Teleoptical Perspectives on Digital Methods: Scientific Claims and Consequences

Julia Pennlert, Björn Ekström, and David Gunnarsson Lorentzen

Introduction

Digital methods, according to Richard Rogers, "strive to make use not only of born-digital data but also the methods that are native to the medium."[1] Following Rogers's description, digitized or born-digital material opens up the possibility for adopting new computer-assisted research methods and techniques that collect, analyze, or describe a specific material. Such tools potentially influence the research process as well as the scientific outcome of a certain investigation or study. The development and use of digital methods within social sciences and the humanities have also had impact on scientific debates and discussions about how these new methods can contribute to the development of knowledge claims and knowledge-making in a contemporary media landscape.[2] As noted by Kenneth M. Price and Ray Siemens, digital technologies affect the ways that texts are available and perceived, which, according to Price and Siemens, leads to a need for the researcher to "adopt altered research methodologies."[3] The notion of change

---

[1] Richard Rogers, *Doing Digital Methods* (London: Sage, 2019).

[2] E.g., Katherine Hayles, *How We Think: Digital Media and Contemporary Technogenesis* (Chicago, IL: University of Chicago Press, 2012).

[3] Kenneth M. Price and Ray Siemens, eds., "Introduction," in *Literary Studies in the Digital Age* (New York: Modern Language Association of America, 2013), https://dlsanthology.mla.hcommons.org.

---

that Price and Siemens identify is intertwined with two important aspects, namely: how a specific material is presented and how this material might be studied with the help from digital technologies. Expressed differently, digitized or born-digital material potentially create, as stated by Price and Siemens, "new" research approaches where certain digital traits are addressed and analyzed.

The descriptions of the benefits of using digital methods are often associated with visual metaphors by for example emphasizing the importance of gaze, scale, or scope.[4] The "altered methodologies" that Price and Siemens describe are also framed as presenting the researcher with new perspectives through interfaces, functions, and applications that the digital methodological tools can utilize. In a similar visual description, Ted Underwood underlines the significance of digital methods as a way of striving to overcome the distance between researchers and research material.[5] Important benefits of these methods are that they, according to Underwood, present new ways of viewing research material. Underwood uses the following formulation to describe their potential: "a single pair of eyes at ground level can't grasp the curve of the horizon, and arguments limited by a single reader's memory can't reveal the largest patterns organizing literary history."[6] For Underwood, digital methods facilitate the researcher in presenting an unbiased overview of a vast material that might otherwise be impossible for an individual to get a hold of and see.[7]

The current discussions of new and digital methodologies have common traits across the social sciences and humanities, in terms of how the research process is affected by adopted methods. Digital methods have thus given scholars new vocabularies and new ways to grasp bigger fields of view.[8] While this is a positive

---

[4] E.g., Katherine Hayles, *How We Think*, 27–31.

[5] Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change* (Chicago, IL, London: University of Chicago Press, 2019).

[6] Underwood, *Distant Horizons*, x.

[7] Cf. Alan Liu, "Theses on the Epistemology of the Digital: Advice for the Cambridge Centre for Digital Knowledge," Alan Liu's website, August 19, 2014, https://liu.english.ucsb.edu/theses-on-the-epistemology-of-the-digital-page.

[8] Cf. Underwood, *Distant Horizons*.

development for technological advancement in social scientific and humanistic methodology, the use of these methods can also be criticized because some processes of knowledge-making may be hidden or made invisible through the utilization of digital methods. From our point of view, methodical approaches and their impact on the scientific process need further investigation not just by focusing on the descriptions and discussions as such in relation to research processes during the 21st century but rather to explore how the methods in use in different ways can be applied when studying and analyzing digitized or born-digital material.

Our aim with this contribution is therefore to discuss how digital methods can serve to provide the researcher with a *teleoptical perspective* on a certain digital material and explain how this perspective affects the researcher's gaze. In this chapter, we intend to investigate how the selection of certain digital methods and perspectives impact the research process, from data collection to selection, analysis, and interpretation.[9] The investigation is based on three case studies where different types of digital material are studied and analyzed according to the following questions:

- How does a selected digital method impact views of a specific body of material?
- What are the methodological consequences of these views?
- How can the concept of teleoptical perspective be used to understand and interpret the limitations and possibilities of digital methods for data collection and analysis?

A Teleoptical Perspective on Digital Methods

Before applying the analytical term *teleoptical perspective*, a short definition is needed. Following the abovementioned visual traits that can be found in the discussions of digital methods and the consequences for the individual researchers when applying them

---

[9] Our focus is not to analyze or discuss the research outcomes such as visualizations. For readers interested in an overview of the analytical vocabulary of information visualization, see Karolina Uggla's chapter in this anthology.

in a specific study, the teleoptical perspective that we propose in this text can be seen as an overarching analytical term. By using the term it is possible for the researcher to make visible and underline the potential relations between the researcher and the selected digital methods. Our purpose with this text is also to argue for an approach to that makes the uses of digital methods more transparent for the research community.[10]

The *teleoptical perspective* can also be seen to work on a more metaphorical level, as a way to formulate and grasp different processes and perspectives attached to various digital tools or methods. As noted by George Lakoff and Mark Johnson, a conceptual metaphor allows for understanding a concept through senses and physical experience.[11] Moreover, the use of a metaphor makes it possible to make abstract theorizing more specific and understandable.[12] In this sense it is therefore possible to think of the teleoptical perspective as a mode of analysis that works both on a concrete methodological level, and makes it possible for the researcher to draw theoretical and methodological conclusions of a conducted study.

In this chapter we also present subterms such as the concepts of *instrument*, *observation*, *lens*, and *filter* as an attempt to grasp the teleoptical perspective afforded by specific selected digital methods. The instrument is in this context defined as the tool through which data is collected and/or analyzed; the observation is what the researcher chooses to investigate through the instrument, such as specific sets of data and/or metadata. The lens is defined as a specification of the dataset, which determines whether to add or subtract data and/or metadata. The filter is a final calibration of specific aspects of the collected data. The filter can be utilized in different parts of the research process, initially to specify the data

---

[10] For a discussion of the critique on how digital methods sometimes can be opaque for others, see Katherine Bode, *A World of Fiction: Digital Collections and the Future of Literary History* (Ann Arbor, MI: University of Michigan Press, 2018), 5.

