070172 UE Methodological course - Introduction to DH: Tools & Techniques (2020W) Memex Edition

Maxim G. Romanov

2020-10-08

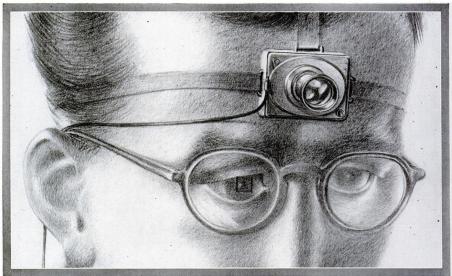
Contents

P	reliminaries	5
Μ	Iemex and Zettelkasten	9
	On Memex	10
	On Zettelkasten	10
1	Lesson 01	11
	1.1 Bibliography Managers	11
	1.2 Zotero	11
	1.3 Homework	17
$\mathbf{S}_{\mathbf{J}}$	yllabus	19
	Course Details	19
	Aims, Contents and Method of the Course	19
	Course Evaluation	20
	Class Participation	20
	Homework	21
	Final Project	21
	Study materials	21
	Software, Tools, & Technologies	22
	Submitting Homework	23
	Schedule	23
	Lesson Topics	24

Preliminaries

This is a collection of relevant materials for a DH course by the Department of History, the University of Vienna.

- Course: 070172 UE Methodological course - Introduction to DH: Tools & Techniques (2020W) Memex Edition
- u:find Link: https://ufind.univie.ac.at/en/course.html?lv=070172&semester= $2020\mathrm{W}$
- Meeting time: Tu 09:00-10:30
- Meeting place: Seminarraum Geschichte 3 Hauptgebäude, 2.Stock, Stiege 9; due to COVID, all meetings will be held online
- $\bullet \;$ Instructor: Dr. Maxim Romanov, maxim.romanov@univie.ac.at
- Language of instruction: English
- Office hours: Tu 14:00-15:00 (on Zoom; please, contact beforehand!)
- Office: Department of History, Maria-Theresien-Straße 9, 1090 Wien, Room 1.10



A SCIENTIST OF THE FUTURE RECORDS EXPERIMENTS WITH A TINY CAMERA FITTED WITH UNIVERSAL-FOCUS LENS. THE SMALL SQUARE IN THE EYEGLASS AT THE LEFT SIGHTS THE OBJECT

AS WE MAY THINK

A TOP U.S. SCIENTIST FORESEES A POSSIBLE FUTURE WORLD IN WHICH MAN-MADE MACHINES WILL START TO THINK

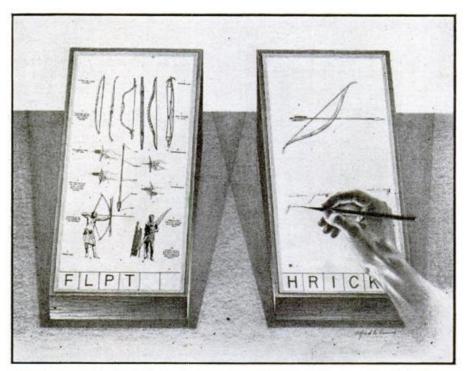
by VANNEVAR BUSH

RECTOR OF THE OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT

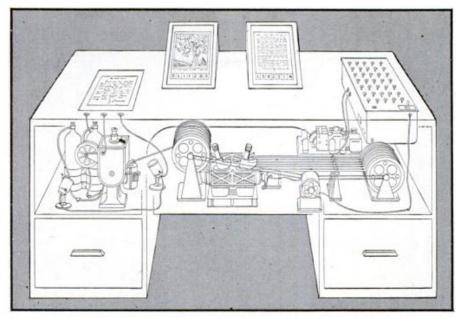
Back in 1945, Vannevar Bush, a Director of the US Office of Scientific Research and Development, proposed a device, which he called memex:

