Scientific writing and bibliographic research

Week 1

Philipp Zumstein, Stefan Weil, Jörg Mechnich
Mannheim University Library

philipp.zumstein, stefan.weil, joerg.mechnich
@bib.uni-mannheim.de

Main Lecturers

- Philipp Zumstein
  - Subject librarian for mathematics + computer science
  - Head of publishing services + research support
  - Open Science specialist
- Stefan Weil
  - Head of the Digital Library Services Department
  - Deputy Subject Librarian for Mathematics and Computer Science
  - OCR and Open Source expert
- Jörg Mechnich
  - Deputy Head of the Digital Library Services Department
  - SysAdmin

Contents:

1. find a topic for a thesis → **Week 1**
2. search, organize literature and data → **Week 2**
3. write a thesis → **Week 3**

How to write a master‘s thesis …

Scientific Writing and Bibliographic Research for Students of Computer Science

… in three weeks?

Course format:

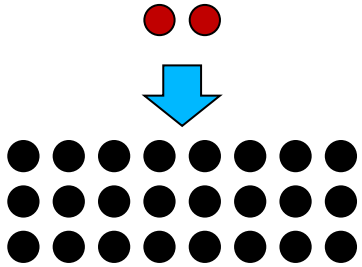
- block course
- 3 weeks
- online

Target audience:

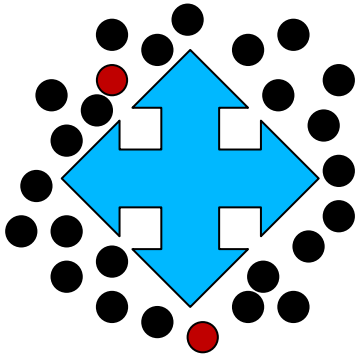
- students of computer science
- … for their master‘s thesis

= YOU!

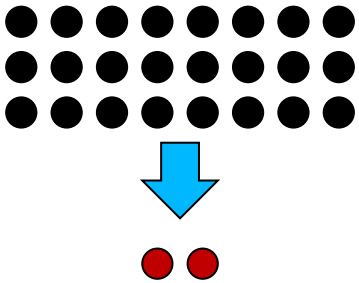
Principal idea



- Lecture
  - Theoretical input



- Workshop
  - Hands-on exercises
  - Individual work and group work



- Open learning
  - Your needs and questions

ASK Everything You Always Wanted to Know About “How to write a thesis”!!

online forum

Getting to know you: Question 1/3

- Mother tongue
  - A: German
  - B: English
  - C: other

Getting to know you: Question 2/3

- Thesis topic
  - A: topic not yet decided
  - B: topic decided
  - C: ideas for a topic

Getting to know you: Question 3/3

- Operating system on your laptop
  - A: Windows 10
  - B: macOS
  - C: Linux
  - D: other

Overall Goal of the Course

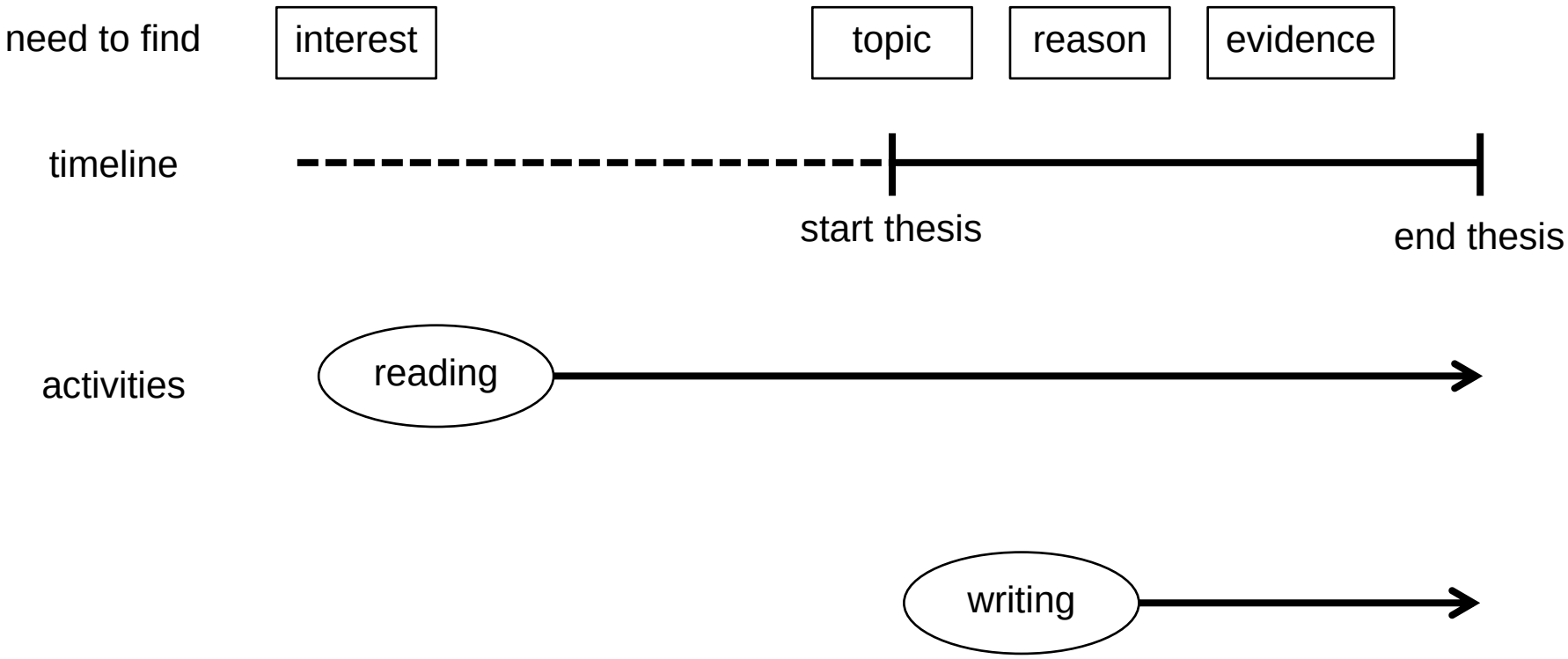
- After the course …
  - … you should know how to write a Master thesis
  - … you should actually performed all necessary steps on a small scale
  - … everybody should be on the same level
- During the course …
  - … you can share your experience from your Bachelor thesis
  - … reflect methods from your Bachelor thesis and compare them
  - … try out new tools or methods without thesis deadline approaching