[11] George Lakoff and Mark Johnson, *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought* (New York: Basic Books, 2010), 45.

[12] Lakoff and Johnson, *Philosophy in the Flesh*, 128.

collection (e.g., through search terms, file types, hashtags), and later on to further specify which data to focus on (e.g., searching within the collected data). Important to note is that all the steps depend on what the researcher aims to do with the data, as well as what is possible to do within the platform studied.[13] The different functions of the concepts, as well as what dimensions they cover, are illustrated by Table 1 below.

The chapter is based on three empirical case studies, which together investigate different aspects of our research questions. The first example discusses how automatic collection of data and metadata from Instagram can be understood using a teleoptical perspective. The second example shows how Twitter conversations concerning vaccination can be viewed from different perspectives, focusing on what aspects can be hidden by the application programming interface (API) that is used to collect data, and how these issues can be solved.[14] The third example deals with what kind of themes and terms are used in published articles from the scientific journal *Digital Humanities Quarterly* during the years 2007–19. In the two first examples, data is collected from social media platforms, hence born-digital material, while the third focuses on digital material published as open-access scientific journal articles. These three cases together demonstrate how digital technology can be viewed as both user-generated content

Table 1: The teleoptical metaphor as a tabulated model.

Instrument	Observation	Lens	Filter
Tool for data collection and/or analysis	Data and/or metadata observed	Command/ option for specifying observed data/ metadata	Command/ option for filtering specified, observed data/ metadata

---

[13] What kind of research endeavours the platform affords might be different from time to time; cf. Tommaso Venturini et al., "A Reality Check(List) for Digital Methods," *New Media & Society* 20, no. 11 (November 1, 2018): 4195–4217, https://doi.org/10.1177/1461444818769236.

[14] Essentially, an API for this end is a set of rules and access points to different types of platform data, such as tweets, users, trends, etc.

and editorial-controlled texts that are selected and curated by platform owners and editors, respectively.

Instagram through the Lens of a Scraper

Studies of activity on Instagram can benefit from a critical perspective on automatic data collection to illuminate an otherwise black boxed procedure of collecting empirical data. This process of bulk data collection is often called web scraping. Web scraping should here be understood as "systematic collection of a specified type of data [from] websites[; t]o use a scraper means to start a little robot looking for a specific piece of code or a specific form of content, automatically archiving it in a dataset."[15] There are several methods for conducting this type of data collection. In this section, we will examine the web scraping method with a command-line application Instagram Scraper.[16] As a case example, the literary access and preservation project the Swedish Literature Bank (Swedish: Litteraturbanken) is investigated. The Swedish Literature Bank is a cooperative project by a number of literary and cultural heritage institutions—among them the Swedish Academy, the National Library of Sweden and the Society of Swedish Literature in Finland—providing digital editions of literary classics in the Swedish language as e-texts, facsimiles, and PDF or EPUB files.[17] For outreach purposes, the material provided by the Swedish Literature Bank is promoted through the project's Instagram account.[18] Using Instagram Scraper, the Swedish Literature Bank's outreach activity is examined and discussed in line with a critical perspective to investigate possibilities and limitations for web scraping social media metadata.

Instagram Scraper allows the automatic data collection of media files and metadata of single Instagram users, a list of such

---

[15] Andreas Birkbak and Anders Kristian Munk, *Digitale Metoder* (Copenhagen: Hans Reitzels Forlag, 2017), 93 (our translation).

[16] Richard Arcega, "Instagram-Scraper: Scrapes an Instagram User's Photos and Videos," GitHub website, https://github.com/rarcega/instagram-scraper.

[17] Litteraturbanken website, "Om Litteraturbanken," https://litteraturbanken.se/om/english.html.

[18] Litteraturbanken on Instagram, https://www.instagram.com/litteraturbanken.

users, or one or several hashtags. Instagram Scraper does not have a graphical user interface but rather relies on the user's familiarity with a command prompt or a terminal window. The options at hand within the command-line interface of Instagram Scraper can be understood as lenses or filters, since they allow the user to select which data to access. Instagram Scraper consists of a total of 25 flags, which can be understood as subcommands.[19] These can be used to specify, for example, the Instagram user through which the scraping operation will be processed, the media types (images, videos, stories) to scrape, whether or not profile metadata should be collected, and whether or not media metadata should be scraped, in JavaScript Object Notation (JSON) format.[20]

The following terminal command (Figure 1) searches images uploaded by the Swedish Literature Bank. It also generates a separate file with post metadata. At the time of scraping, the total number of Instagram posts by the Swedish Literature Bank adds up to 558.[21] The command returns 553 posts—including only image uploads—as well as associated metadata.

The terminal command can be broken down into the following concepts using the teleoptical metaphor (Table 2).

Figure 1: A terminal command for scraping media and metadata from the Swedish Literature Bank's Instagram account through Instagram Scraper. Copyright: Authors. License: CC BY 4.0.

Table 2: The terminal command described using the teleoptical metaphor.

Instrument	Observation	Lens	Filter
Instagram scraper	Litteraturbanken	Media/metadata	Media type: image

---

[19] Arcega, "Instagram-Scraper."
[20] A format commonly used to serialize data, for instance in an API.
[21] November 14, 2019.

In the example command above, Instagram Scraper is the teleoptical instrument, while the following string—in this case the handle *litteraturbanken*—is the observation. The scholar's attention is thus pointed toward the digital surrogates of media uploaded to the Swedish Literature Bank's Instagram account. Additional flags added to this command can be understood as including or excluding data in terms of lens and filter perspectives. One lens is in use, namely "—media-metadata." This indicates that the lens is set to collect, apart from the media files, additional metadata associated with all of the Swedish Literature Bank's Instagram posts. These metadata is saved to a file titled "Litteraturbanken.json" within an automatically created folder titled "Litteraturbanken." The metadata retrieved includes information related to, for example, the types of media, the dimensions of media files, web links to the media file on Instagram's servers, the number of likes, the text written as a description for the media, any hashtags used for a post, timestamps, and view counts (if the media file is a video). The flag "—media-types image" command can be seen as a restrictive filter that tells Instagram Scraper to only search for media in image format. Additional lenses and filters can be combined and added to the data collection command through the use of flags and further zooming in or out on the observations.[22]

An excerpt of the metadata collected from the larger number of retrieved Instagram posts uploaded by the Swedish Literature Bank can be seen in the JSON file in Figure 2 below. Orange-colored text are keys and purple-colored text are values.