Consider a future device ... in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory. ... The owner of the memex, let us say, is interested in the origin and properties of the bow and arrow. Specifically he is studying why the short Turkish bow was apparently superior to the English long bow in the skirmishes of the Crusades. He has dozens of possibly pertinent books and articles in his memex. First he runs through an encyclopedia, finds an interesting but sketchy article, leaves it projected. Next, in a history, he finds another pertinent item, and ties the two together. Thus he goes, building a trail of many items. Occasionally he inserts a comment of his own, either linking it into the main trail or joining it by a side trail to a particular item. When it becomes evident that the elastic properties of available materials had a great deal to do with the bow, he branches off on a side trail which takes him through

textbooks on elasticity and tables of physical constants. He inserts a page of longhand analysis of his own. Thus he builds a trail of his interest through the maze of materials available to him. And his trails do not fade. Several years later, his talk with a friend turns to the queer ways in which a people resist innovations, even of vital interest. He has an example, in the fact that the outraged Europeans still failed to adopt the Turkish bow. In fact he has a trail on it. A touch brings up the code book. Tapping a few keys projects the head of the trail. A lever runs through it at will, stopping at interesting items, going off on side excursions. It is an interesting trail, pertinent to the discussion. ... — The Atlantic, July 1945.



MEMEX IN USE is shown here. On one transparent screen the operator of the future writes notes and commentary dealing with reference material which is projected on the screen at left. Insertion of the proper code symbols at the bottom of right-hand screen will tie the new item to the earlier one after notes are photographed on supermicrofilm.



MEMEX in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference.

Memex and Zettelkasten

The *memex* machine is often thought of as a precursor of the Internet, where information is interconnected. Vannevar Bush, however, seems to have viewed more as a personal machine knowledge organization system, to use a modern term. A system that would facilitate coordination of relevant pieces of information into organized sequences that he himself called "trails".

The idea of a personal knowledge device is still of great relevance and of great importance to scholars and scientists whose job is to construct such trails on a daily basis. Needless to say that historians will particularly benefit from having such a machine at their disposal (as Bush's example about the Turkish bow and the English longbow indicates, see (?)).

While it is not necessary to adhere slavishly to Bush's vision, we can definitely use his vision as a springboad to develop something similar; something that will allow us to navigate the massive volumes of information, which grew significantly since the 1940s.

In the following two sections you will find some relevant materials on memex and the history of this idea. Here, however, I want to take some time to think about how the design of our own memex should look like. What do we want from it? What can we reasonably achieve?

Before I procede to that, I would like to dwell on another relevant and, in my opinion, closely connected idea — that of *Zettelkasten*. On its own, there is nothing particularly interestin and exciting about it, as the word refers to a rather unexciting piece of furniture: a "slip-box", or a "card-box". However, this term became closely associated with Niklas Luhmann, a German professor of sociology (U Bielefeld), who is considered one of the most prolific scholars of the 20th century. Like others, Luhmann himself attributed his productivity to his working method and the knowledge organization system which he implemented and systematically used throughout his career (See, ??).

NB: Detailed bibliography can be found at the end of the section and in the References section (see, TOC).

On Memex

• (?) is the first article—"As We May Think"—that Vannevar Bush published on memex in *The Atlantic*. This animation imagines the way memex would have functioned (produced by the organizers of the Brown/MIT symposium).

- (?) is the book that came out of the symposium held at the The Brown/MIT Vannevar Bush Symposium in 1995, celebrating the 50th anniversary of Bush's groundbreaking article "As We May Think". The book includes several Bush's articles on memex that show the evolution of his thinking about this device. This book is difficult to find; someone made an EPUB version of it (also shared via Slack). Recordings of the symposium are available on YouTube:; other videos of this symposium can also be found at the Video Archive of The MIT/Brown Vannevar Bush Symposium https://www.dougengelbart.org/content/view/258/000/. I highly recommend you watch Paul Kahn: A Visual Tour of Vannevar Bush's Work; other presentations are very interesting as well.
- (?) is an recent experiment (MA Thesis in Design), trying to create a version of memex.

On Zettelkasten

Niklas Luhmann on his Zettelkasten

• Niklas Luhmann was open about his working method, which he discussed in his interviews (?), and in some of his academic articles(?).

