New York City at institutions like Fordham, the New School, Columbia, NYU, CUNY, and more, all centered on the Digital Humanities. Digital Scholarship Librarian and fellow DHC team-member Madiha and I have been dashing around the city like Anne Hathaway in The Devil Wears Prada, but instead of "fashion" or "having a really mean boss" we've been fired up over learning new skills and discovering new projects in DH happening all over our city.

NYCDH week kicked off with an all-day event of panel discussions, lightning talks, and presentations of work, hosted at Fordham University. I don't have world enough or time to share with you all the incredible projects presented at this kickoff event, but I'd like to touch on a few conversations from Monday.

Keynote speaker Meredith Broussard opened things up with a talk about AI — what AI isn't (sorry Data from Star Trek, you may be my favorite character but you're pure Hollywood mythmaking), what AI is (really complicated statistical analysis), and how AI can reflect the biases of both the people creating it and the society in which it is created. She warned against the hubristic refusal to acknowledge that technology is not the inherently perfect saviour of humanity it is sometimes thought of as but is in fact as flawed as we are; a term she calls "technochauvanism" and traces to the overwhelmingly white, male, and ivy-league educated founding fathers of the field of computer science.

Next, Barnard Library's own Jenna Friedman took the stage, along with our neighbor from across the street, Columbia Digital Scholarship Librarian Alex Gil, and Juber Ayala and Wendy Hayden, of the Newark Public Library and Hunter College, respectively, for a roundtable on Information, Democracy, Archives, and Absence. This roundtable raised a lot of fascinating points, and sparked a conversation that I believe is still continuing on twitter, about absence in the archive. Should, for example, all data be publicly accessible? And when we talk about democratizing the archive, how do we discuss the complicated interwoven roles of librarian, archivists, and professors? Jenna, for one, made the excellent point that if a library's collection is uncomfortably un-diverse, it may not have been entirely at the discretion of the librarian, but also reflective of the curricular needs of professors. In other words, it's all related! And it's on all of us to interrogate our library and archival spaces.

I want to conclude with highlighting one of the neatest projects presented on Monday. Katy Gero of Columbia University won one of NYCDH's gradute student project prizes, for her tool, metaphoria, aimed at generating metaphorical connections between concepts. You can explore metaphoria yourself here, but what I found most interesting is the ways Katy described poets working with metaphoria. Some poets, she explained, mostly wrote on their own, turning to metaphoria if they were in a jam or needed some specific input, using it much like one would use a thesaurus. Others took a metaphoria suggestion as a starting point to riff on at length in their own words. A third poet build a poem more or less entirely out of metaphoria outputs, effectively co-writing the piece with the tool. I found this range compelling as an example of the elasticity of Digital Humanities tools — there are hundreds of ways to relate to technology in our work, and it's always a treat to get to peek in on some of them.
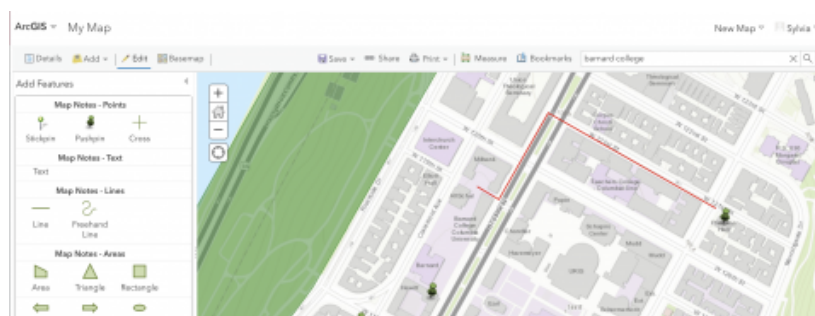
NYCDH week has continued with a full schedule of workshops with DH tools and concepts, taking place all over the city — Madiha and I have attended workshops learning about data epistomologies, wikidata, DH in the classroom, and many many more. All of the workshops are free and open to anyone interested in the Digital Humanities. Next year, I hope to see you there!
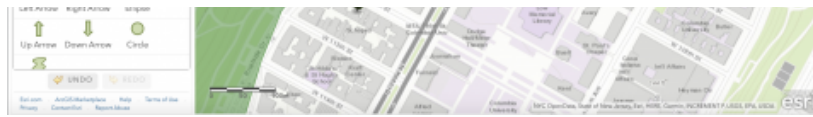
DHC Weekly 2/1/19: Maps!