Figure 2 provides the metadata for a certain Instagram image post through key-value pairs. For example, line 272 describes the type of media of the post, namely a graphic image. Lines 275–276 depict the dimensions of the image uploaded in the post, namely 937 × 750 pixels. Line 278 returns the display URL (uniform resource locator, i.e., the web address) for the image uploaded. Line 280 returns the amount of likes for the post; in this case, the post has generated 53 likes. Line 286 provides a representation of the full text description accompanying the image upload. Line 292 displays the number of comments this particular post has

---

[22] Arcega, "Instagram-Scraper."

```
271
272            "__typename": "GraphImage",
273            "comments_disabled": false,
274            "dimensions": {
275                "height": 937,
276                "width": 750
277            },
278            "display_url": "https://scontent-arn2-1.cdninstagram.com/vp/41d467212702f09bd780bdabc694cb6c/5E680190/t51.2885-15/e35/
        67946139_168262120966347_3853995595589639362_n.jpg?_nc_ht=scontent-arn2-1.cdninstagram.com&_nc_cat=103&ig_cache_key=MjExNTUyNDM3NTA0OTk2NDU0Mw%3D%3D.2",
279            "edge_media_preview_like": {
280                "count": 53
281            },
282            "edge_media_to_caption": {
283                "edges": [
284                    {
285                        "node": {
286                            "text": "Dagens boktips från Litteraturbankens redaktion är Kerstin Söderholms diktsamling Mörkret och människan från 1941.
        \n#lbsommarlästips"
287                        }
288                    }
289                ]
290            },
291            "edge_media_to_comment": {
292                "count": 0
293            },
294            "gating_info": null,
295            "id": "2115552437504996456",
296            "is_video": false,
297            "media_preview": "ACEqG88bQSSo9qiLoP4j09+nrSSO/K7dyjHrz/nrTNMJXZwMgdcEHP5Z6erNSA/XA43NkcfX0z6VYQccEnPPPvVTBXDE/nzxj+NAf/wBVTRuxONuFx1/L/
        wCuPw96AJ6KXKKTg6VQecICxM3D1eoH9Owq+Olc51jPIwxxS7c/gBTAvLeLJXiozcnkdiWHvitasC2WZkjwuH3HPtgHr9a36AEoooopAA6VlPBIXYgAD0RyOTnqff0rVHSkz+ph5SUWWuO3kEisQMA/
        3hx/iT/U1rUmxeuB+VLTASiiipA//Z",
298            "owner": {
299                "id": "2206796874"
300            },
301            "shortcode": "B1b2jXoiIwT",
302            "tags": [
303                "lbsommarlästips"
```
271,9   0%

Figure 2: Metadata excerpt in JSON from a single Instagram post uploaded by the Swedish Literature Bank. Copyright: Authors. License: CC BY 4.0.

received, in this case none. Line 303 returns the hashtag that have been used in the upload for this post: "#lbsommarlästips."[23] This metadata has been collected based on the lens in place, namely media/metadata, and the filter, the image media type. This in turn denotes that all video media are excluded from the data collection, as the teleoptical filter is not focused on this particular media format. This extracted metadata can in turn be used as a basis for data analysis.

Thus, Instagram as a platform provides access to users' media uploads as well as metadata, collected through Instagram Scraper using a series of commands with additional flags specifying which data to obtain. Altering the data and metadata itself is not possible for the researcher. However, the options through which additional metadata and information can be chosen, added, ordered, combined, and filtered does indeed provide these types of choices through which the scope of the data and the collection procedure can be controlled and refined. Some of these flags are very much like tick boxes in a graphical user interface. By providing the flag

---

[23] Translation from Swedish: Summer reading tips from the Swedish Literature Bank.

"—media-metadata," the user metaphorically *ticks* a box telling the program to include media metadata and save it to a designated file when performing the scraping process.

Instagram Scraper hence places a perspective on the user's uploaded media files as well as the metadata assigned to either these media files or the user's profile as a whole. The inclusion of specific lenses enhances certain observations on a more granular (i.e., metadata through the command "—media-metadata"), while other lenses enhance other details. Certain lenses and filters enable the collection of certain data but limit or obscure other kinds of data. Acquiring datasets from this perspective facilitates qualitative analyses through investigations of the content of visual data as media files and textual data as media file descriptions, user biographies, hashtags, and comments, as well as a focus on material features.[24] Quantitative analyses can, for instance, include natural language processing techniques for topic modeling post descriptions in order to understand the text that accompanies image or video uploads.[25] Statistical analyses are also made possible through counting and visualizing the number of likes or view counts. Furthermore, computer vision techniques can be applied to the media files, which also provides a way to automatically identify objects in the Instagram media uploads.

Limitations to data collection occur as Instagram Scraper does not collect data on single users' practices like clicking, following, searching, scrolling, liking, commenting, previewing, sharing, bookmarking, tagging, reporting, searching, or discovering content. The tool also does not provide information about other users, tags, followers, or other forms of interaction. The teleoptical view is thus limited to providing data and metadata on users' media, rather than a full perspective on practices within Instagram. Practices related to the reporting of media deemed

---

[24] E.g., Janet Vertesi and David Ribes, eds., *Digital STS: A Field Guide for Science & Technology Studies* (Princeton, NJ: Princeton University Press, 2019); Paul Dourish, *The Stuff of Bits: An Essay on the Materialities of Information* (Cambridge, MA: MIT Press, 2017).

[25] E.g., Ted Underwood, "A Genealogy of Distant Reading," *Digital Humanities Quarterly* 11, no. 2 (June 27, 2017).

inappropriate fall out of the context as well, as the practices are hard to trace, and reported media deemed inappropriate by Instagram are supposedly blocked and removed from any web servers. In what follows, the teleoptical perspective will be further discussed in terms of Twitter conversation and retweet networks.

Filtering Twitter for Conversations and Patterns

Whereas Instagram is centered on visual media, their metadata, and image-based social interaction, Twitter emphasizes publicly visible individual messages and the conversation that these generate. When investigating Twitter, collecting data is arguably the most challenging part of the research process.[26] This undoubtedly makes choices of instruments and selection of suitable lenses and filters crucial. Technically speaking, the Twitter Streaming API pushes a stream of tweets for a client to filter in real time,[27] where a filter specifies what content to collect, for example through keywords or hashtags.[28] The chosen keywords can be seen as representing the observation, for example a topic. First, there are semantic challenges associated with the choice of keywords, because words and hashtags might represent different topics or aspects than those the researcher intended to study. Second, different qualities of the topic can be studied through the selection and combination of filters. Such qualities may include communication patterns, trends, and conversations. If the aim is to study conversations, a challenging issue arises as a consequence of the affordances of the API.