Others on Luhmann's Zettelkasten

...

Chapter 1

Lesson 01

1.1 Bibliography Managers

Bibliography managers make your life easier when it comes to collectin, organizing and maintaining bibliographical references and your library of electronic publications (most commonly as PDFs). Additionally, they are an indispensable writing tool as they take care of formatting (and reformatting) references and bibliographies in any writing project that you might undertake. There are plenty of different programs out there with their advantages and disadvantages (for example, Mendeley, RefWorks, Citavi, Endnote, Papers, Zotero, and quite a few more). We will use Zotero—it is being developed by scholars for scholars; it is free and open source; it does pretty much everything you might possibly need from a program of this kind.

1.2 Zotero

1.2.1 Getting Started

- Zotero can be installed from here: https://www.zotero.org/download/; the page will offer you a version suitable for your operating system, but you should also see the links to versions for specific systems (Mac OS, Windows, Linux).
- During installation Zotero should automatically integrate into your browser (like Chrome or Firefox) and into your word processor (MS Word, LibreOffice, GoogleDocs are supported). It is possible that you may have to do that manually.
 - Zotero Connector for Chrome can be installed from the same page (https://www.zotero.org/download/)

 detailed explanations on how to use word processor plugins can be found here; you can use Zotero with MS Word, LibreOffice and Google Docs; in case you cannot get your plugin activated, check the Troubleshooting Section.

1.2.2 Main Functionality

You need to be able to do the following tasks with your Zotero in order to take full advantage of its functionality.

(If you prefer video tutorials, you can check a series of Zotero Tutorials prepared by McGill Library; there are also plenty other tutorials on YouTube:)

1. Adding bibliographical records (and PDFs)

- Using Zotero Connector: the easiest way to add a reference is from a browser with Zotero connector. This can be done practically from any library or journal database (e.g., Uni Wien Library, Worldcat.org, JSTOR); simply click the connector button while you are on a page with a publication that you want to add to your Zotero database. PDF may be automatically downloaded, if available; keep in mind that in places like JSTOR you need to agree to terms before this function will work; what you need to do is to download one PDF manually from a JSTOR page, where you will be asked to agree to terms of their services;
- Drag-and-dropping PDFs into Zotero; this however works only when Zotero can parse relevant bibliographical information from a PDF; This might be a good way to start if you already have lots of PDFs that you want to add to Zotero.
- Using Unique Identifiers: you can use ISBN or DOI numbers.
- *Using Import*: you can import bibliographical data from another application or from bibliographical files (formats, like RIS, which you can download from most libraries as well).
- Manually: you can manually add and fill in a record as well.

2. Write-and-cite

- Detailed Instructions: MS Word, LibreOffice and Google Docs; you can also check the video tutorial.
- Add a citation
- Customize a citation (by adding prefixes, suffixes, page range for a specific reference, etc.).
- Change citation style.
 - For example, change from *Chicago Manual of Style* to *Universität Wien Institut für Geschichte* (Yes, there is this specific citation style

1.2. ZOTERO 13

for Zotero: https://www.zotero.org/styles?q=id%3Auniversitat-wien-institut-fur-geschichte); in order to do that you need to download the IfG style and install it into Zotero.

- You can find lots of different citation styles here: https://www.zotero.org/styles; to add a new style to Zotero:
 - * download the style you want.
 - * Open Zotero. Go to Preferences (under Zotero, Edit, or Tools—depending on your system).
 - * Click the "Cite" button.
 - * Click the "Styles" tab.
 - * Click the + button at the bottom right.
 - * Select the style file you saved in the first step.
- Generate and update bibliography in your paper.
- NB: If you use Zotero plugin for adding your citations, they remain connected to Zotero and can be automatically reformatted; you can also dragand-drop any bibliographical record into any text editor—the reference will be formatted according to the currently selelected style, but it will not be connected to Zotero and cannot be reformatted automatically later.