Topics

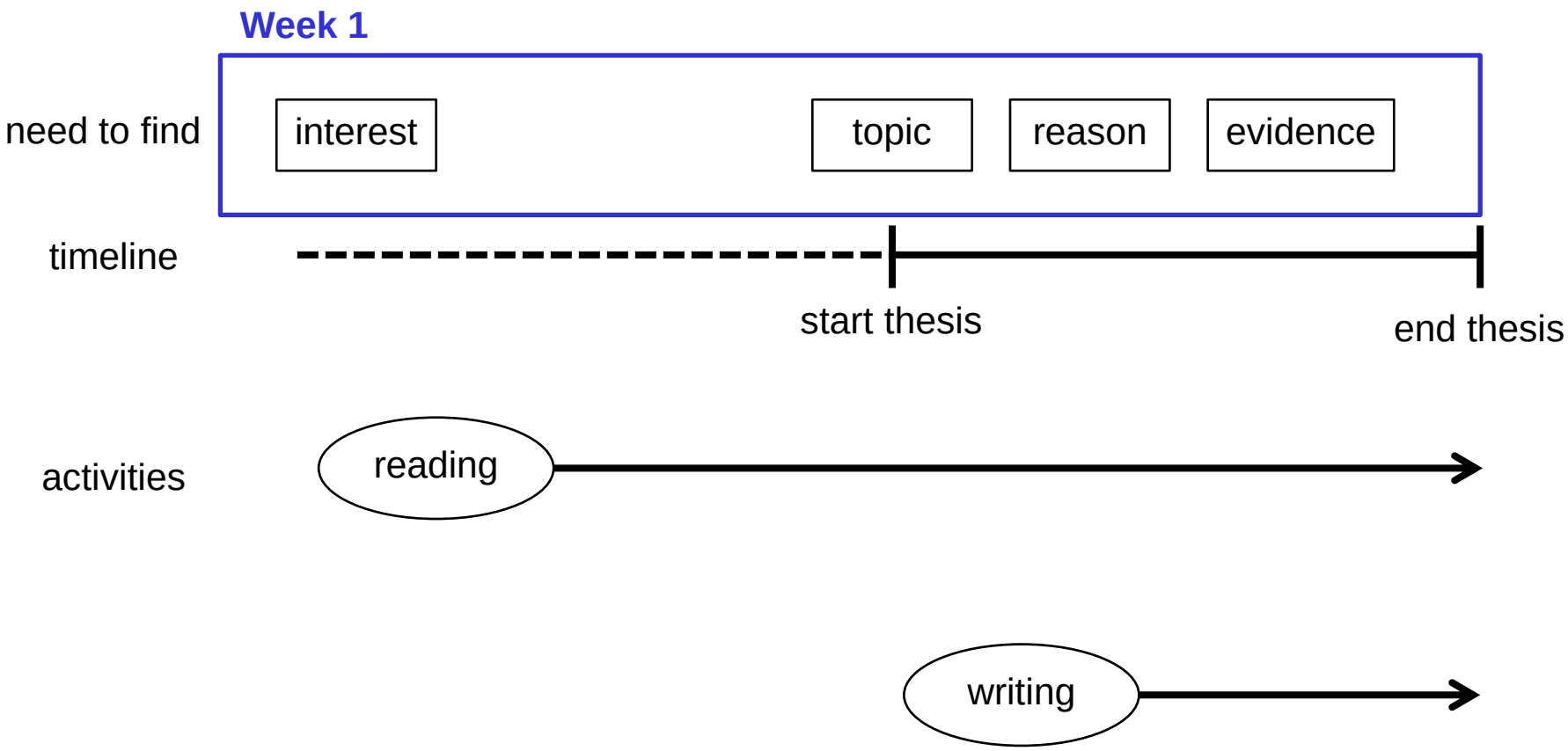
- Scientific process and scientific writing
- Bibliographic research methodology
- Research databases
- Tools

- Educational objective
  - At the end of this course, you...
    - Understand the **process** of scientific work
    - Learned and tried basic strategies for bibliographic research
    - Know and used the most important research databases in your field
    - Installed and used exemplary tools to support the work process
    - Prepared a short overview of your research topic „ready for press“
  - Thus, you can after the course (at least on a formal level) start writing your thesis!

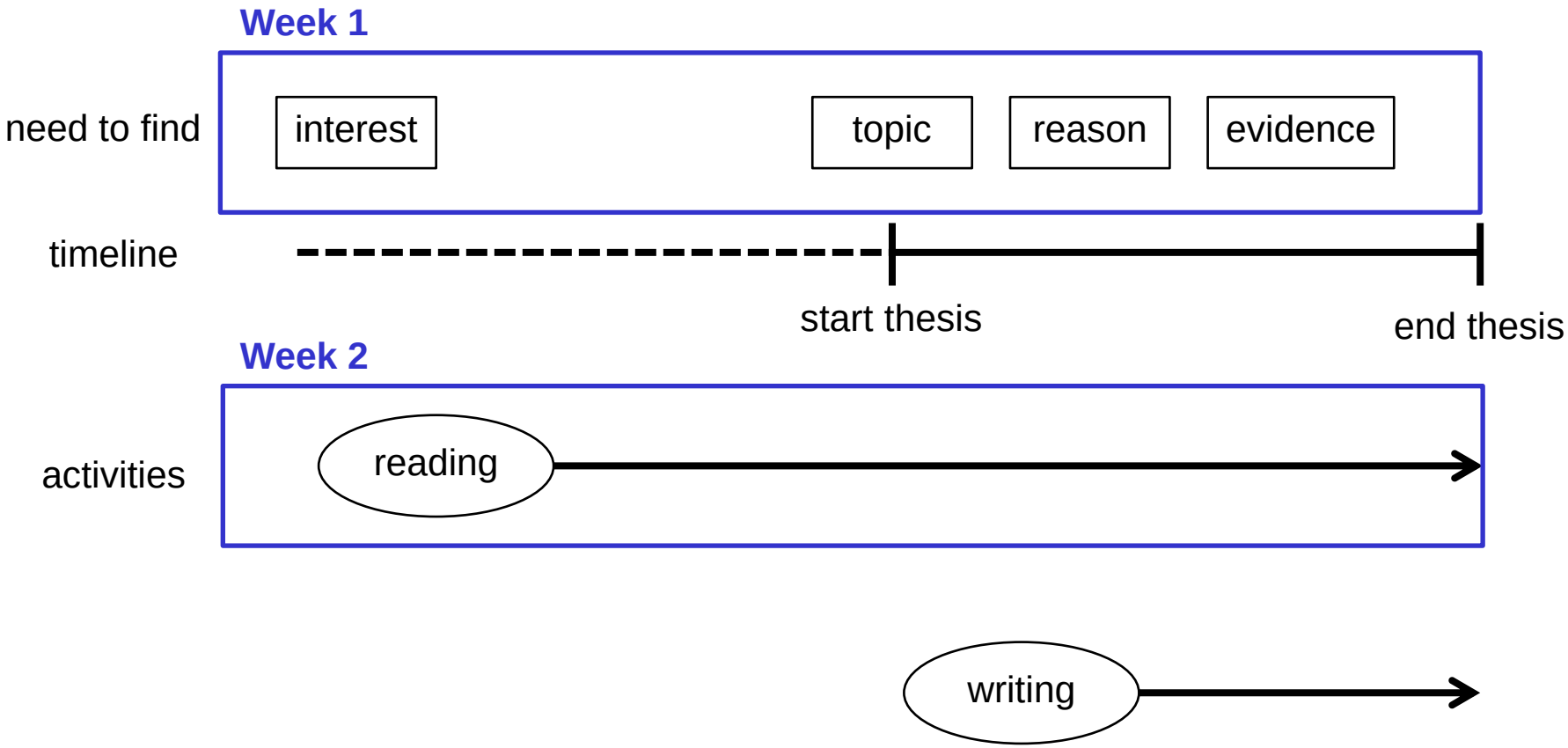
Overview


need to find
interest
topic
reason
evidence
timeline
start thesis
end thesis
activities
reading
writing