Following a conversational thread as it evolves is challenging because not all tweets in a thread include the selected keyword. A relevant option to consider is the user filter, which collects

---

[26] David Gunnarsson Lorentzen and Jan Nolin, "Approaching Completeness: Capturing a Hashtagged Twitter Conversation and Its Follow-On Conversation," *Social Science Computer Review* 35, no. 2 (April 1, 2017): 277–286, https://doi.org/10.1177/0894439315607018.

[27] The Streaming API does not offer an archive to search in, but through the Search API it is possible to retrieve tweets that are a few days old.

[28] The *track* parameter in the API, which can contain up to 400 keywords. The tweet is collected if it includes either of the selected keywords.

tweets if a selected user is either the sender or recipient.[29] However, not all tweets in the thread are posted by or to a user in the filter. Hence, choosing one filter leads to incomplete datasets, as only a certain set of objects are visible to the researcher (e.g., keyword-matching tweets). A proposed solution to this problem is to combine two filters, starting with a static set of hashtags and—as tweets are collected—build a dynamic list of the most active users in the dataset.[30] By matching tweet content with the list of keywords and the "in_reply_to_status_id_str" metadata field with tweet IDs in the database, both keyword-matching tweets and replies (with or without keywords) can be collected using continuous scripts.[31]

In what follows, data collection and selection will be illustrated through conversations around the topic of vaccination (Table 3). The first step is data collection through the combination of a keyword and a user filter.[32] Thirty-six keywords related to vaccination were selected as the first filter and, as data was collected, the 2,000 most active users over the last 48 hours were selected as the second filter.[33] The sizes and contents of the filters decide to what extent it is possible to study the intended object of research, in this case the conversations around the topic. If too much content has been produced, the user filter might not cover all users in the dataset that have been replied to. If the user filter contains too many highly active or popular users, there is a risk that an API cap will be reached or the computer may not be able to keep up with the stream.[34] The lenses represent what is focused on within the

---

[29] The *follow* parameter in the API, which can contain up to 5,000 user IDs. The tweet is collected if the selected user is the sender or receiver.

[30] Lorentzen and Nolin, "Approaching Completeness: Capturing a Hashtagged Twitter Conversation and Its Follow-On Conversation."

[31] Note that the Twitter API is currently under development. One consequence is that conversations will be easier to collect as it is possible to search with a conversation ID. However, this does not mean that conversations are automatically collected through the API. The researcher still needs a filter for keyword matching and a filter for the conversation.

[32] Data collected November 19–26, 2018.

[33] The user filter was updated programmatically every 120 seconds.

[34] The API cap is set at 1% of all tweets produced at any given time.

Table 3: The demarcations and filters for conversation-oriented data collection.

Instrument	Observation	Lens	Lens	Filter	Filter
Custom made Python application based on Tweepy	Vaccination as a topic on Twitter, November 19–26, 2018	Micro level (conversations)	High detail level (metadata captured)	#antivaccination, #vaccine, #vaccination[35] (static)	List of 2,000 user IDs (dynamic)

A   B
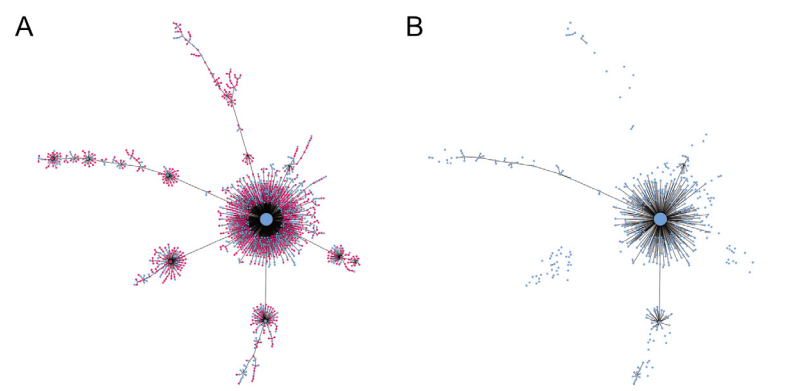
Figures 3A–B: Two conversation networks where each dot is a tweet and the connection between two dots represents a reply. A: The full conversation. B: The part of the conversation where all tweets include either of the selected keywords. Replies without keywords in magenta, tweets with keywords in blue. Copyright: Authors. License: CC BY 4.0.

observation, in this case conversations about vaccination, guiding which metadata tied to tweets and users is captured.

The differences between the keyword and conversation-based approaches are illustrated in Figures 3A and 3B. In each of the figures, one tweet is represented by a node and edges are drawn between one tweet that is replying to another tweet. Figure 3A shows the structure of one conversation including all collected tweets in the thread, that is, what is collected using both

---

[35] Only a subset of the used keywords is included in the table.

filters combined. Figure 3B represents the same conversation but includes only tweets matching keywords. The fully connected conversation stretches longer and includes far more tweets, 1,416 compared to 509 (395 of them connected) in the second example. In the total dataset, 43% of the 302,328 tweets did not include a selected keyword, and 97 conversations with at least 100 tweets were found. The implications of these data qualities are that, when the contents and dynamics of the conversations are studied, the keyword-based approach would omit large parts of the discussions, including reactions to the collected tweets, and that the analysis of the topic would be biased toward the selected keywords. Additionally, the contexts of the disconnected tweets are unknown without access to the rest of the conversation.

As a sample, the 172,327 tweets collected with the keyword filter might constitute a substantial body of data, and the structure in Figure 3B represents a conversational thread as there are tweets connected as replies. When the contents of the tweets are analyzed, the researcher might be given the false impression of completeness, that is, that the data represented in the study is all data connected to the topic of vaccination. However, it is important to note that, due to the abovementioned issues regarding filter options, incomplete retrieval of tweets during data collection is still a risk, even though two filters have been used.