3. General Maintenance and Organization

- Zotero can [automatically] rename PDFs using metadata, although the default function is not very robust (see, *Zotfile* plugin below).
- You can create "collections" and drag-and-drop publications relevant to a specific topic or project you are working on.

1.2.3 Additional Functionality: Plug-Ins

There is a variety of third-party plugins that you can add to Zotero for additional functionality. The list of plugins can be found at https://www.zotero.org/support/plugins. To install a plugin, you need to download its .xpi file to your computer. Then, in Zotero, click "Tools \rightarrow Add-Ons", then drag the .xpi for the plugin onto the Add-Ons window that opens. Two plugins will be of particular interest to us: Zotfile and BetterBibTeX.

1.2.4 Zotfile

Zotfile (http://zotfile.com/) is a Zotero plugin to manage your attachments: automatically rename, move, and attach PDFs (or other files) to Zotero items, sync PDFs from your Zotero library to your (mobile) PDF reader (e.g. an iPad, Android tablet, etc.) and extract annotations from PDF files.

This plugin is particularly helpful for organizing PDFs on your hard drive. By default, Zotero saves PDFs in a computationally safe, but humanely incomprehensible manner: each PDF, even if it is renamed from bibliographical metadata and is human readable, it is still placed into a folder whose name is a random sequence of characters. Zotfile allows you to organize PDFs in a more human-friendly manner. The first screenshot below shows Zotero default mode, while the second one shows Zotfile mode: essentially, Zotfile creates a folder for each author and PDFs of all publications by that author get placed in that folder. You can sync this folder with Dropbox or other cloud service and access it from your tablet or phone.

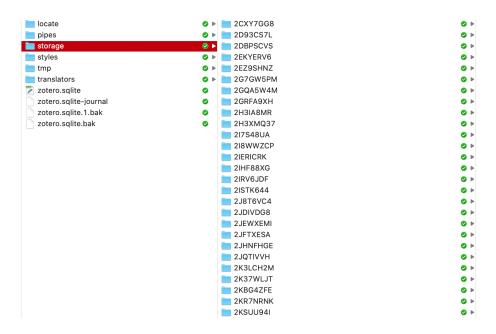


Figure 1.1: Zotero default organization.

1.2.5 Better BibTeX for Zotero

For a moment this will not be an immediately useful plug-in, but it is the most important one for our Memex project. This plugin exports bibliographical data into a bibTeX format, which is very easy to process with python scripts (it also generates *citation keys* which can be used for citation in markdown, which we will cover later). The two screenshots below show how the same record looks in Zotero preview and in the bibTeX format.

1.2. ZOTERO 15

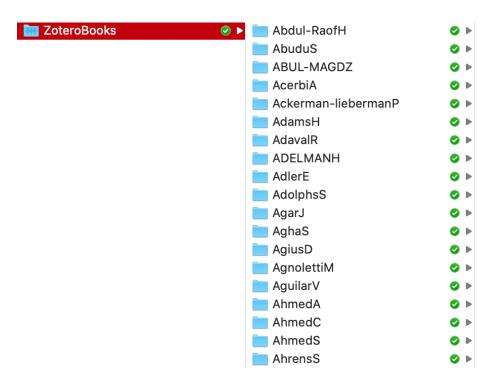


Figure 1.2: Zotfile organization.

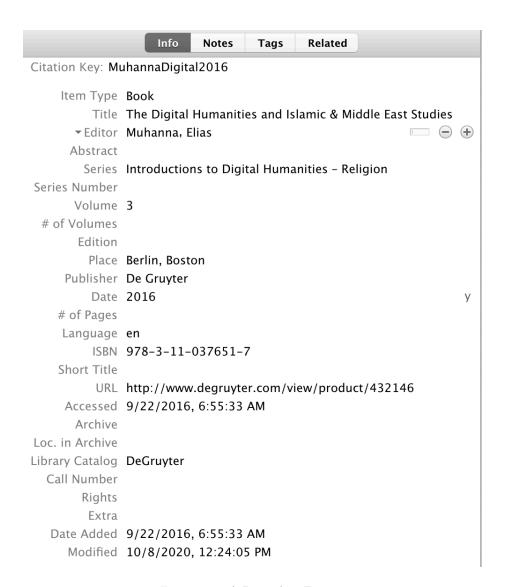


Figure 1.3: A Record in Zotero.