FEBRUARY 5, 2019
CATEGORIES: DHC WEEKLY, GUIDES AND DEMOS!, SYLVIA KORMAN
TAGS: ARCGIS, ART HISTORY, GUIDES

This semester, the DHC and me (postbacc Sylvia) in particular have the profound joy of working with Art History professor Elizabeth Hutchinson and her class American Monuments. An early assignment for the class is inspired by the Southern Poverty Law Center's report "Whose Heritage?: Public Symbols of the Confederacy" (read the report here). The report, which tracks and accounts for over 2000 monuments, commemorated schools, courthouses, and roads, and other public symbols of the Confederacy, is a masterclass in the power of data visualization, from timelines plotting the consecrations of these symbols against major events in the Jim Crow and Civil Rights Eras to a map with each monument, park, and building plotted onto it. This is only one example project of the kind of power a map can have. But how can you or I make a map with as much power as those created by the SPLC?

The answer is ArcGIS, a geographic information system used to create and work with maps.

ArcGIS lets you choose a basemap (think google maps vs. google earth), but even cooler it lets you import layers added by other users on top. So say, for example, that I'm interested in looking at national parks. Sure I could go in to my map and painstakingly trace the boundaries of every national park…or, I could just see if there's a layer I can import.

What do you know, there is! In addition to adding pre-made layers, I can add points, sections, and captions to my map by adding map notes. So, for example, I can mark every Barnard residence hall I lived in, plus the route I used to walk to get to class when I lived in Plimpton.

I swear it felt twice as long as it looks on the map.

With features like Story Maps, ArcGIS can be used to make beautifully eloquent maps that show geographical changes over time, or compare different overlays of information. Check out this map on America's Protected Areas, or this one that compares Earth at Night imagery from 2012 to 2016 to tell the story of human movement and ambition. Maybe they'll spark some inspiration to get mappin'!

DHC Weekly 1/25/19: Data Visualization

JANUARY 5, 2019
CATEGORIES: DHC WEEKLY, GUIDES AND DEMOS!, SYLVIA KORMAN
TAGS: GUIDES, OPERA, TIMELINEJS

Hello Digital Humanities fans! This week on the blog I want to play a little with some user-friendly data visualization, in the form of Excel charts and TimelineJS.

As my data set this week, I'm taking a look at a world near and dear to me – that of professional opera. As probably very few of you are aware, it is now that time of year when American opera companies start to announce their seasons, which means it is also that time of year when opera fans and musicologists and arts journalists have serious conversations about representation on the operatic stage, and the overwhelmingly conservative and regressive nature of the operatic canon. Because production costs are so much higher and talent pools so much smaller and more specialized, opera companies tend to pull from a much more restrictive canon than theater companies – but just how restrictive is the operatic canon? I took the recently announced 2019-2020 seasons of one major American house, the Lyric Opera of Chicago, as well as the 2018-2019 season of the most prominent American opera house, the Metropolitan Opera (their season for next year has not yet been announced), and am going to look at a couple different ways to visualize their diversity (or lack thereof).

One of the first decisions to be made when putting together a dataset is what sort of information to include; I knew I wanted to look at how old the works being performed were, how many (if any) were written in languages other than the big three of Italian, German, and French, and — optimistic fool that I am — I added a column for the gender of the composer. In Excel, my dataset for the Lyric looked like this:

Simple data visualization is actually quite easy in Excel – from this dataset I pulled out the information on the languages:

By selecting these cels and choosing the pie chart option under the "Insert" tab, I get a nice pie chart:
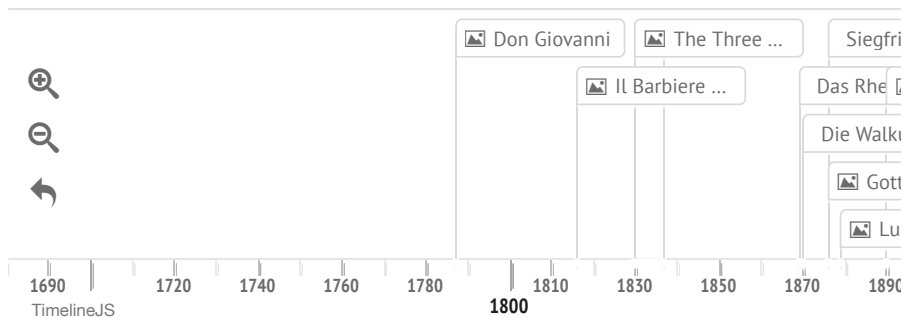
Let's compare this to the language breakdown for the Met, which has a significantly larger season (24 operas rather than 12):

What jumps out to me immediately is that the Lyric has no French operas programmed for next season, whereas a quarter of this year's Met productions are in that language. I am more surprised, however, by these seasons' fundamental similarity – even with vastly different resources and different scopes, they are both significantly Italian, about 40% of each season.