To highlight the necessity of using the two filters when studying conversations on Twitter, an analogy can be drawn between Twitter and a discussion forum. In the latter, the context of the posts is clearly visible, making it possible to analyze what the discussion is about. The affordances of the Twitter API make it challenging to collect tweets in their contexts. The example illustrates that, when viewing activity through different filters, a more complete picture of the conversational structure appears.

The conversation-based analysis enables a microlevel focus. Another way to analyze Twitter data is to focus on macrolevel patterns. In the following example, a dataset collected for the purpose of outlining communication patterns over time was utilized. This observation was made over eight weeks, when the instrument was programmed to search for tweets matching a set of keywords

Table 4: Demarcations and filters for macrolevel data collection.

| Instrument	Observation	Lens	Filter
Custom Python application based on Tweepy	Vaccination as a topic on Twitter, eight weeks starting at January 14, 2019	Macro level, retweet patterns	Keywords: vaccine, vaccination, protective immunity, immune responses, immunogenicity

(Table 4).[36] Here, the first step is to create a zoomed-out overview of the communication patterns.

Here, the topic is represented through a series of retweet networks, in which each node is a user. A connection is drawn between two users if one redistributes the other user's content. Retweet networks tend to show clustering tendencies in topics that are polarized.[37] This makes them suitable for analysis of potential controversies, as users with similar viewpoints are likely to be grouped in clusters.[38] Through the use of the visualization application Gephi, networks were created using the force-directed layout algorithm *Force Atlas 2*. Figure 4 shows focused views of the observation where the largest connected part of users with at least two connections are kept.

---

[36] Only a subset of the keywords used is included here.

[37] Michael D. Conover et al., "Political Polarization on Twitter," in *Proceedings of the Fifth International Conference on Weblogs and Social Media* (Menlo Park, CA: AAAI Press, 2011), 89–96; David Gunnarsson Lorentzen, "Polarisation in Political Twitter Conversations," eds. Axel Bruns and Katrin Weller, *Aslib Journal of Information Management* 66, no. 3 (January 1, 2014): 329–341; Leo G. Stewart, Ahmer Arif, and Kate Starbird, "Examining Trolls and Polarization with a Retweet Network," in *Proceedings of WSDM Workshop on Misinformation and Misbehaviour Mining on the Web (MIS 2)* (New York: ACM, 2018), unpaginated, https://faculty.washington.edu/kstarbi/examining-trolls-polarization.pdf.

[38] Noortje Marres, "Why Map Issues? On Controversy Analysis as a Digital Method," *Science, Technology & Human Values* 40, no. 5 (2015): 655–686.
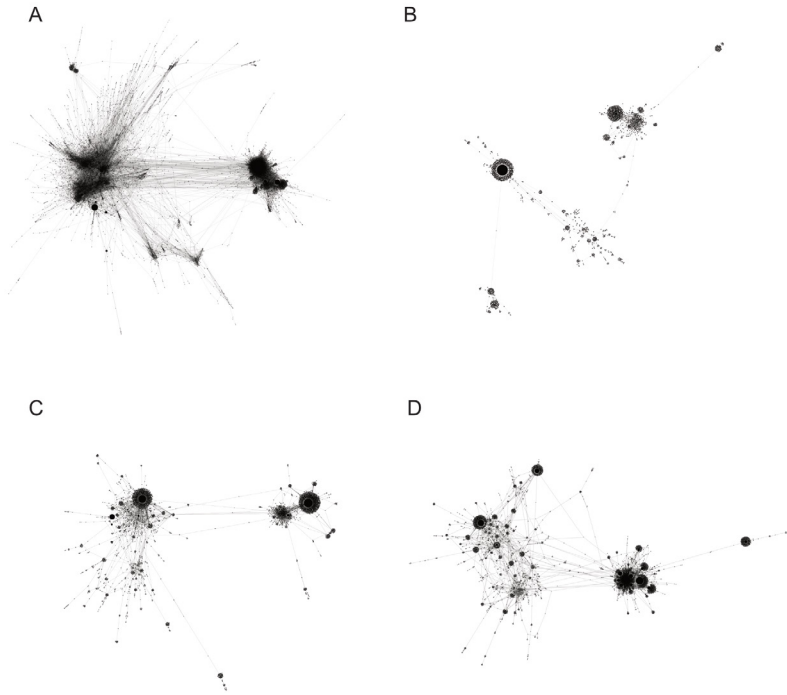
Figures 4A–D: Four filtered views within the vaccination retweet observation, each representing a time-delimited slice of the retweet connections among users. A: aggregated network representing all eight weeks. B: week two. C: week five. D: week eight. Copyright: Authors. License: CC BY 4.0.

Figure 4A represents an aggregation of all eight weeks. It includes 25,157 users sharing 46,336 connections and has two large clusters separated by what appears to be a wide bridge. At the fringe areas, some medium-sized clusters appear. Focusing on weeks two, five, and eight (Figures 4B, 4C, and 4D), various patterns appear. Whereas the networks from weeks five and eight have similar polarized structures, week two does not. It consists of a few loosely connected islands, one of which includes a heavily retweeted user that is barely involved in the topic during the rest of the period. This network is also the smallest of the three, with its 2,414 nodes and 2,982 connections (week five: 4,745 and 6,314; week eight: 5,982 and 8,682). While the aggregated network shows clear clustering tendencies indicating that vaccination

is a polarized topic, a one-week sample of Twitter activity within a given topic does not capture the characteristics of the topic on its own. The researcher thus needs to collect data for a longer time period and analyze the aggregated view as well as the individual views to capture polarizing tendencies, if any, as well as structural dynamics of the topic. Differences in characteristics regarding for example the number of participants, communication patterns, and network structures illustrate the problem of analyzing a snapshot, which in itself might not represent the activity particularly well.

Summarizing this section, it is important to address the issue of what these kinds of figures represent, when each figure represents a different perspective, rather than a whole. Through digital methods and visualization tools, different views can be presented, each of them giving a picture of a specific set of activities. By compiling several views side by side, a more representative picture of the discussions appears. Seemingly, the two large clusters linked with a wide bridge (Figure 4A) indicate different viewpoints colliding, which is a setting in which one could expect conversations to evolve. By following up the structural analysis with a qualitative reading of the tweets, representations of viewpoints can be identified. However, learning about the actual viewpoints requires different methods, such as interviews with the users. Content analysis of threads, both quantitative and qualitative, are recommended to understand the conversation dynamics. For large volumes of tweets, other relevant methods to consider are identification of linguistic patterns and topic modeling, which will be discussed in the next section.