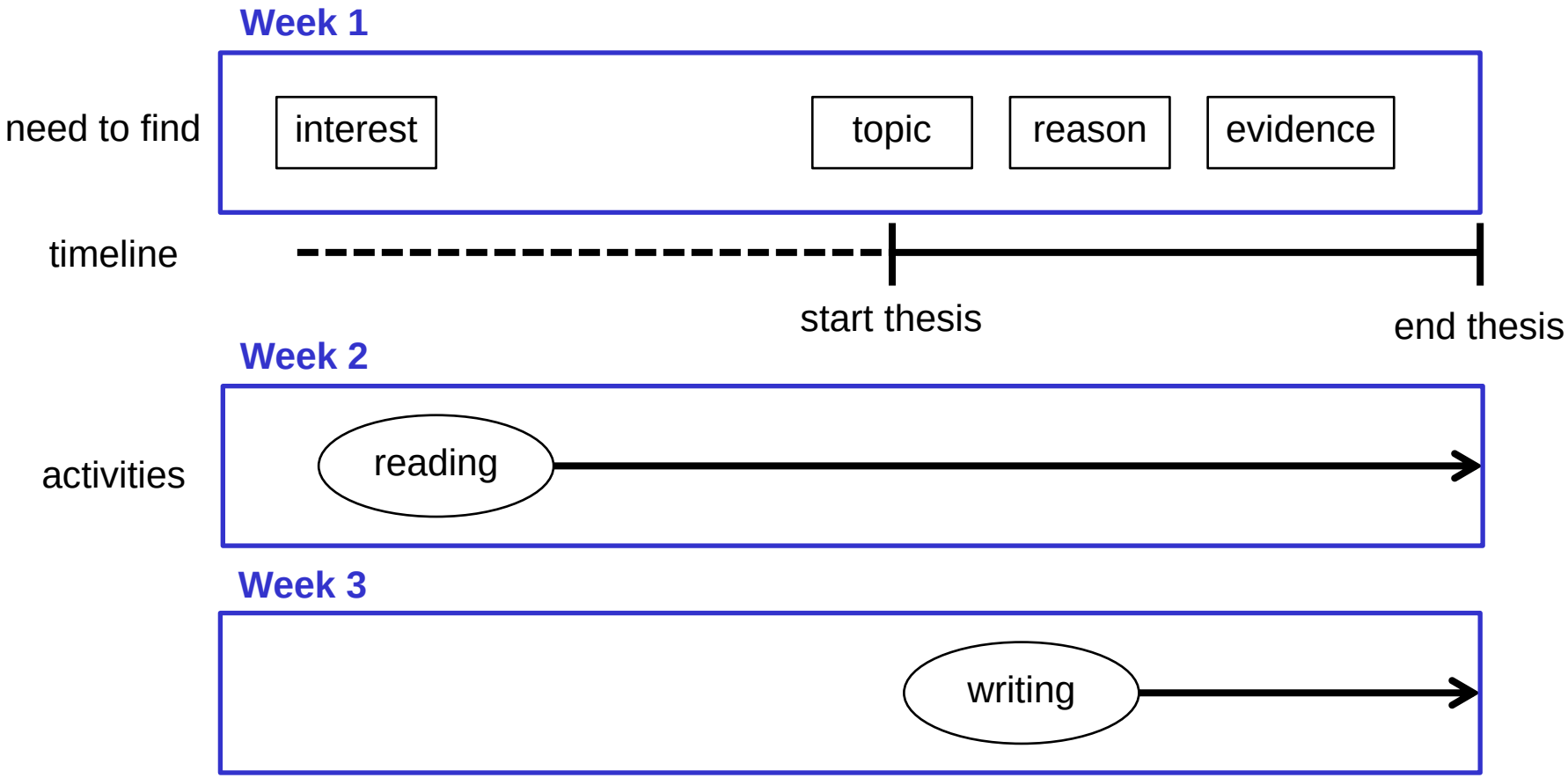
Overview


Week 1
need to find
interest
topic
reason
evidence
timeline
start thesis
end thesis
activities
reading
writing

Overview


Week 1
need to find
interest
topic
reason
evidence
timeline
start thesis
end thesis
Week 2
activities
reading
writing

Overview


Week 1
need to find
interest
topic
reason
evidence
timeline
start thesis
end thesis
Week 2
activities
reading
Week 3
writing

Course parts

- Everybody should check that they are **enrolled in Ilias** as well to this course!
- In Ilias **one week** contains
  - Several videos with the presentation (2-4 hours per week)
  - Additionally all slides as a PDF
  - Hands-on exercises (2-4 hours per week)
  - Forum for all your questions
- Live Sessions
  - Ask us questions
  - Small exercises for working collaboratively in a group
  - Zoom room

Fixed dates

	Monday	Tuesday	Wednesday	Thursday	Friday
Week 1	**10:00-10:30** Introduction				**13:30-15:00** Questions/ Recap Week 1
Week 2					**13:30-15:00** Questions/ Recap Week 2
Week 3		**14:00-15:00** Questions/ LaTeX Problems		**14:00-15:00** Questions/ LaTeX Problems	**13:30-16:30 EXAM**

Exam and Grades

- Exam
  - From 13:30 to 16:30
  - Open book exam
  - Combination of the exercises and the theoretical stuff
  - Practical test
  - Requirements:
    - laptop/computer
    - internet connection
    - the tools you were introduced in the course and/or you have used in the exercises
- Grades and ECTS points
  - will be entered into the university system later
- Possible to inspect your exam after the correction and grading (send us an email)