The next question I had was about how old these works are, and while I could have sorted them into centuries and created more pie charts, I instead decided to use a program called TimelineJS to create timelines for these seasons – I suspected being able to experience and interact with the gulf between the operatic classics of the 19th and 18th century and the few new or contemporary works in these seasons might prove interesting.

TimelineJS is extremely user-friendly – it generates a Google Sheets page that you plug your data points into, and then generates a timeline from them:Here's what my Lyric Opera timeline turned out looking like:

LYRIC OPERA 2019-2020 ❯

And here's the Met's

MET OPERA 2018-2019

This is a timeline of the operas from the Met's current season

›

	Don Giovanni		La Fille du	D	Samson e
	La Clemenza ...		Rigolet	Aida	Otell
			Les	Carm	Iol
			M	Siegfri	L
			D	Gotterda	

1690 1730 1760 1790 1830 1860 1890

TimelineJS)0 1800 190

Again, as with our pie charts, what strikes me here is that despite the Met having fully twice the productions of the Lyric, the distribution across centuries is more or less the same – with one or two Mozart operas from the 18th century and one or two contemporary works flanking a veritable glut of 19th century opera (admittedly this is probably a little skewed by both of these seasons happening to contain full productions of Wagner's four opera Ring Cycle).

The Lyric is probably the second most prominent opera house in the United States, after the Met – going into this exercise, I was unsure of whether the Met's larger season would allow them more room to take chances on less mainstream works or if the Lyric's lessened scrutiny in the opera world would push them further from conservatism. What I found is that neither of these were the case – the Lyric's season and the Met's are in many ways built along the same lines, albeit at different scales. To my own question of how restrictive is the established operatic canon, I would say that at the country's largest houses, the answer is *extremely*. This is a perhaps a less than shocking conclusion, to me, a person who has followed these sorts of season announcements in the past, but the value of the visualizations is that I could show them to anyone and they, too, would be able to see the lack of variation in these seasons.

I'd like to leave you with one last set of charts, that kind of say it all when it comes to the lack of diversity in the opera world. Remember my columns for composer gender? Well, here they are, graphed:

Yeah...

DHC Weekly 1/18/19: Text Analysis and JSTOR

Hello Digital Humanities fans, and welcome to part two of Things Sylvia Learned at the MLA Conference!

This week I want to talk about a new text analysis tool that JSTOR is developing, and the ways it might broaden the idea of a "text" to include the critical thought that surrounds it, as well as destabilizing that critical thought's location within a static discipline or field. Perhaps you, like me, use the JSTOR website regularly without having noticed the "Tools" tab, hiding out next to "Advanced Search" and "Browse." Last week at the MLA, the kind people staffing the JSTOR booth were good enough to give me an introduction to the tools in beta that are already accessible and useable on the site. One of these tools, the Text Analyzer, which allows a user to search JSTOR by uploading a document, which an algorithm digests and turns into search keywords, has been giving me a little trouble (these tools are still in beta!) but while I navigate the troubleshooting process, I want to talk about the other new tool, the JSTOR Understanding Series.

The JSTOR Understanding Series takes (public domain) texts in their entirety, and displays them with a little number next to each line – this number is the number of articles on JSTOR that cite or quote that line. So a pretty famous section of Romeo and Juliet, for example, might look like this:

Romeo's line "she speaks," perhaps unsurprisingly, is quoted only once, whereas Juliet's much more famous "What's in a name? That which we call a rose / By any other word would smell as sweet," has a significantly higher count, with 78 articles quoting the full line and 51 pulling just the pithier "a rose / by any other word…" Clicking on the numbers lets you see the articles that quote the line or passage:
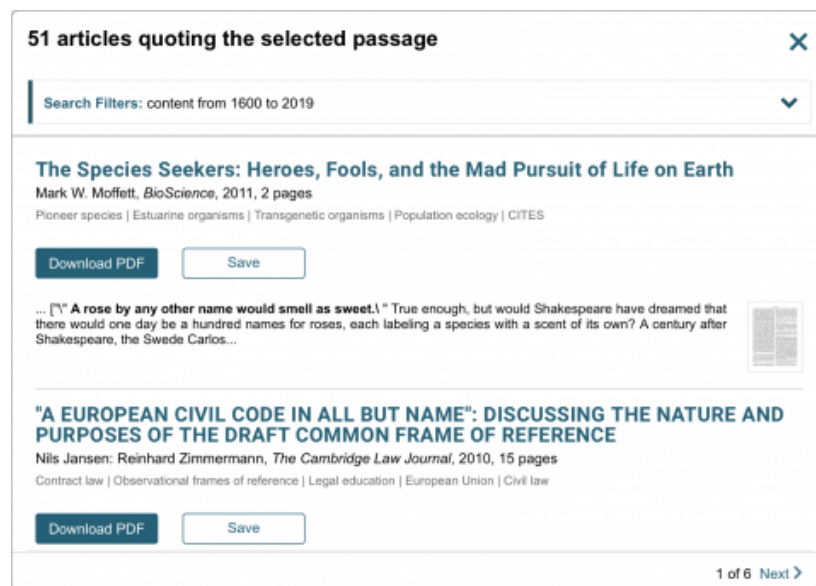
The value of this tool is multifaceted. I, for example, would have loved to have known about it when I was writing my senior seminar paper as a Barnard English major. The assignment was not only to present a holistic reading of a play by Shakespeare, but also to situate that reading within the existing field of criticism, a difficult task considering that the field of Shakespeare scholarship has had 400 years to become a dense and sprawling forest in which it is remarkably difficult for an undergraduate to find their way. To search, say, "Antony and Cleopatra" on a library database or JSTOR is to uncover far more reading material than is possible to comb through, even for a thesis paper. With a tool like this, however, a student can allow their own reading of and interests within a text to guide them towards the relevant critical material.

The JSTOR Understanding Series tool effectively broadens the idea of "the text", at least in the context of critical writing, in which the text could somewhat loosely be defined as the material under the lens of thought, to include the existing scholarship around it. One could, of course, argue that this broadening is already implicit in an assignment that asks a student to engage with existing literary criticism, as that student is hardly bringing different skills or attitudes to their close readings of the source text than they are to the critics – fundamentally, the work of interpretation needed to advance a reading of a play is pretty similar to the work of interpretation needed to agree or disagree with an argument.

But the JSTOR Understanding Series goes beyond the idea of transferable skills and actually creates, digitally, a composite text in which it is possible to read the source and the critics more or less simultaneously. This is true even if you never open up and browse the list of citing articles – suddenly, a line's relative weight in the world of scholarly writing is immediately evident, replacing the line number as the most visible and immediate organizing principle scaffolding the way the text is read. Romeo and Juliet's reception is as present on the page as its poetry is. And once you start clicking those hyperlinked citation numbers, the relationship between each line of the source text and the article that cites it is so immediate as to all but erode the barrier between primary text and secondary critic altogether.

The barrier between source and criticism is not, however, the only barrier being eroded; the JSTOR Understanding Series compiles *every* article that quotes a given line, its algorithm is heedless of disciplinary silo-ing or relevance. The result is that the critical thought being welcomed into the realm of "source text" is dizzyingly interdisciplinary. For an example, let's take a look at those 51 citations of "a rose / by any other word would smell as sweet."

A glance at just the first page of results for this line reveals, among other things, two papers from science journals about literal plants, a paper from a law journal, an article called "What's in a Name.com?: the Effects of Name Changes on Stock Prices and Trading Activity", an article on a Washington Irving-penned satirical periodical that once quoted the line, as well as several of the expected papers on the play itself published in journals of English literature or drama.

Obviously, there are immediately apparent downsides to the catholic nature of the algorithm. If I were writing my seminar paper on Romeo and Juliet and wanted to quickly locate my own reading of this line vis-a-vis a body of scholarship I would be pretty frustrated at having to scan and discard articles that are not really about Shakespeare. I imagine this problem becomes far less irksome with lines that are much written-about but less culturally iconic. For example, the entries, numbering 20, for Romeo's claim that Juliet, "hangs upon the cheek of night / As a rich jewel in an Ethiop's ear," turn up more of the expected articles in Shakespeare Quarterly and The English Journal, as well as books tagged with the keywords "Black feminist theory".

However, even when the algorithm is casting its most frustratingly wide net, it still reveals something productive about the nature of Romeo and Juliet, namely, that this is a text that can expand to encompass not only a critical context of scholarly thought but a cultural context of interdisciplinary allusion. Shakespeare is not just Shakespeare with the JSTOR Understanding Series. Rather, the text of Romeo and Juliet must in some way include the botanist making a point about the way plant species are named and the lawyer writing probably extremely intelligently about some stuff that I tried to skim and summarize and could not understand one single word of.