Observing and Analyzing Patterns in *Digital Humanities Quarterly*

In this section, an investigation of the articles published in *Digital Humanities Quarterly* (*DHQ*) will be presented. *DHQ* is not a social media platform like Twitter and Instagram but rather a scientific online journal where the published articles have undergone a peer-review process and after that been published by the editors. By scraping, collecting, compiling, and visualizing data using a variety of digital methods, it is possible to investigate what themes, topics, or specific terms are covered in the articles. This

third case study will explore the extent to which digital methods can be helpful in finding and charting patterns in all 308 articles published in *DHQ* during the years 2007–19.[39] *DHQ* is an "an open-access, peer-reviewed, digital journal covering all aspects of digital media in the humanities."[40] The journal is also related to and associated with the field of digital humanities and, in the section *DHQ on Digital Humanities*, it is stated that digital humanities can be defined as:

> a diverse and still emerging field that encompasses the practices of humanities research in and through information technology, and the exploration of how the humanities may evolve through their engagement with technology, media and computational methods.[41]

This last example will focus on how themes, terms, and topics found in published articles in *DHQ* can be said to relate to the journal's own definition of digital humanities, as an intersection of technology, computer-assisted methods, and/or media.[42] This is done in two steps, the first being to choose a suitable instrument for collecting the chosen material for observation, in this case the XML editions of the published articles. The second step is to compile the material into one text corpus and then to choose a digital method for charting and visualizing the themes and topics present in the material. These steps of the research processes and which instruments, lens, and filters have been used are presented in the table below (Table 5).

When the material is collected and compiled, it is possible to use various computer-assisted tools to visualize themes, subjects, frequencies, and relations between terms by looking at specific words or concepts. In the figure below (Figure 5), the material has been visualized through Voyant Tools, "a web-based text reading and analysis environment" suitable for working with text collections in different formats.[43] As for processing the material, it is

---

[39] The investigated texts are published under the section "Articles," and hence reviews and editorial texts are not included in the selection.

[40] *Digital Humanities Quarterly* website, "About DHQ," January 6, 2020, http://www.digitalhumanities.org/dhq/about/about.html.

[41] *Digital Humanities Quarterly* website, "About DHQ."

[42] *Digital Humanities Quarterly* website, "About DHQ."

[43] Voyant Tools website, "Getting Started," https://voyant-tools.org/docs/#!/guide/start.

Table 5: Description of the collection, selection, and specification of the investigated dataset.

Instrument	Observation	Lens	Filter
Python-Library Beautiful Soup	XML editions of *Digital Humanities Quarterly* journal articles	Journal issues between 2007, vol. 1, no. 1 and 2019, vol. 13, no. 2	\<p> elements within \<body> element

also possible to adjust the stopwords through a stopword list, which will exclude common words. In Figure 5, the words *digital* and *humanities* were added to the stopword list, among other common words, and are thus not included.

The visualization below presents an overview of the most frequently occurring words in the articles, such as "text," "new," "work," "research," "data," and "information." The fact that these words can be found in the material is no surprise, since the scientific discussion of the digital humanities as a discipline is often described as an intersection between humanities and digital technology, and hence a combination of text, data, and information.[44] As pointed out by Todd Presner et al., digital humanities can be seen as a universe where print-based media are no longer "the normative medium"[45] for knowledge production or for the object of study. In the figure above, words related to print-based media formats such as "book," "codex" and "print" do not appear at all. Instead words like "text," "information," and "work" appear. By zooming in on the smaller words, for example "model," "project," "process," and "tools," the compiled material also presents a picture of what Matthew G. Kirschenbaum identifies as a "common methodological outlook" that characterizes research within the

---

44 E.g., Patrik Svensson, "The Landscape of Digital Humanities," *Digital Humanities* 4, no. 1 (2010): http://digitalhumanities.org/dhq/vol/4/1/0000 80/000080.html.

45 Todd Presner, Jeffrey Schnapp, and Peter Lunenfeld, "The Digital Humanities Manifesto 2.0," (2009) 391.Org blog, June 22, 2009, https://391 .org/manifestos/2009-the-digital-humanities-manifesto-2-0-presner -schnapp-lunenfeld.

Figure 5: Word cloud showing the most significant words in articles published in *Digital Humanities Quarterly* 2007–19. Copyright: Authors. License: CC BY 4.0.

digital humanities.[46] These words can also be said to underline the research process in academic investigations that have an affinity with the digital humanities and can also contribute to a larger discussion about how studies in the digital humanities should be conducted and understood.

It is, however, possible to problematize the view of the material given above. Figure 5 can be used as an illustration of significant words in the material, but the visualization does not reveal potential relations between the words. To be able to grasp the connections between the terms, topic modeling can be useful. A topic can be described as "a collection of words that have different probabilities of appearance in passages discussing the topic."[47] By choosing 100 topics and the 10 most significant words in every topic, it is possible to chart and explore the potential connections between the words. To summarize the result of the topic

---

[46] Matthew G. Kirschenbaum, "What Is Digital Humanities and What's It Doing in English Departments?" *ADE Bulletin* 150 (2010): 55–61, 56, https://doi.org/10.1632/ade.150.55.

[47] Ted Underwood, "Topic Modeling Made Just Simple Enough," The Stone and the Shell blog, April 7, 2012, https://tedunderwood.com/2012/04/07/topic-modeling-made-just-simple-enough.

modeling accuracy, we have selected a couple examples of the topics revealed below:

- Topic 4: http, archive, search, library, users, resources, scholarly, collection, metadata, user.
- Topic 26: technology, users, word, tool, figure, user, objects, textual, http, visual.
- Topic 46: visualization, design, literary, visual, users, space, user, image, images, narrative.
- Topic 61: projects, search, archive, corpus, users, literature, literary, national, google, objects.
- Topic 89: space, infrastructure, landscape, humlab, visualization, cyberinfrastructure, technology, science, past, design.

To be able to explore the relations between the topics and the words in each topic, visualizing using Gephi can be useful. This step can be described with a table that in detail specifies the teleoptical perspective of the digital method in use (Table 6).

The result of these steps is illustrated by the network graph below (Figure 6). The figure shows the relations and positions between words/terms/concepts accessed through the compiled material of the scientific journal *DHQ*.