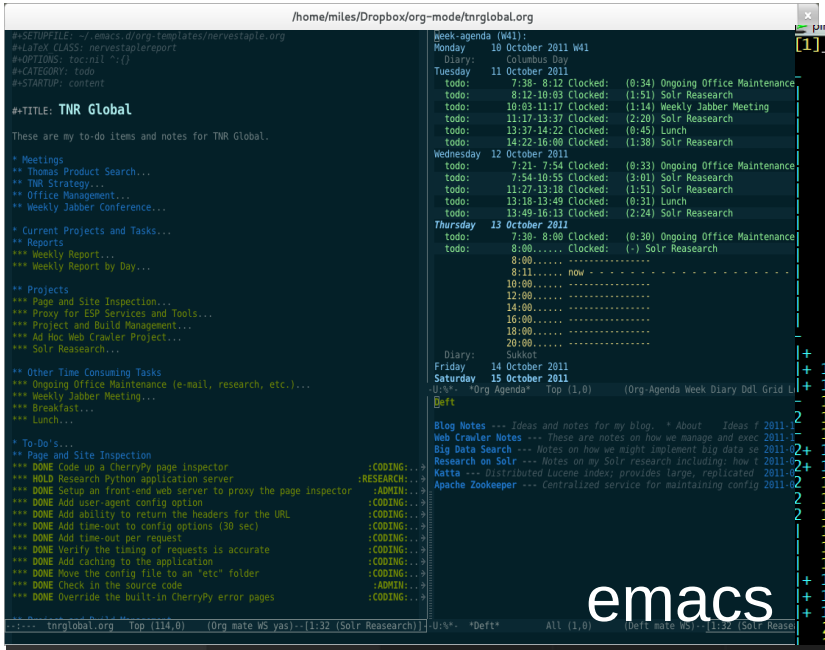
Questions

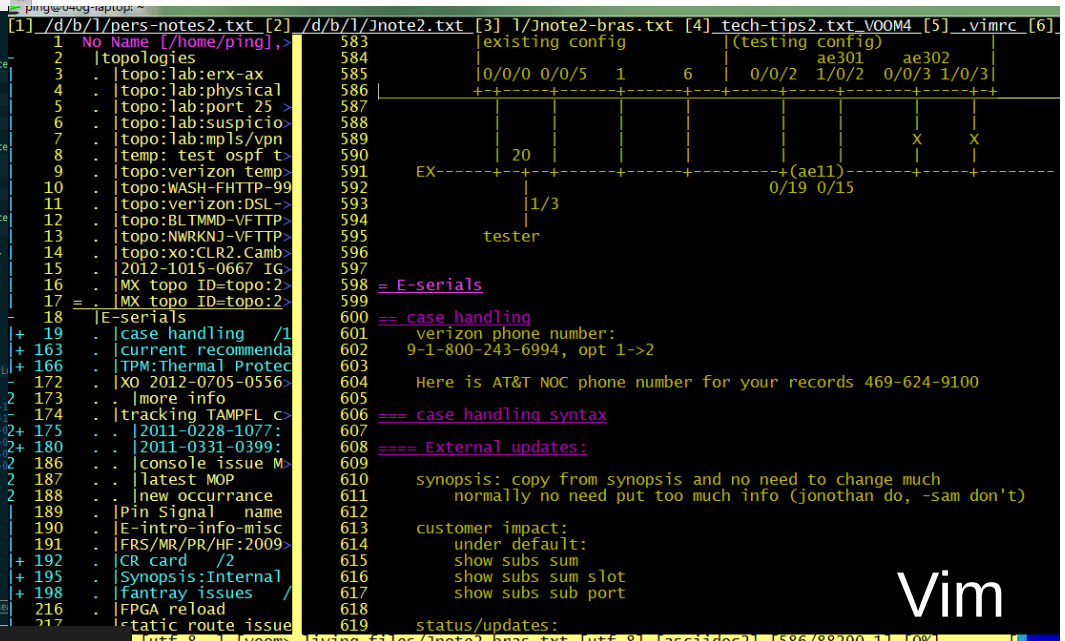
Tools and Tips

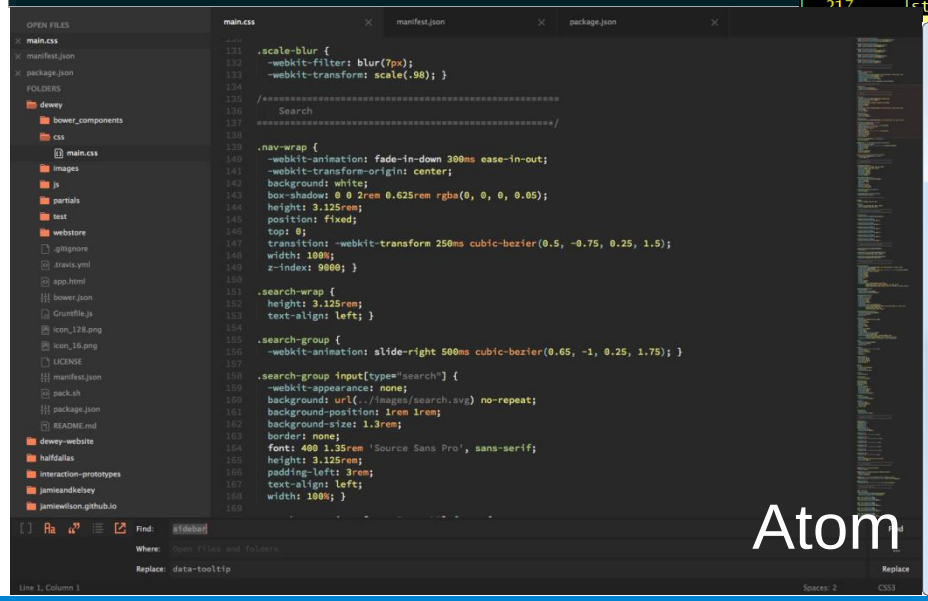
Note taking

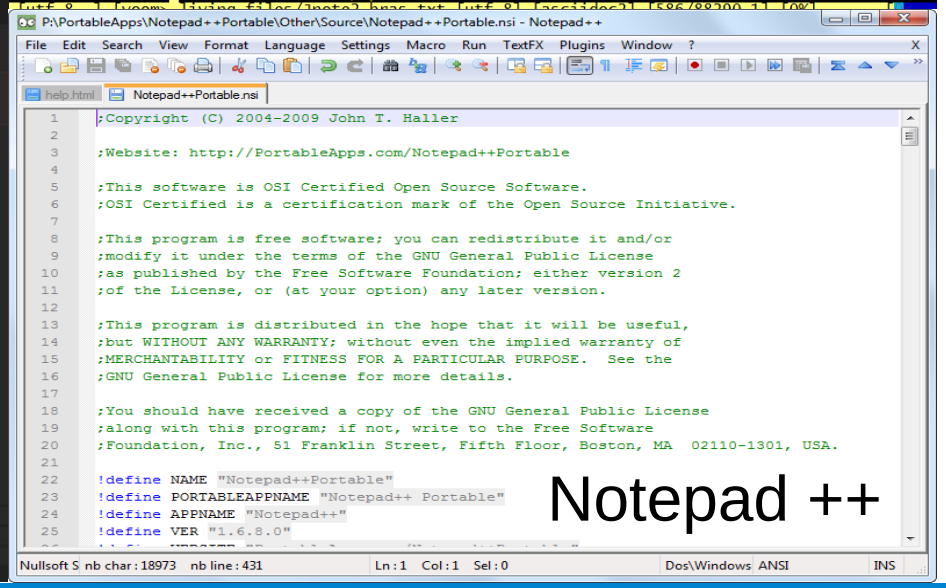
- Taking notes is an important part of your scientific work.
- Paper and pencil are still useful in some situations, but normally electronic notes are better. Think of some reasons.
- Text editors, personal wikis, mind mapping software, …
- Wikipedia has some good starting points with more information:
  https://en.wikipedia.org/wiki/Note-taking
  https://en.wikipedia.org/wiki/Comparison_of_notetaking_software
  https://en.wikipedia.org/wiki/List_of_mind_mapping_software

Note taking – Text editors + Version Control







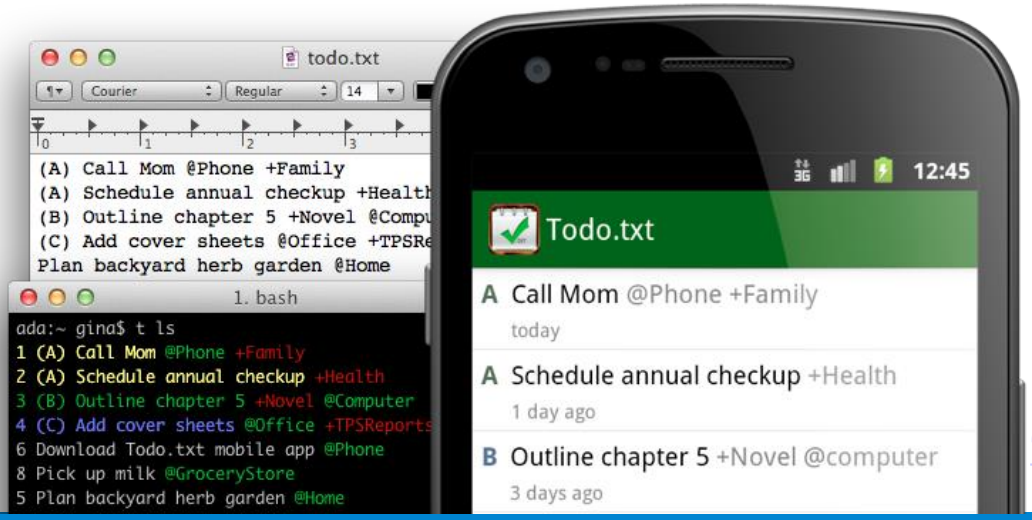


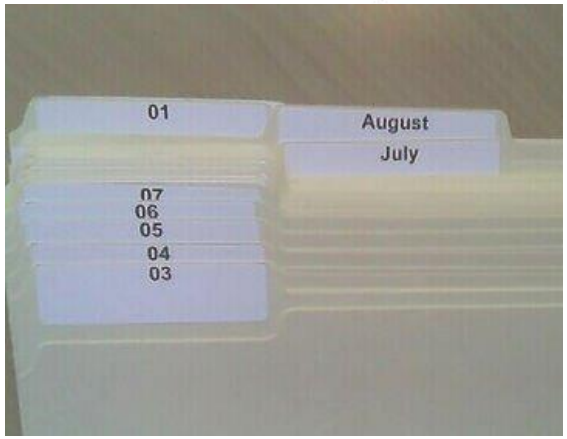
Taking notes – Plain Text Productivity

- Dokuwiki on a Stick:

  https://www.dokuwiki.org/install:dokuwiki_on_a_stick

- Plaintext Productivity: http://plaintext-productivity.net/
- Todo.txt: http://todotxt.com/
- Emacs Org-Mode: http://orgmode.org/