Cambridge Law Journal, **69**(1), March 2010, pp. 98–112
doi:10.1017/S000819731000019X

"A EUROPEAN CIVIL CODE IN ALL BUT NAME": DISCUSSING THE NATURE AND PURPOSES OF THE DRAFT COMMON FRAME OF REFERENCE

NILS JANSEN AND REINHARD ZIMMERMANN*

"What's in a name? That which we call a rose
By another name would smell as sweet."
(Romeo and Juliet, II, ii, 1–2)

I can't even tell which of the words in this article's title are proper nouns.

This blurring of critical and cultural contexts is actually made more visible by the JSTOR Understanding Series' greatest limitation in these early stages of its operation – the extremely white, Western, and male nature of its corpus, which comprises pretty much solely works of classic American and British literature and thought like the complete works of Shakespeare, the King James Bible, the United States Constitution, and so on. As of right now, the only works you can read on JSTOR in this digitally-dual form are the works of the Western canon, so, the works with the most culturally ubiquitous language that are the most likely to be used as epigraphs or brief allusions by writers in any and all fields.

I imagine the Understanding Series is mostly being thought of the way I pictured my senior-thesis-writing self as using it – as a way to conduct very specific research tailored to the places in a text that the researcher is already invested in. Or, perhaps, as a way of telling at a glance which parts of a text have been historically been the most contentious or placed under the most scrutiny. I, however, find the Understanding Series most interesting and compelling for the ways it destabilizes the idea of a text altogether. Changing technology changes the way we read (think, for example, of the idea of a stable and immutable text of a poem only taking root as privately-circulated handwritten copies stopped being the primary form of publication). Tools such as the JSTOR Understanding Series seem to me to be technologies with the potential to change how we think, relate to, and handle texts, as readers and as scholars. To online editions of highly allusive works like Ulysses or Moby Dick that allow for new levels of simultaneously presented text and contextual information, the Understanding Series promises to add a whole new layer of text-as-text-plus. And I, for one, am excited to see what kinds of readings will follow!

DHC Weekly 1/11/19: Glitch and the Grotesque at the MLA

JANUARY 5, 2019
CATEGORIES: DHC WEEKLY, SYLVIA KORMAN
TAGS: LITERATURE & POETRY, MEDIA

Hello and Happy New Year from the DHC! I (DHC Post-Bac Sylvia Korman) am recently returned to campus from the MLA's 2019 conference and excited to share with you over the next couple weeks some tools and reflections picked up over the course of an unseasonably warm winter weekend in the Windy City.

This week, I want to consider what it means to apply digital humanities methods to the study of texts or concepts from way before digital technologies existed. At MLA, Professor Gina Bloom of UC Davis gave a paper at a session all about applying digital concepts such as "glitch" and "noise" to Shakespeare that took up performance as the interface between digital methodology and pre-digital texts. She discussed two examples of recent Shakespeare projects that involve motion-capture technology.

The first was the Royal Shakespeare Company's 2016 partnering with Intel for an expensive, high-tech, 400th anniversary production of (what else) The Tempest, in which the actor playing Ariel was transformed via mo-cap into a digital double act, shared between the actor, Mark Quartley, and a projected avatar. The process was fraught with glitch anxiety – the production seemed to be dependent on absolute mastery of the technology creating the digital Ariel. One glitch, and the RSC Tempest could suffer the same public embarrassment as the ignoble Met Opera Ring Cycle, in which Robert LePage's feared and loathed set, called the "machine", would stick, freeze and, on one memorable occasion, project the Windows logo onstage instead of Valhalla. The RSC crew also nicknamed their tech, calling the array of computers that allowed for the real-time manipulation of the Ariel avatar the "beast."

Mark Quartley as Ariel

The RSC approach is definitely technological, but it fails to engage the potential of digital methods to challenge our cultural assumptions. If the technology is the beast, Bloom argued, the production's theater-makers become tamers, reproducing the problematic dynamics of The Tempest itself – alienated and othered, the technological body of the mo-cap avatar is more Caliban than Ariel, an abject beast to be mastered by the pinnacles of human art, as wielded by the Prospero figure of director Greg Doran.

On the other end of the spectrum is Dr. Bloom's augmented reality Shakespeare-performance game Play the Knave. Players choose a scene from Shakespeare, and then select sound design, a recreation of a historical or contemporary theater to perform in, and an avatar to represent them on that stage. Then, as they act out the scene, a Kinect, a webcam-style motion capture add-on for the Xbox console, records and maps onto their avatars the gestures they perform.