Through the color coding of word co-occurrences, it is possible to get a chart of connections between the identified topics. In that

Table 6: Description of the visualization, selection, and specification of the investigated dataset.

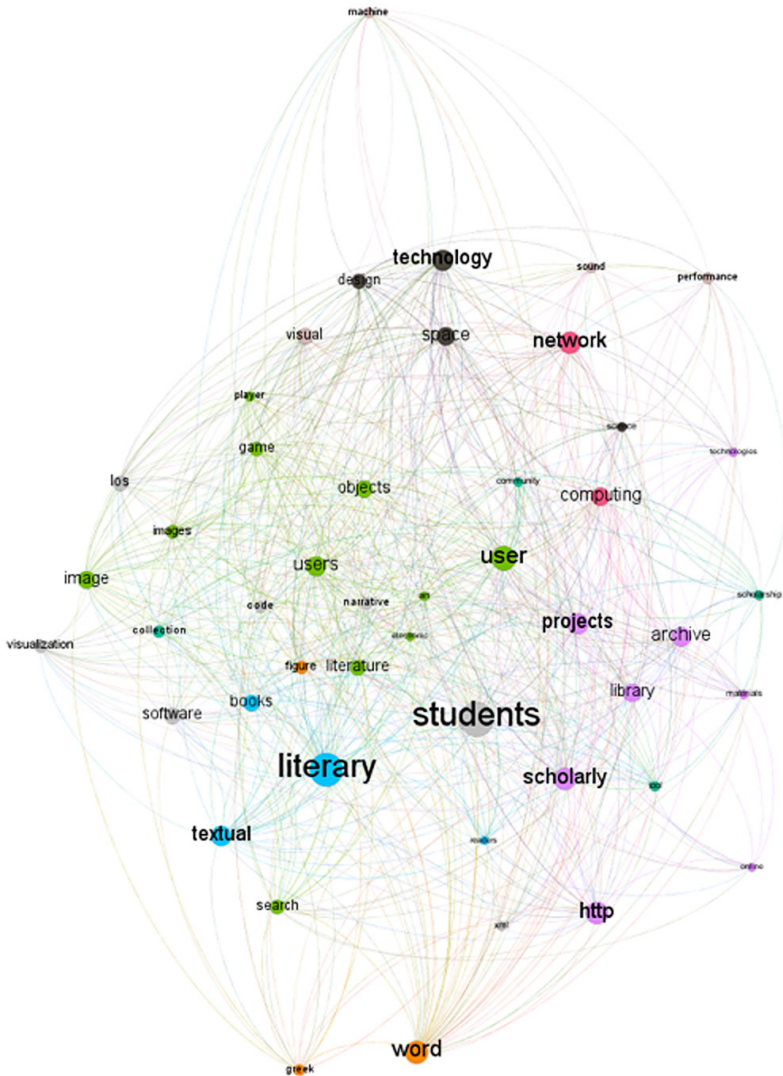| Instrument	Observation	Lens	Lens	Filter
Gephi	Topic modeled *Digital Humanities Quarterly* journal articles	Force Atlas 2 visualization algorithm	Word co-occurrence clustering through modularity function	10 most significant words from 100 topics. Words occurring in at least five topics are included in the visualization.

Figure 6: Relations and positions between words in published articles in *DHQ* 2007–19. Copyright: Authors. License: CC BY 4.0.

sense, the visualization in Figure 6 presents an overview of proximity conditions between words by, for example, locating words like "literary" (in green) and "students" (in gray) in relation to one another. However, the network graph also shows where there is no connection between the different topics and where the topics remain intact, as for example with the black topic covering words

like: "machine," "technology," "design" and, "space." As a result, the network graph gives the researcher significant clues about the different relations between topics in the compiled material.

In a more general perspective, the investigated articles from *DHQ* are aligned with the journal's own definition of the digital humanities as an "engagement with technology, media and computational methods."[48] However, the visualizations above also make it possible to observe that the words, themes, and topics that are covered in the compiled material also show how "digital humanities hardly make up an uncontested or well-defined landscape."[49]

By investigating *DHQ* through different digital tools, it is possible to collect, compile, and visualize different aspects of the themes, terms, and words used in the scientific articles published in the journal. This can be done through various different digital methods and visualization tools that illustrate different aspects of the material. What these figures, however, do not reveal is how a specific word in a topic can be interpreted and analyzed more closely, within its full textual context. The teleoptical view as an observation across a given timespan is, similarly to the investigations of Instagram and Twitter, limited, because it fails to provide a close analysis of how words are used within their specific contexts, within individual articles.

Discussion and Conclusions

In this chapter, we explored if and how a teleoptical metaphor can be used as an analytical term for understanding and illuminating how a certain digital method can affect the steps of a research process as well as the results of a certain study. This was done by three case examples where concepts such as *observation*, *instrument*, *lens*, and *filter* were used in the processes of collecting, visualizing, and analyzing born-digital or digitized material. Our investigations were conducted through scraping the Instagram account of the Swedish Literature Bank, visualizing vaccine

---

[48] *Digital Humanities Quarterly* website, "About DHQ," January 6, 2020, http://www.digitalhumanities.org/dhq/about/about.html.

[49] Patrik Svensson, "The Landscape of Digital Humanities."

discussions on Twitter, and topic modeling and visualizing journal articles in *Digital Humanities Quarterly*. In what follows, these investigations are discussed further, in a wider context, with regard to the initial research questions of *how* digital methods alter scientific research practices and *what* the consequences are for employing certain digital methods.