1.3. HOMEWORK 17

```
@book{MuhannaDigital2016,
   title = {The Digital Humanities and Islamic \& Middle East Studies},
   editor = {Muhanna, Elias},
   date = {2016},
   volume = {3},
   publisher = {De Gruyter},
   location = {Berlin, Boston},
   url = {http://www.degruyter.com/view/product/432146},
   urldate = {2016-09-22},
   file = {/Users/romanovienna/Dropbox/3.Reading/Zotero/ZoteroBooks/
   MuhannaE/Muhanna 2016 - The Digital Humanities and Islamic & Middle
   East Studies.pdf},
   isbn = {978-3-11-037651-7},
   langid = {english},
   series = {Introductions to Digital Humanities - Religion}
}
```

Figure 1.4: The Same Record in BibTeX Format.

1.3 Homework

- collect 30-50 bibliographic records into your Zotero (ideally with PDFs); the number may seem like a lot, but you will see that you can do that it will take only about 30 mins on JSTOR; those of you who are already using Zotero must already have more than 50 records in your databases.
- clearly, you should be collecting items that are relevant to your fields of study and your research; organize them into folders, if that is necessary;
- create Bibliography and email it to me (this is one-click operation; try to figure on your own how to do this; asking on *Slack* counts);
- make sure that you are comfortable with the main functionality of Zotero; that you have the discussed plugins installed; to get comfortable with the main functionality, you should practice each listed procedure at least a couple of times.

1.3.1 Submitting homework:

- Homework assignment must be submitted by the beginning of the next class;
- Email your homework to the instructor as attachments.
 - In the subject of your email, please, add the following: 071172-L01-HW-YourLastName-YourMatriculation
 where YourLastName is your last name and YourMatriculationNumber
 is your matriculation number.

Syllabus

- Course: 070172 UE Methodological course Introduction to DH: Tools & Techniques (2020W) Memex Edition
- Instructor: Dr. Maxim Romanov, maxim.romanov@univie.ac.at
- Language of instruction: English
- Office hours: TBA
- Office: Department of History, Maria-Theresien-Straße 9, 1090 Wien, Room 1.10

Course Details

- u:find Link: https://ufind.univie.ac.at/en/course.html?lv=070172&semester=2020W
- Meeting time: Tu 09:00-10:30
- Meeting place: Seminarraum Geschichte 3 Hauptgebäude, 2.Stock, Stiege
 9; due to COVID, all meetings will be held online

Aims, Contents and Method of the Course

Back in 1945, Vannevar Bush, a Director of the US Office of Scientific Research and Development, proposed a device, which he called memex:

Consider a future device ... in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory. ... The owner of the memex, let us say, is interested in the origin and properties of the bow and arrow. Specifically he is studying why the short Turkish bow was apparently superior to the English long bow in the skirmishes of the Crusades. He has dozens of possibly pertinent books and articles in his memex. First he runs through an encyclopedia, finds

an interesting but sketchy article, leaves it projected. Next, in a history, he finds another pertinent item, and ties the two together. Thus he goes, building a trail of many items. Occasionally he inserts a comment of his own, either linking it into the main trail or joining it by a side trail to a particular item. When it becomes evident that the elastic properties of available materials had a great deal to do with the bow, he branches off on a side trail which takes him through textbooks on elasticity and tables of physical constants. He inserts a page of longhand analysis of his own. Thus he builds a trail of his interest through the maze of materials available to him. And his trails do not fade. Several years later, his talk with a friend turns to the queer ways in which a people resist innovations, even of vital interest. He has an example, in the fact that the outraged Europeans still failed to adopt the Turkish bow. In fact he has a trail on it. A touch brings up the code book. Tapping a few keys projects the head of the trail. A lever runs through it at will, stopping at interesting items, going off on side excursions. It is an interesting trail, pertinent to the discussion. ... — The Atlantic, July 1945; YouTube: https://www.youtube.com/watch?v=c539cK58ees.