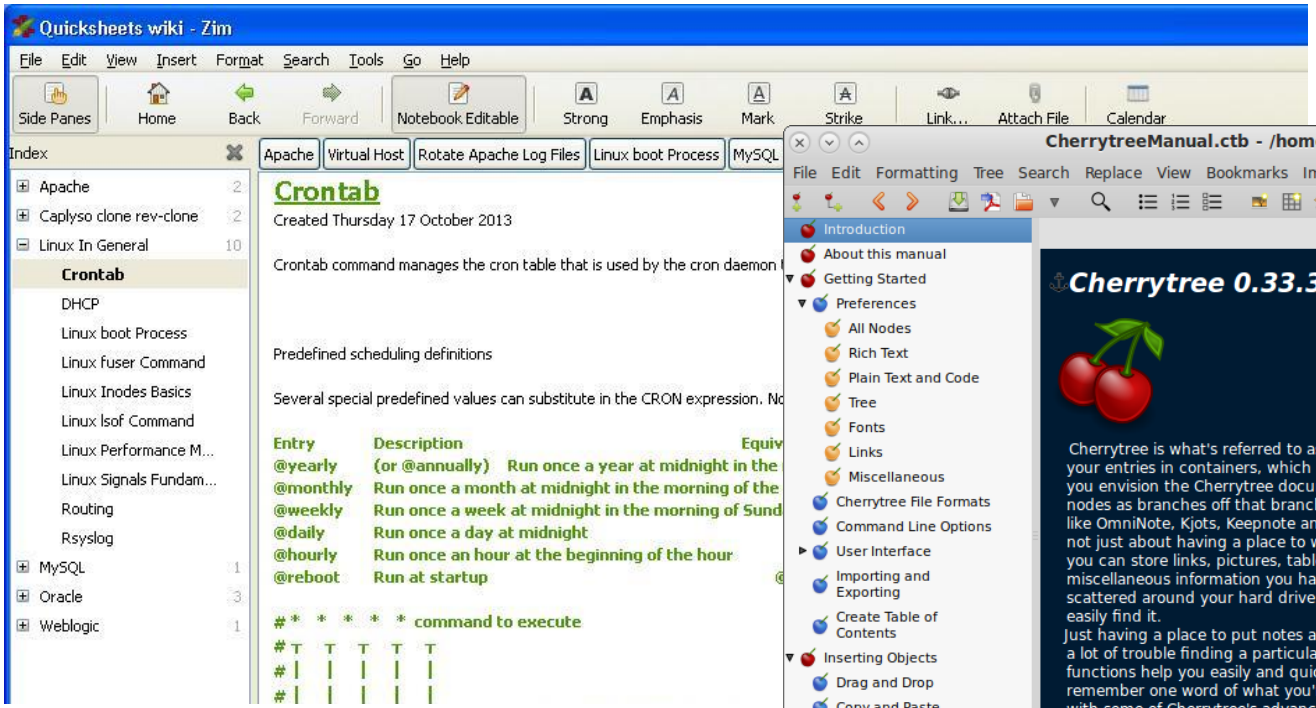


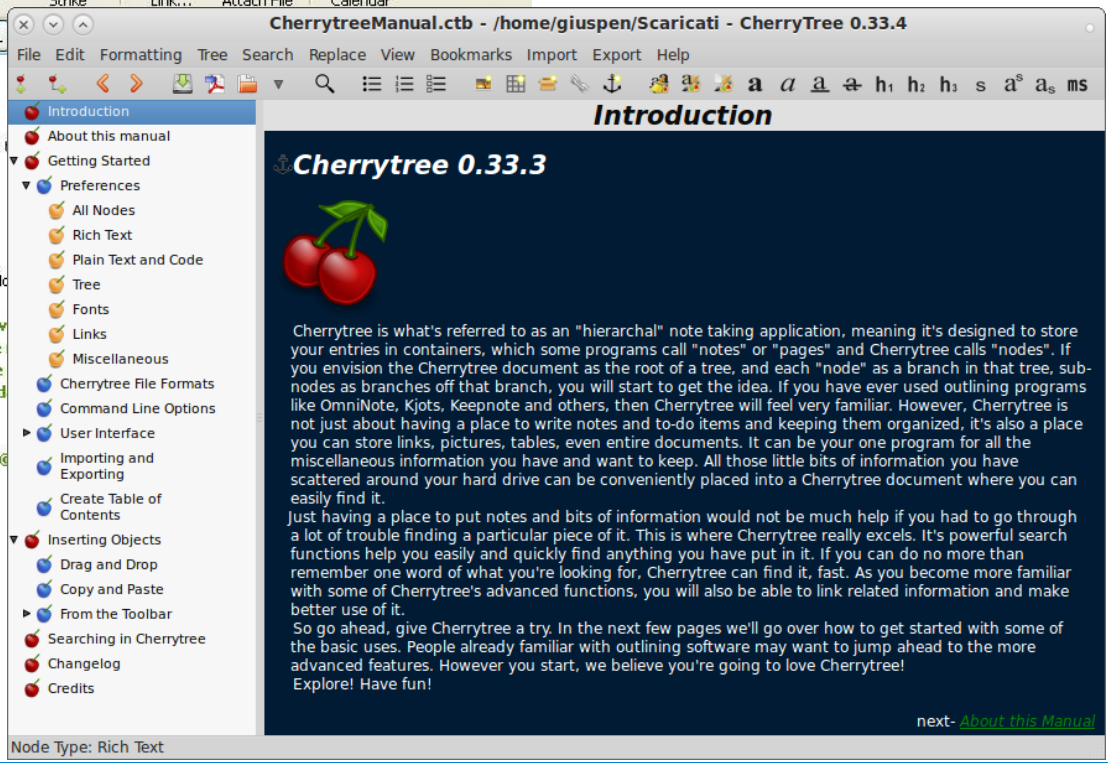


"43 Folders" CC-BY-SA, © by Vlauria@en.wp

Taking notes – Outliner / Desktop Wiki

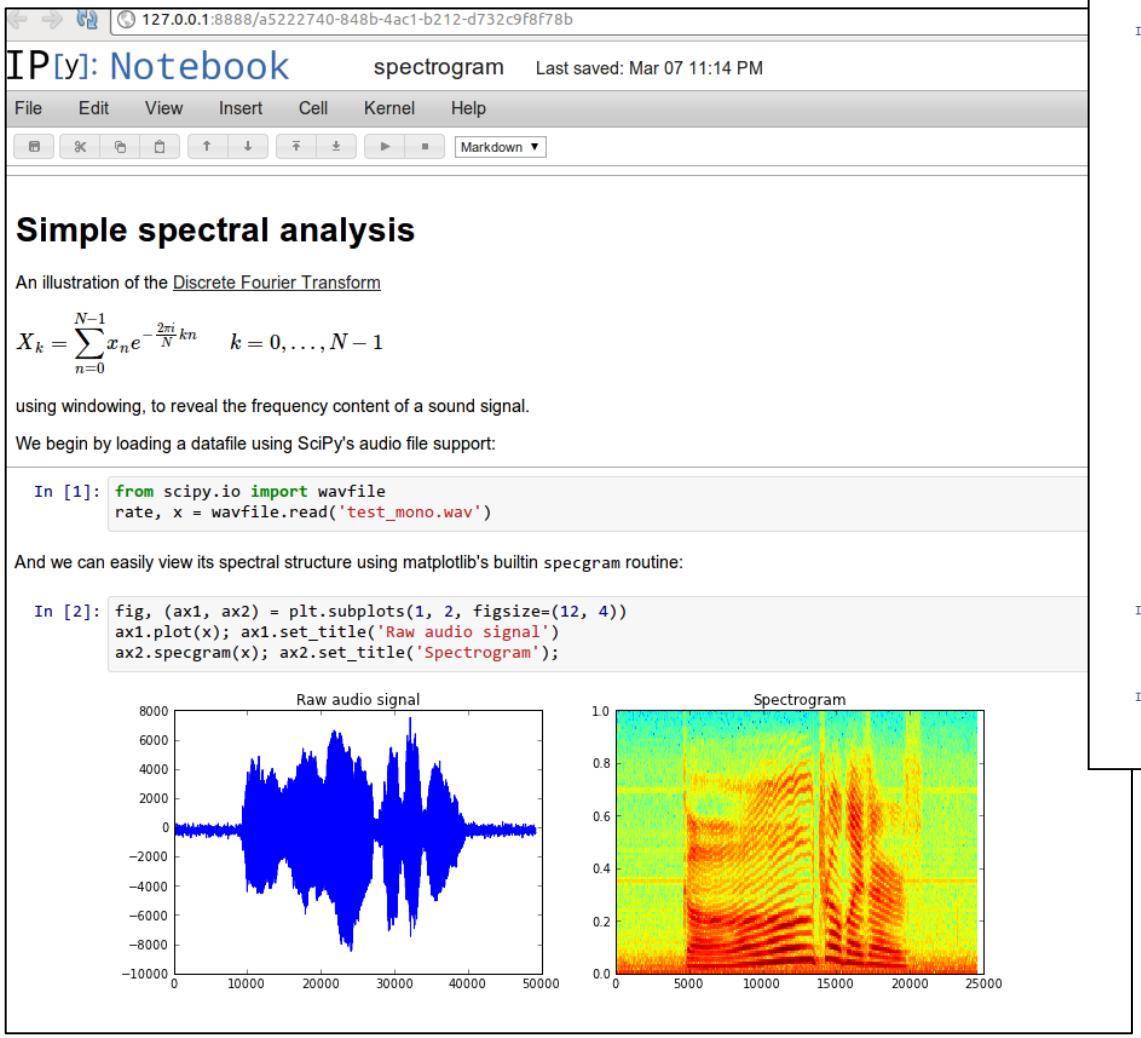
- Zim: http://zim-wiki.org/
- CherryTree: http://www.giuspen.com/cherrytree/