The teleoptical view of scraping data and metadata from Instagram can be understood in operative terms through Instagram Scraper.[50] The scope is aligned, lenses and filters are fixed, and observations can be sighted depending on how these lenses and filters have been determined. Understanding web scraping as a digital method in teleoptical terms makes automatic data collection understandable, but also problematized with regard to which phenomena can be sighted. This method enables, for instance, distant reading of Instagram content, including descriptions, comments, or hashtags, which can help a researcher to understand word choices, grammatical patterns, or common topics between one or several Instagram accounts or hashtags. However, practices of—for instance—clicking, following, searching, previewing, and liking content cannot be understood solely through the utilization of Instagram Scraper, and therefore reveal a limitation of the web scraping method. The Twitter example shows that one filter gives a partial view of the context of the topic. The combination of two filters gives access to discussions connected to a certain topic. Another aspect relates to what the studied view actually represents. As with the Instagram example, it is important to acknowledge the limitations of the methods, and to bear in mind that the contents are representations or manifestations of opinions rather than actual opinions. Asking participants about their opinions and intentions would be one way to ground the results in other data sources.[51] Another issue that the researcher needs to be aware of is that, while large datasets need to be filtered so that they can be studied, the choice of filters influences the findings. Also, focusing on a shorter time period within the dataset will not capture a representative view of a topic. The research process involves multiple choices with regard to scope, filters, and

---

[50] Arcega, "Instagram-Scraper."
[51] Cf. Venturini et al., "A Reality Check(List) for Digital Methods."

lenses, all of which influence the findings and the scope of those findings. By collecting, organizing, and visualizing the scientific articles published in *DHQ*, it is possible for the researcher to get an overview of topics covered in the journal, as well as how these topics relate to each other. These kinds of overviews can work as a point of departure for a certain study, making it possible to grasp the notion of digital methods as offering a variety of ways to present different views or scopes, compared to closer scrutinization or manual textual analysis. However, in our example, the presented case studies do not cover individual traits in a specific text, nor how a specific article relates to a certain theme or topic.

By applying a teleoptical perspective on the three examples, we have shown how the research process—from selection and collection to analysis of data—can be made transparent. This has been done with the ambition to problematize their application, including how digital methods shape and limit the creation, interpretation, and framing of knowledge. Methodical transparency is key to providing an understanding of the contributions of digital methods as well as to make research processes visible.

Acknowledgments

References

Birkbak, Andreas, and Anders Kristian Munk. *Digitale Metoder*. Copenhagen: Hans Reitzels Forlag, 2017.

Bode, Katherine. *A World of Fiction: Digital Collections and the Future of Literary History*. Ann Arbor, MI: University of Michigan Press, 2018.

Conover, Michael D., Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. "Political Polarization on Twitter." In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 89–96. Menlo Park, CA: AAAI Press, 2011. https://www.aaai.org/ocs/index.php /ICWSM/ICWSM11/paper/view/2847.

Digital Humanities Quarterly website. "About DHQ." January 6, 2020. http://www.digitalhumanities.org/dhq/about/about.html.

Dourish, Paul. *The Stuff of Bits: An Essay on the Materialities of Information*. Cambridge, MA: MIT Press, 2017.

Gunnarsson Lorentzen, David, and Jan Nolin. "Approaching Completeness: Capturing a Hashtagged Twitter Conversation and Its Follow-On Conversation." *Social Science Computer Review* 35, no. 2 (April 1, 2017): 277–286. https://doi.org/10.1177/0894439315607018.

Gunnarsson Lorentzen, David. "Polarisation in Political Twitter Conversations," edited by Axel Bruns and Katrin Weller. *Aslib Journal of Information Management* 66, no. 3 (January 1, 2014): 329–341. https://doi.org/10.1108/AJIM-09-2013-0086.

Hayles, Katherine. *How We Think: Digital Media and Contemporary Technogenesis*. Chicago, IL: University of Chicago Press, 2012.

Kirschenbaum, Matthew G. "What Is Digital Humanities and What's It Doing in English Departments?" *ADE Bulletin* 150 (2010): 55–61. https://doi.org/10.1632/ade.150.55.

Lakoff, George, and Mark Johnson. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books, 2010.

Litteraturbanken on Instagram. "Litteraturbanken.se." https://www .instagram.com/litteraturbanken.

Litteraturbanken website. "Om Litteraturbanken." https://litteratur banken.se/om/english.html.

Liu, Alan. "Theses on the Epistemology of the Digital: Advice for the Cambridge Centre for Digital Knowledge." Alan Liu's website,

August 19, 2014. https://liu.english.ucsb.edu/theses-on-the-epistemo
logy-of-the-digital-page.

Marres, Noortje. "Why Map Issues? On Controversy Analysis as a
Digital Method." *Science, Technology & Human Values* 40, no. 5
(2015): 655–686.

Moretti, Franco. *Distant Reading*. London: Verso, 2013.

Presner, Todd, Jeffrey Schnapp, and Peter Lunenfeld. "The Digital
Humanities Manifesto 2.0 (2009)." 391.Org blog, June 22, 2009.
https://391.org/manifestos/2009-the-digital-humanities-manifesto
-2-0-presner-schnapp-lunenfeld.

Price, Kenneth M., and Ray Siemens, eds. "Introduction." In *Literary
Studies in the Digital Age*. New York: Modern Language Association
of America, 2013. https://dlsanthology.mla.hcommons.org.

Rogers, Richard. *Doing Digital Methods*. London: Sage, 2019.

Stewart, Leo G., Ahmer Arif, and Kate Starbird. 2018. "Examining
Trolls and Polarization with a Retweet Network." In *Proceedings
of WSDM Workshop on Misinformation and Misbehaviour
Mining on the Web (MIS 2)*. New York: ACM, 2018. https://faculty
.washington.edu/kstarbi/examining-trolls-polarization.pdf.

Svensson, Patrik. "The Landscape of Digital Humanities." *Digital
Humanities* 4, no. 1 (2010). http://digitalhumanities.org/dhq/vol
/4/1/000080/000080.html.

Underwood, Ted. *Distant Horizons: Digital Evidence and Literary
Change*. Chicago, IL, London: University of Chicago Press, 2019.

Underwood, Ted. "A Genealogy of Distant Reading." *Digital Human-
ities Quarterly* 11, no. 2 (June 27, 2017). http://www.digital
humanities.org/dhq/vol/11/2/000317/000317.html.

Underwood, Ted. "Topic Modeling Made Just Simple Enough." The
Stone and the Shell blog, April 7, 2012. https://tedunderwood
.com/2012/04/07/topic-modeling-made-just-simple-enough.

Vertesi, Janet, and David Ribes, eds. *Digital STS: A Field Guide for
Science & Technology Studies*. Princeton, NJ: Princeton University
Press, 2019.

Venturini, Tommaso, Liliana Bounegru, Jonathan Gray, and Richard
Rogers. "A Reality Check(List) for Digital Methods." *New Media*

& Society 20, no. 11 (November 1, 2018): 4195–4217. https://doi
.org/10.1177/1461444818769236.

Voyant Tools website. "Getting Started." https://voyant-tools.org/docs
/#!/guide/start.