The memex machine is often thought of as a precursor of the Internet. Be it as it may, the idea of a personal knowledge device is still of great relevance and of great importance to scholars and scientists whose job is to construct such trails on a regular basis. Needless to say that *historians* will benefit greatly from having such a machine at their disposal. The course will introduce you to basic, intermediate, and some advanced computational techniques, which will allow you to build and maintain your own digital memex machine.

No prior programming experience is expected (we will be learning Python). Each class session will consist in large part of practical hands-on exercises led by the instructor. Laptops are required for the course. We will accommodate whatever operating system you use (Windows, Mac, or Linux), but it must be a laptop rather than a tablet.

Course Evaluation

Course evaluation will be a combination of in-class participation (30%), weekly homework assignments (50%), and the final project (20%).

Class Participation

Attendance is required; regular participation is the key to completing the course; all students must come with their laptops; homework assignments must be submitted on time (some can be completed later as a part of the final project, but

1.3. HOMEWORK 21

this must be discussed with the instructor whenever the issue arises); the final project must be submitted on time.

Homework

Just as in research and real life, collaboration is a very good way to learn and is therefore encouraged. If you need help with any assignment, you are welcome to ask a fellow student. If you do work together on homework assignments, then when you submit it please include a brief note (just a sentence or two) to indicate who did what.

NB: On submitting homework, see below.

Final Project

The final project is your own memex machine, which can help you with your studies and your research. Your final project must include all working scripts that will allow you in the future to continuously expand your memex machine by adding new readings into the mix. You are most welcome to work on this final project in groups, but everybody is required to produce their own working machine.

Study materials

MAIN TEXTBOOK

- Zelle, John M. Python Programming: An Introduction to Computer Science. Third edition. Portland, Oregon: Franklin, Beedle & Associates Inc, 2017. (access via Moodle); (?)
 - We will focus primarily on learning how to work with python, which is one of the most popular programming languages used in digital humanities. We will use several resources and the emphasis will be on you studying on your own: partially, this is because of time constraints, but more importantly, you will need to acquire a skill of learning on your own. No worries, I will provide necessary help whenever needed.
 - This textbook will be our main resource. It is well written and will help you to wrap your heads around important computer science concepts; this reading is crucial and without it many interactive tutorials out there will not be particularly helpful. Each chapter has assignments and self-test multiple choice sections;

 Supplementary materials are available at the publisher's website, where you can download example code and end-of-chapter solutions; additionally, you can find videos with complimentary instructions

- Additional:

- * Paul Vierthaler's "Hacking the Humanities Tutorials" (Python+): https://www.youtube.com/playlist?list=PL6kqrM2i6BPIpEF5yHPNkYhjHnFYWh17
- * https://www.codecademy.com/learn/learn-python :: you can use this free interactive Python course; it, however, uses Python 2.x, while the main textbook focuses on Python 3.x; the course is still a good supplementary practice.

ADDITIONAL MATERIALS

- https://www.codecademy.com
 - Codecademy has a series of free course that you are encouraged to use for specific skills and technologies:
 - * https://www.codecademy.com/learn/learn-how-to-code
 - * https://www.codecademy.com/learn/learn-python
 - * https://www.codecademy.com/learn/learn-html
 - $*\ https://www.codecademy.com/learn/introduction-to-regular-expressions$
 - * https://www.codecademy.com/learn/learn-css
- https://programminghistorian.org/lessons/
 - "Programming Historian" offers a number of tutorials for aspiring digital humanists. These will be assigned to you as reference materials. You also are encouraged to explore those tutorials that are not included into the course.

Software, Tools, & Technologies

The following is the list of software, applications and packages that we will be using in the course. Make sure to have them installed by the class when we are supposed to use them.