Play the Knave Performance (Inside the Game 2)

🔳 https://youtube.com/watch?v=h2jimn3tqEA

As you can no doubt see from the video, Play the Knave is glitchy – avatar's limbs twitch and shudder, phasing jerkily through their own bodies, the floor, their scene-partners. But these glitchy avatars are far from the failure implied by the RSC's attitude. Rather, the relentless glitchiness of the Kinect gameplay brings the gesturing body of the player and the glitching body of the avatar into productive apposition. Here, we, too, are glitchy; the player does not have the RSC's option of disavowing the avatar as beast, but must identify it as somehow fundamentally their own body. As we modify our own movements to achieve a smoother animation, we step further away from the typically valued world of naturalistic gesture and closer to a collaborative blending of player and avatar.

Professor Bloom's idea, that we should embrace the glitch, that a glitchy body in virtual space offers us something vital in its relationship to our own bodies, resonated with me. In fact, it resonated with me in that most specific and ineffable of ways: it made me laugh. To illustrate Play the Knave'sidiosyncratic gameplay, Bloom played a video not unlike the one embedded above, and, reader, it was *funny*. The way the avatars' limbs would disappear and reappear, getting caught in little loops of bizarre, unintentional-looking movement, had the whole room of jet-lagged academics giggling into their legal pads. Why do glitching bodies make us laugh? I've been mulling it over it all week, and I think I'm ready to answer that question. But first, I want to establish that video game glitches *do* (and always have) make us laugh. I want to talk about some memes.

As the venn diagram of "people interested in viral online content" and "people interested in video games" has always been more or less a circle, many memes are born from replicating, warping, or imitating popular video games. The latest example of this is how currently extremely-online-teens are recreating the character dances from the hugely popular Fortnite (dances that are themselves largely stolen from predominantly Black culture-creators), but in 2010 we were all about geddan. Geddan was a meme created by users on the Japanese video-sharing site Nico Nico Douga from footage from the Nintendo 64 James Bond game Goldeneye, which, when the game cartridge was inserted slightly askew, had a glitch that turned the characters' bodies into Hieronymous Bosch-meets-Katamari Damacy maelstroms of twitching, flailing limbs. Then, another user added the song "Promise", by Kohmi Hirose to footage of the glitch.

To Watch: https://www.youtube.com/watch?v=SP5c_MEs9mo&feature=youtu.be

Unfortunately, the best compilation of the resulting meme, a four-way split-screen masterpiece, has been taken offline following a copyright claim from Victor Entertainment, the corporation holding the rights to "Promise." There are still, however, some videos floating around on youtube that escaped the same fate, as well as a Know Your Meme primer. Basically, people would take a bunch of pictures of themselves in crazy poses, and then stop-motion animate them together to get a "live action" imitation of the Goldeneye glitch, set, of course, to Hirose's high energy J-pop banger.

https://youtu.be/XOPh44nMijo

The result is glitch-as-choreography – a dance craze only joinable with video editing software and a lot of dedication. That dedication is not to be undersold – making a geddan video took effort, both the physical effort of jumping around to get a wide enough range of images and the time effort of editing all those images together. Nevertheless, hundreds of people put the effort in, in pursuit of making a "normal" human body do the crazy things a virtual body can do when it glitches. Play the Knave closes the loop of player imitating glitch. With mo-cap, the player's body directly generates the avatar; the glitching avatar is the player's own body. The popularity of geddan reinforces Bloom's own sense that Play the Knave's glitches, and the unavoidable identification between avatar and player they imply, are not bugs but features, realizing a deep and irresistible human affinity – and desire – for glitch. Fundamentally, when we see a digital body really going wacko, we don't want to master it, smooth out its imperfections and misfires, like the RSC would have us believe, but, rather, laugh at it and then, hopefully, become it.

I keep returning to our laughter when confronted with our own affinity for glitch, because I think it points to a useful framework through which to think about glitching bodies, Mikhail Bakhtin's ideas of the carnivalesque and the grotesque. Describing the folk humor of the Middle Ages and Early Modern period, Bakhtin compares the destabilizing, anarchic ritual of the carnival, which "built a second world and a second life outside officialdom" (Bakhtin, 6) and the official feasts, "whether ecclesiastic, feudal, or sponsored by the state …sanctioned the existing pattern of things and reinforced it" (Bakhtin, 9). This is more or less the difference between Play the Knave's motion capture and the RSC/Intel collaboration's: the former takes the high-culture property of Shakespeare's plays and brings them into the sphere of the unofficial, where Hamlet can be recited by a tiny, twitching alien avatar while young people point and laugh at it; the latter in its flawlessness allows for no slippage of identification between the physical and the digital and in its narratives of human mastery only reinforces antique notions of canonicity and power.