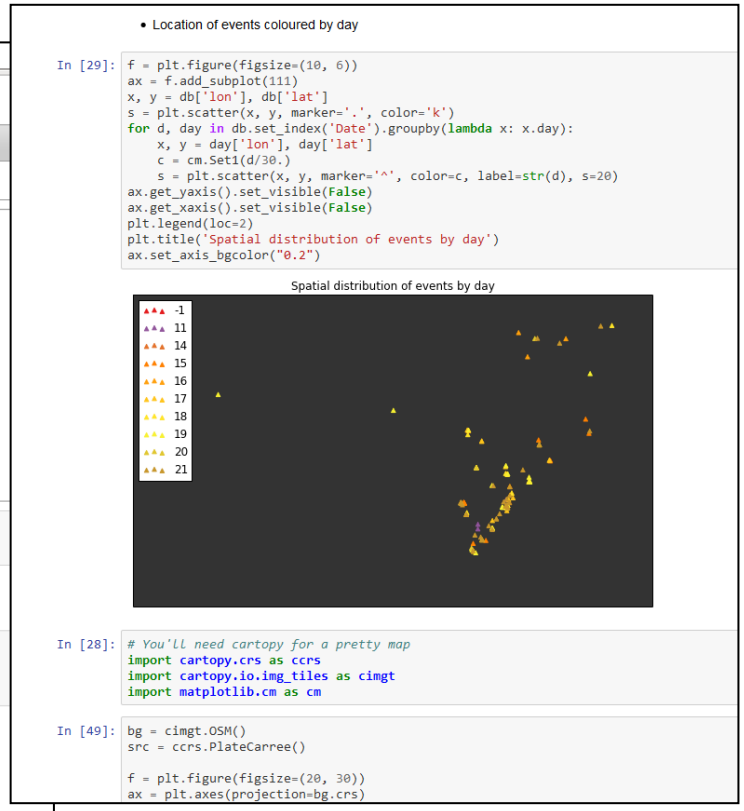




Taking notes - REPL

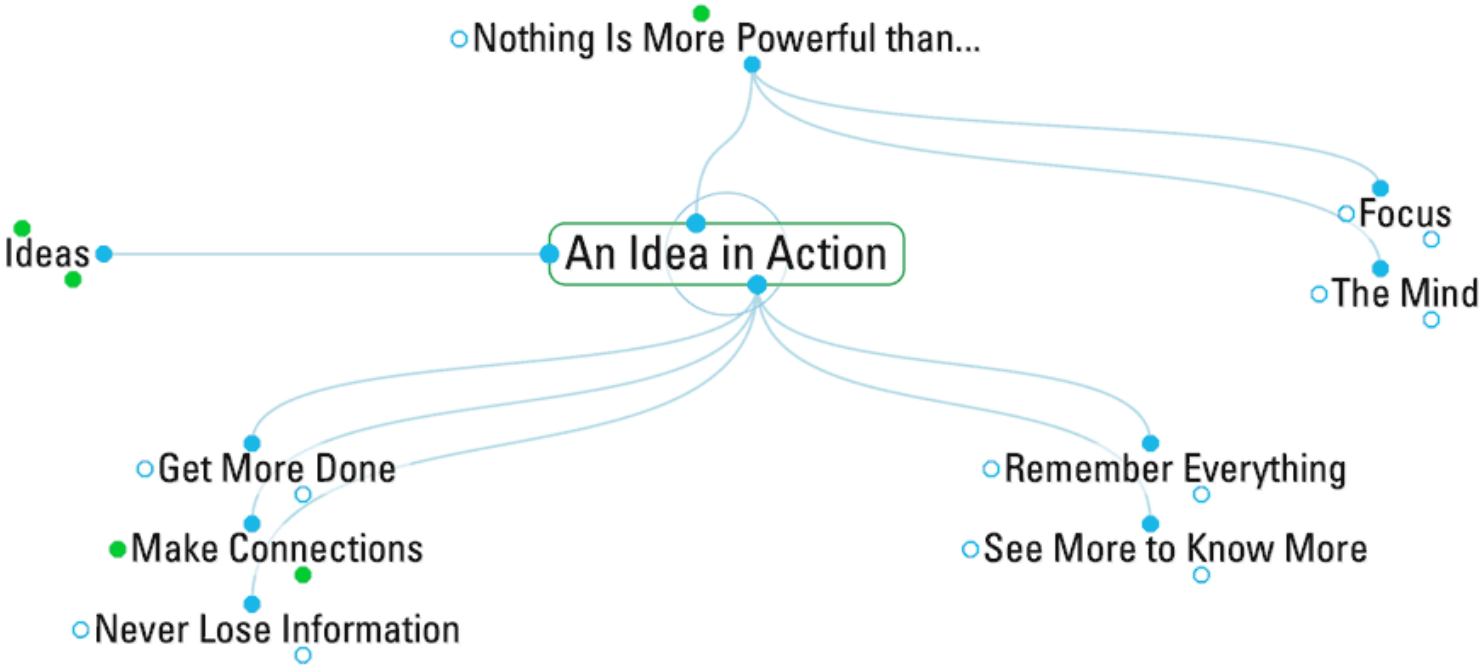
- Jupyter Notebook: http://jupyter.org





Taking notes – Mind Mapping

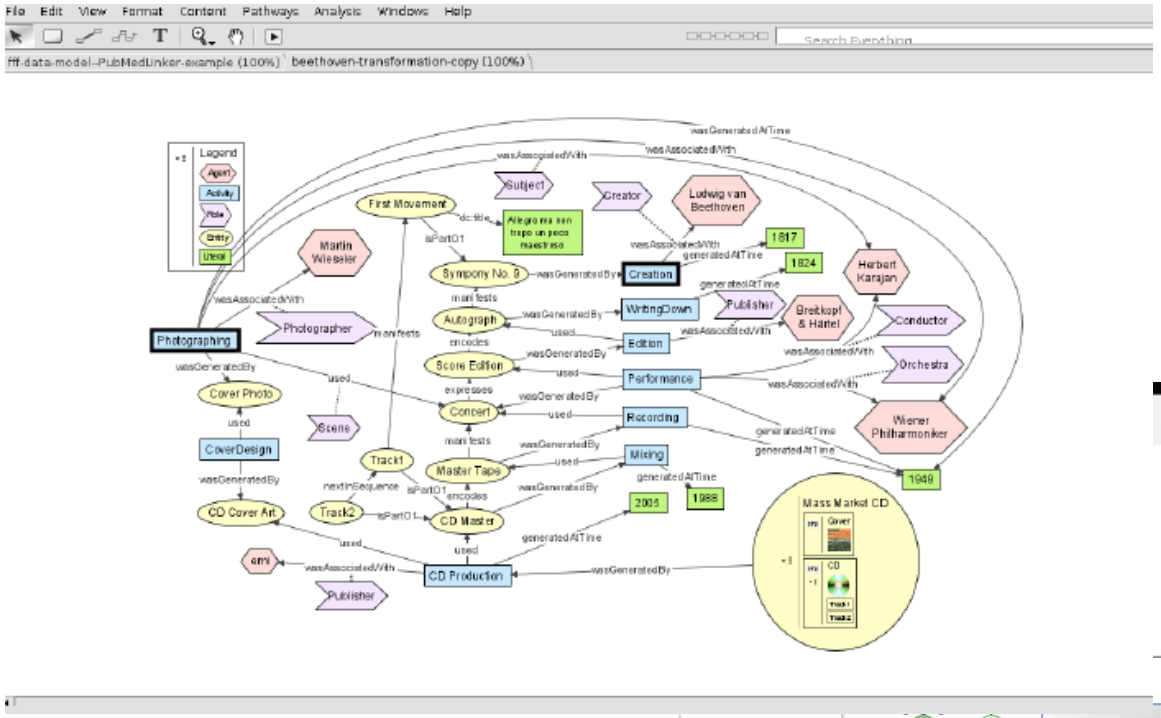
- mindmaps https://www.mindmaps.app/
- The Brain https://www.thebrain.com/