- Zotero, https://www.zotero.org/
- [Mac] Terminal / [Windows] Powershell (both are already on your machines)
- Python https://www.python.org/, install the latest 3.x version
- git and https://github.com/, version control system
- pandoc (https://pandoc.org/), markdown, bibTex (bibliographical format for LaTeX)

1.3. HOMEWORK 23

• Regular expressions; (Sublime Text, https://www.sublimetext.com/ is a text editor which supports regular expressions)

- Wget (https://www.gnu.org/software/wget/), a free software package for retrieving files
- Understanding formats: [TEI] XML, csv/tsv, json, yml, etc.
- Creating: HTML, css, tiny snippets of Javascript

Submitting Homework

- Homework assignments are to be submitted by the beginning of the next class;
- For the first few classes you must email them to the instructor (as attachments)
- Later, you will be publishing your homework assignments on your websites
 and sending an email to the instructor informing that you have completed
 your homework and providing a link to the blogpost with the homework
 report that you created.
 - In the subject of your email, please, use the following format: CourseID-LessonID-HW-Lastname-matriculationNumber, for example, if I were to submit homework for the first lesson, my subject header would look like: 070112-L01-HW-Romanov-12435687.
- DH is a collaborative field, so you are most welcome to work on your homework assignments in groups, however:
 - You must still submit it. That is, if a groups of three works on one assignment, there must be three separate submissions emailed from each member's email.

Schedule

Location: Seminarraum Geschichte 3 Hauptgebäude, 2.Stock, Stiege 9; due to COVID, all meetings will be held online via video-conferencing

- Tuesday 06.10. 09:00 10:30
- Tuesday 13.10. 09:00 10:30
- Tuesday 20.10. 09:00 10:30
- Tuesday 27.10. 09:00 10:30
- Tuesday 03.11. 09:00 10:30
- Tuesday 10.11. 09:00 10:30
- Tuesday 17.11. 09:00 10:30
- Tuesday 24.11. 09:00 10:30
- Tuesday 01.12. 09:00 10:30

- Tuesday 15.12. 09:00 10:30
- Tuesday 12.01. 09:00 10:30
- Tuesday 19.01. 09:00 10:30
- Tuesday 26.01. 09:00 10:30

Lesson Topics

- === CORE TOOLS & METHODS ===
- [#01] Introduction & Roadmap; Managing Bibliography with Zotero
- [#02] Getting to Know the Command Line; Getting Started with Python
- [#03] Version Control and Collaboration
- [#04] Sustainable [Academic] Writing
- [#05] Constructing Robust Searches / Optional: Basics of Webscraping
- [#06] Understanding Structured Data and Major Formats
- === BUILDING MEMEX ===
- [#07] Parsing and Manipulating Bibliographic Data
- [#08] Processing PDFs: OCR
- [#09] View and Display: Simple HTML-based Interface
- [#10] Summarizing Textual Data: Keyword Extraction
- [#11] Finding Connections: Similarity Measures
- [#12] Processing Everything Together: Batch Processing and re-Processing
- [#13] Improving the Overall Memex Design: What Else Can We Add?

Note: one of the classes might be canceled; this will be announced separately. Lesson materials will be appearing on the website shortly before each class. Lessons will be accessible via the *Lessons* link on the left panel.

Bibliography

- Bush, V. As We May Think.
- Luhmann, N. Kommunikation mit zettelkiisten. In Baier, H. and Noelle-Neumann, E., editors, Öffentliche Meinung und sozialer Wandel: Für Elisabeth Noelle-Neumann = Public opinion and social change, pages –228. westdt. Verl.
- Luhmann, N. and Baecker, D. Archimedes und wir: Interviews. Number 143 in Internationaler Merve-Diskurs. Merve.
- Nyce, J. M. and Kahn, P., editors. From Memex to Hypertext: Vannevar Bush and the Mind's Machine. Academic Press.
- Park, H. Memex and Beyond: A Design Trajectory from Vannevar Bush's Memex.
- Zelle, J. M. Python Programming: An Introduction to Computer Science. Franklin, Beedle & Associates Inc, third edition edition.