It is worth noting that the Royal Shakespeare threw all of their officially sanctioned cultural power at their mo-cap project – their money, their access to collaborations with high-profile corporate entities, their status as *the* Shakespeare company best suited to a lavish 400th anniversary celebration. And yet, the success of the project still seemed to rest on the Ariel avatar's perfection. This speaks to the power of what Bakhtin calls the grotesque body, the body that is "not separated from the rest of the world. It is not a closed, completed unit; it is unfinished, outgrows itself, transgresses its own limits" (Bakhtin, 26). The grotesque body, often at the heart of the carnival festivities, consumed and excreted, it was exaggerated, expressive, twisted and deformed – very like, in other words, the bodies created in a video game glitch.

The undeniable zenith of glitch-grotesquery is the GMod video, one of the most crass, crude, and anarchic versions of video-game-glitch-as-viral-humor, which became popular around the same time as geddan. Garry's Mod, or GMod, was a sandbox style "game" (it has no gameplay, rules, or goals) through which players could access and manipulate characters, objects, and textures from any of the games made by Valve, including the then culturally dominant titles of Portal, Half Life 2, and Team Fortress 2. People used GMod primarily for making odd little slices of video game surrealism, making characters, usually from TF2, flail their constantly glitching limbs about, spouting gibberish from distinctively warped faces, created essentially by telling the game's physics engine not to bother keeping facial expressions normalized to a face's usual scope of physical motion. (Warning for some cartoonish body horror on this next video – nothing gory, just a little warped!)

Heavy and Scout think they are birds

The GMod videos, while they might seem unwatchably inane and noisy now, were pretty wildly popular in their time; the embedded video, "Heavy and Scout think they are birds" has 2,652,589 views on youtube. Unsurprisingly, considering the demographics making and consuming them, GMod videos also usually involved cartoonish violence and crude sexual humor — the humor of the grotesque. The contorted faces created in the GMod videos are strikingly similar to those of architectural grotesques:

As Bakhtin describes, the focus here is on "those parts of the body that are open to the outside world, that is, the parts through which the world enters the body or emerges from it, or through which the body itself goes out to meet the world" (Bakhtin, 26): the open mouth, the protruding nose, the genitals. The warped profanities of the GMod avatars, like the masked reveler, the circus freak, or the clown, celebrate all that is culturally debased about the body by offering our own bodily excesses, functions, and excretions back to us, exaggerated and to be laughed at.

As the tides of online humor have turned from lol-so-random inanity to something a little cooler and more deadpan, the madcap grotesquery of the GMod video has fallen out of favor. But people still love glitches, and the spirit of the online-carnivalesque lives on in in the popularity of the notoriously glitch-heavy games of studio Bethesda (just search "Bethesda glitch" on youtube and you will find hours of gamers celebrating these games not despite but *for* their recurring glitches) and in series like Monster Factory, in which brothers Justin and Griffin McElroy use the character-creation screens of video games to make and play as hyper-exaggerated "monster" characters.

A pretty typical Monster Factory creation.

The appeal of Monster Factory is precisely that of Bakhtin's carnivalesque – like the Feast of Fools, in which the social hierarchies of medieval Europe were mocked in their reproduction, inverted, with the lowest of the social order standing in for the highest, Monster Factory draws its humor from the co-opting of structure to produce anarchy. The character-creation sequences of the games, part of the ordered form in which the game is meant to be played, become fairgrounds in which the game is made to look ridiculous as it must cope with wildly exaggerated, carnivalesque bodies. And still, in the series' significant online following, the same affinity of the geddan memers for the glitchy avatars persists. Fans cosplay as Monster Factory creations, using costumes and makeup to, again, make their own bodies match the grotesque bodies created in the games.

The humor of the grotesque and the anarchy of the carnival, Bakhtin argues, was increasingly suppressed in the modern era, forced out by aesthetics of beauty and ideas of taste. But I think it might potentially be back, if only virtually, in the world of video game glitch humor. If this is the case, then Dr. Bloom's Play the Knave project may well be successful on a level beyond that of bringing the player into identification with the digital; the game also brings the player into identification with the Renaissance, a period in which the humor of the carnivalesque was much more present and accessible than it is today. Our laughter, when confronted with a glitchy body, may well have more in common with the laughter at Shakespeare's fools than we know. And

this too is a sort of carnivalesque construction: after all, the carnival collapses difference – between peasant and king, performer and spectator, then and now.