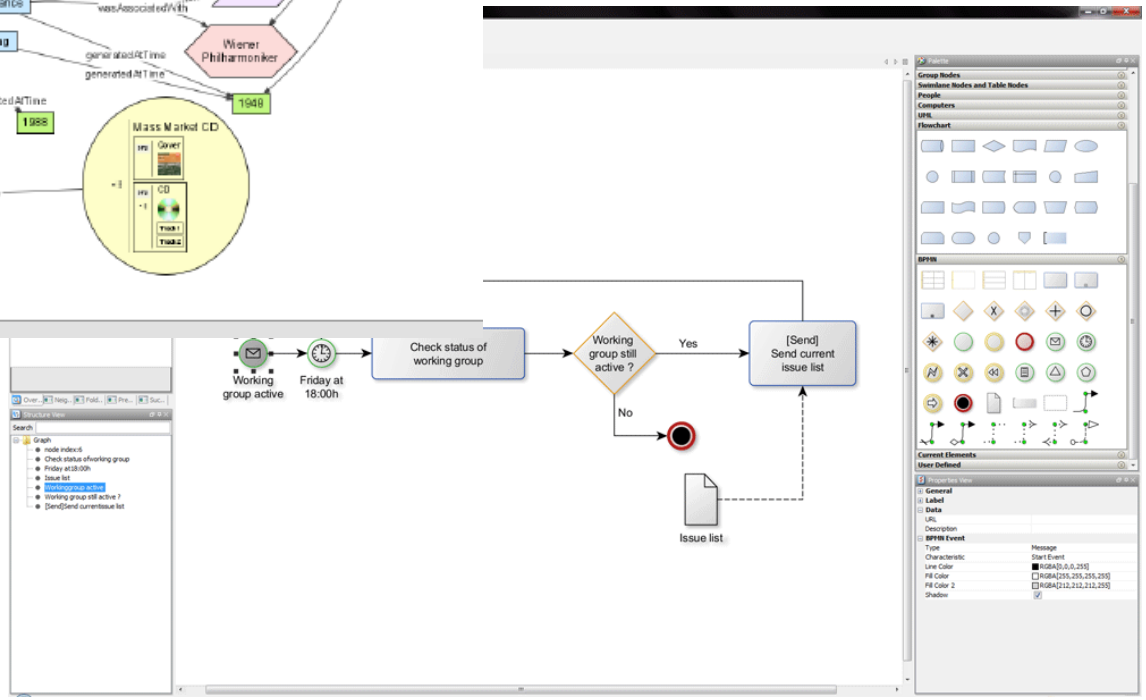
cf. Pensieve

Taking notes – Drawing Graphs

VUE: http://vue.tufts.edu/



yEd Graph Editor

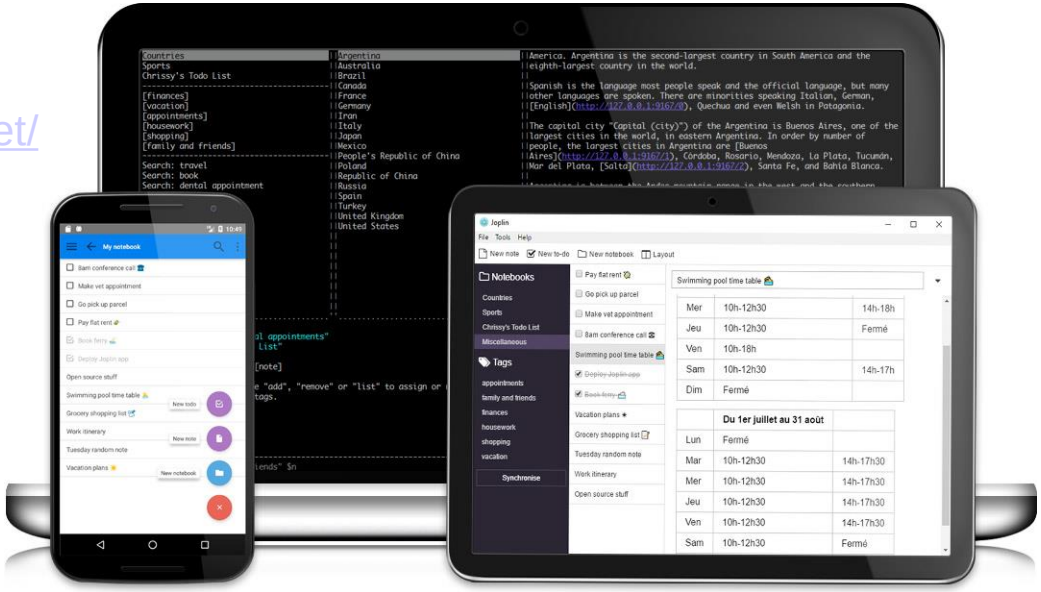


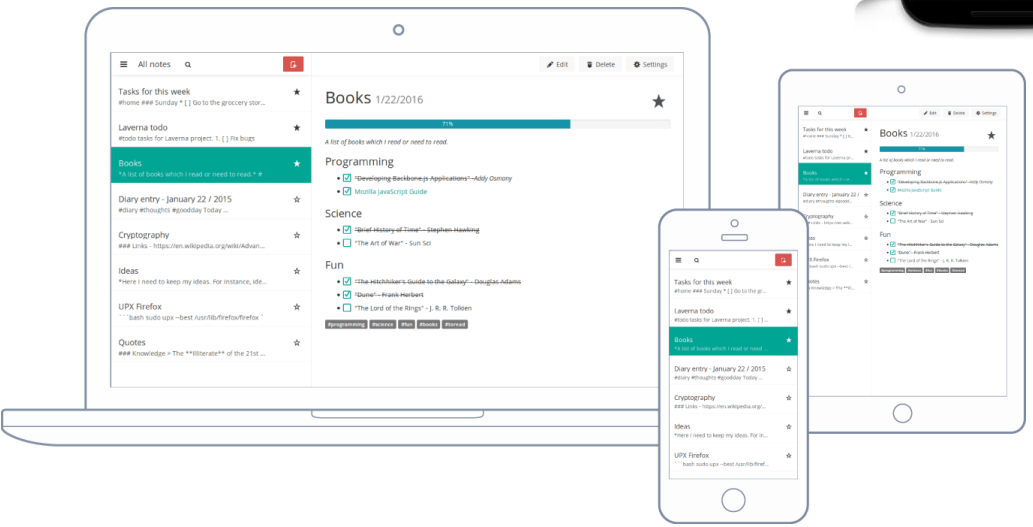
Taking notes – Secure and private Synchronization everywhere