That's all for this week! Next week I will be reflecting on the MLA some more, by exploring some cool new tools JSTOR has in beta that the folks at their booth were kind enough to introduce me to! In the meantime, the Barnard Digital Humanities Center twitter is up and running! Pop on over there and let me know what your personal favorite piece of glitch humor or art is!

DHC-son's Greetings!

DECEMBER 3, 2018
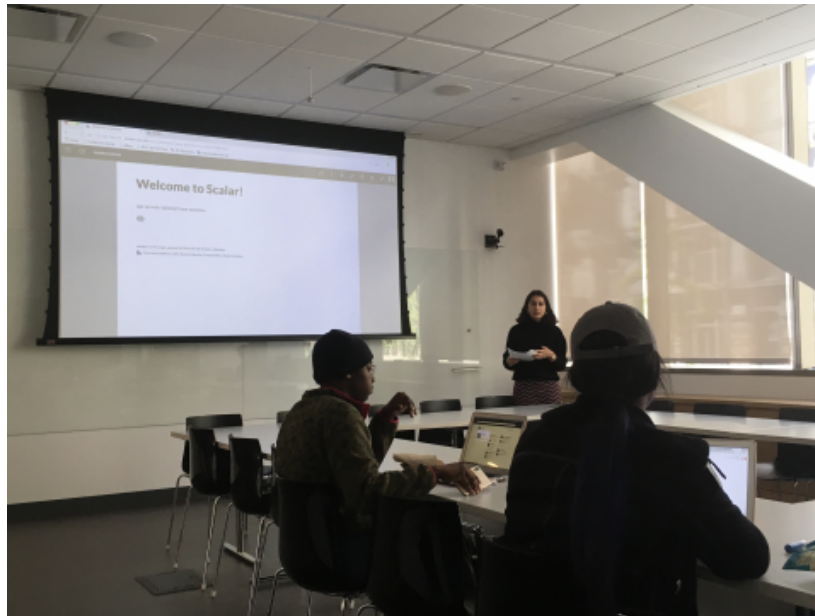**CATEGORIES: DHC WEEKLY, SYLVIA KORMAN**

It's finals week here at Barnard, and as students look back at their semesters for exams, we're taking a look back at ours at the DHC – and what a semester!

At the grand opening of the Milstein Center for Teaching and Learning, we got to meet and chat with so many students, alumnae, and members of the Barnard community about the Digital Humanities at Barnard! One of my favorite things about the grand opening was getting to showcase past Barnard DH projects to Barnard alumnae who majored in the humanities, and talking with them about how humanities study is both different and the same in the era of digital platforms and methodologies. A particular favorite of the many Art History alumnae who stopped by for the opening was the Parquet to Parterre video lecture (Catherine Turocy and Barnard College, 2015), which illustrates the relationship between the gardens and the court dances at Versailles by overlaying garden plans onto footage of dancers from the New York Baroque Dance Company.

Meeting and greeting at the Grand Opening!

With the Milstein open for business, the DHC was ready to start doing what it does best — working directly with students and faculty on DH projects. Overwhelmingly, Fall 2018 was the Semester of Scalar at the DHC. We ran four workshops acquainting students in Professor Elizabeth Hutchinson's North American Art and Culture class and Professor Kim Hall's Ntozake Shange & Digital Storytelling class with Scalar, a digital publishing platform from the Alliance for Networking Visual Culture (learn more about Scalar here, on the ANVC site!). Scalar is cool because it's customizable, collaboration-friendly and supports all kinds of media, so Professor Hutchinson's class was able to post and annotate pictures from an exhibit of Native American art at the Met Museum, and Professor Hall's class was able to create multimedia digital exhibits engaging with the work and life of poet, playwrite, and Barnard alumna Ntozake Shange.

Madiha teaching Professor Hall's class the Scalar basics

Next semester in the DHC, we'll be adding more public events, including a workshop series on data privacy led by our own Digital Scholarship Librarian, Madiha Choksi, so watch this space for more info on how you can lock down your online info. Also, you should watch this space for more posts from me, Sylvia, our Postbaccalaureate Fellow! I'm going to be checking in weekly to talk about tools and events in the wide world of the Digital Humanities – I hope you'll join me!