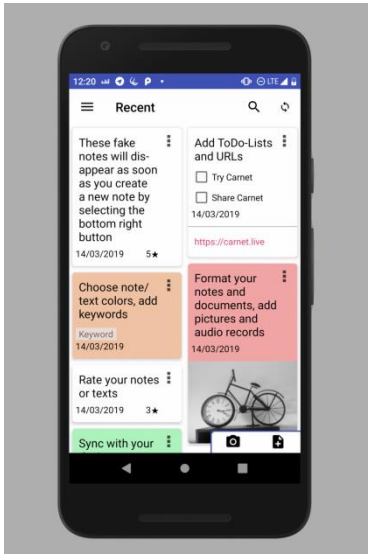
Joplin https://joplin.cozic.net/

Boostnote https://boostnote.io/

Laverna https://laverna.cc/index.html

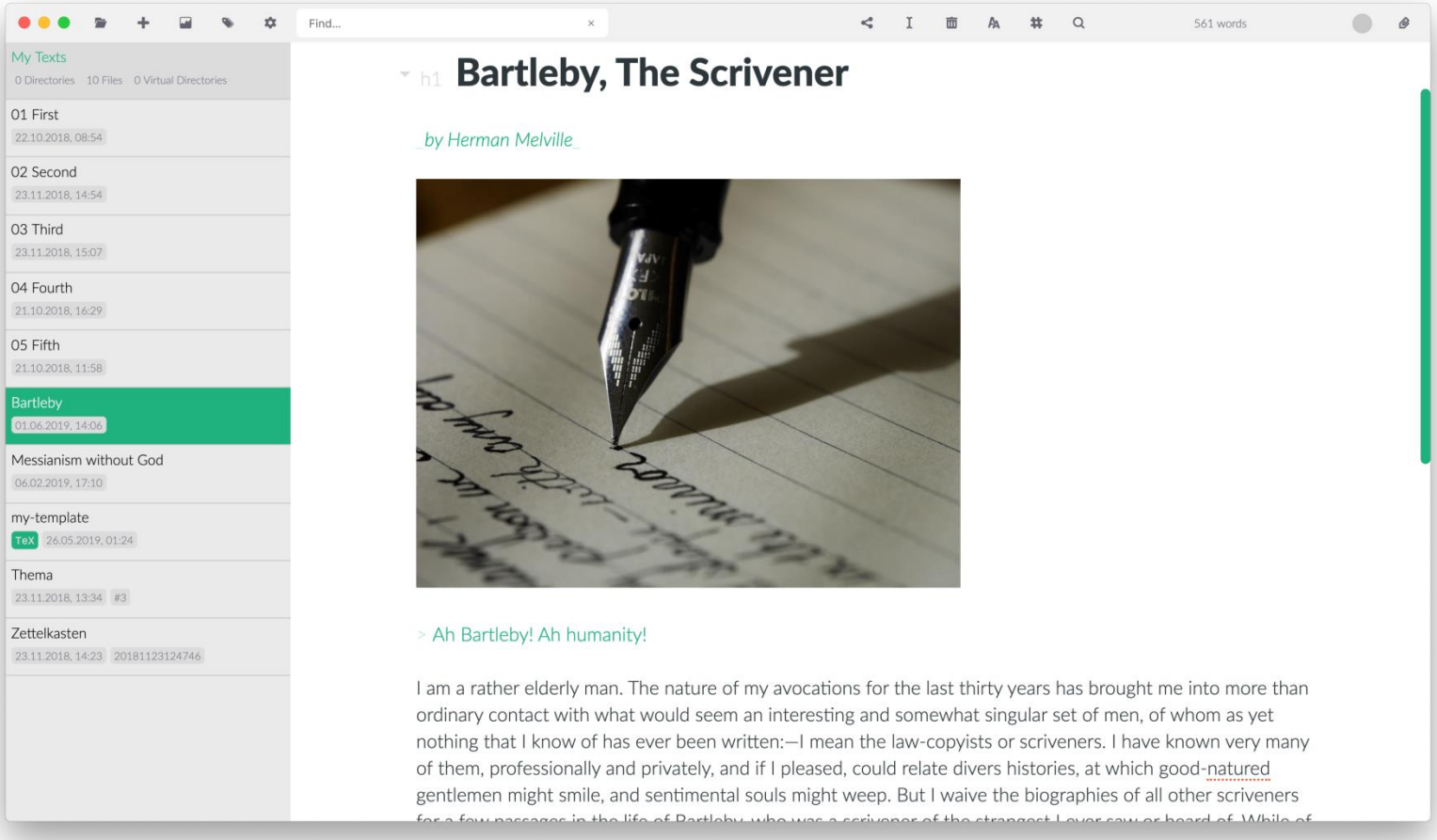






Carnet https://getcarnet.app/

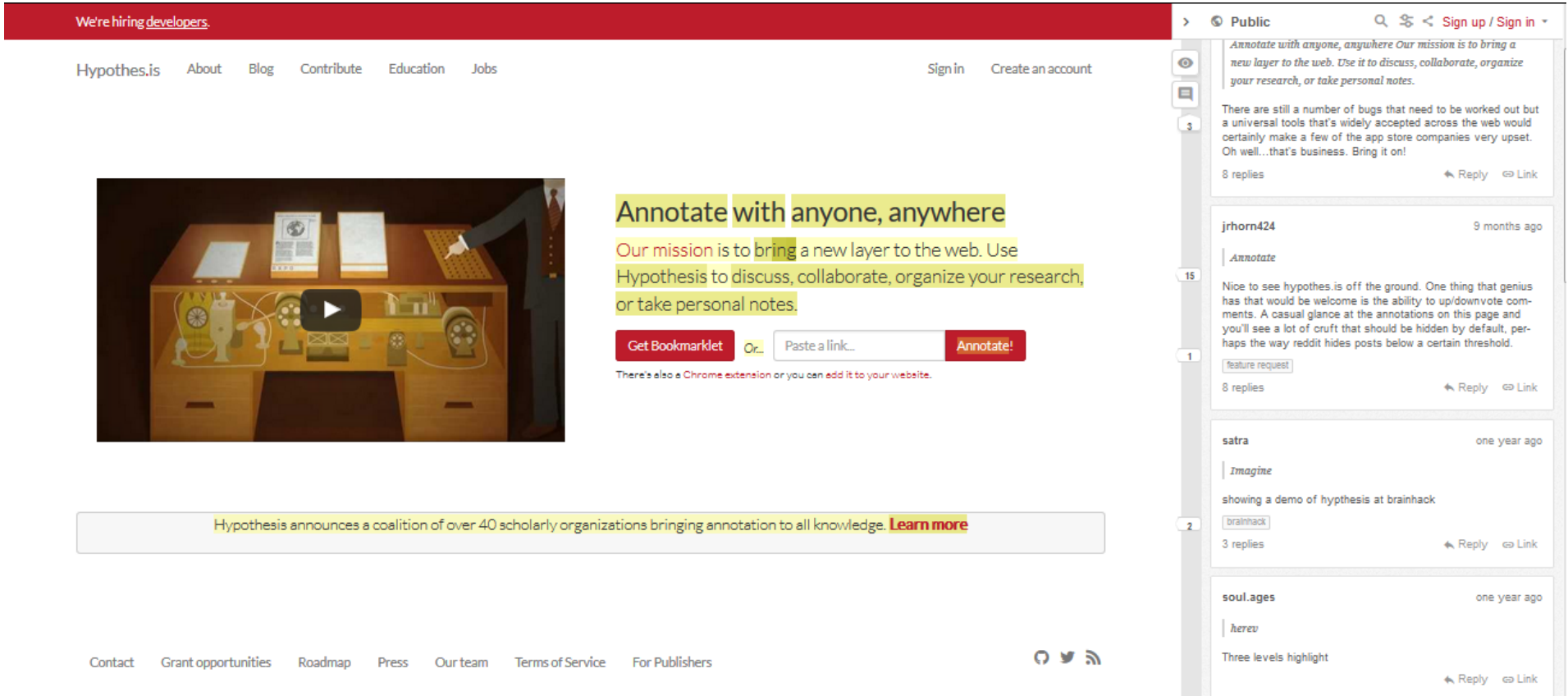
Taking notes – other Markdown editors



Zettlr https://www.zettlr.com/

Taking notes – Annotating the Web

Hypothesis: https://hypothes.is/



Revision control and backup

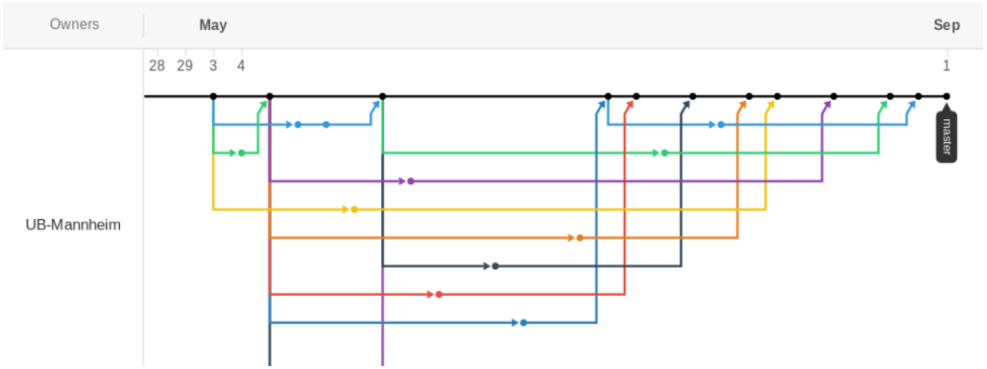
- Recommendation: (~~Subversion~~), GIT or Mercurial
  - Some groups already provide the infrastructure
  - Your adviser can access your code and text
  - Backup is done by the system administrator

- **Alternative**: free cloud storage, e.g. Dropbox, Google Drive, amazon cloud drive, iCloud, Telekom Mediencenter

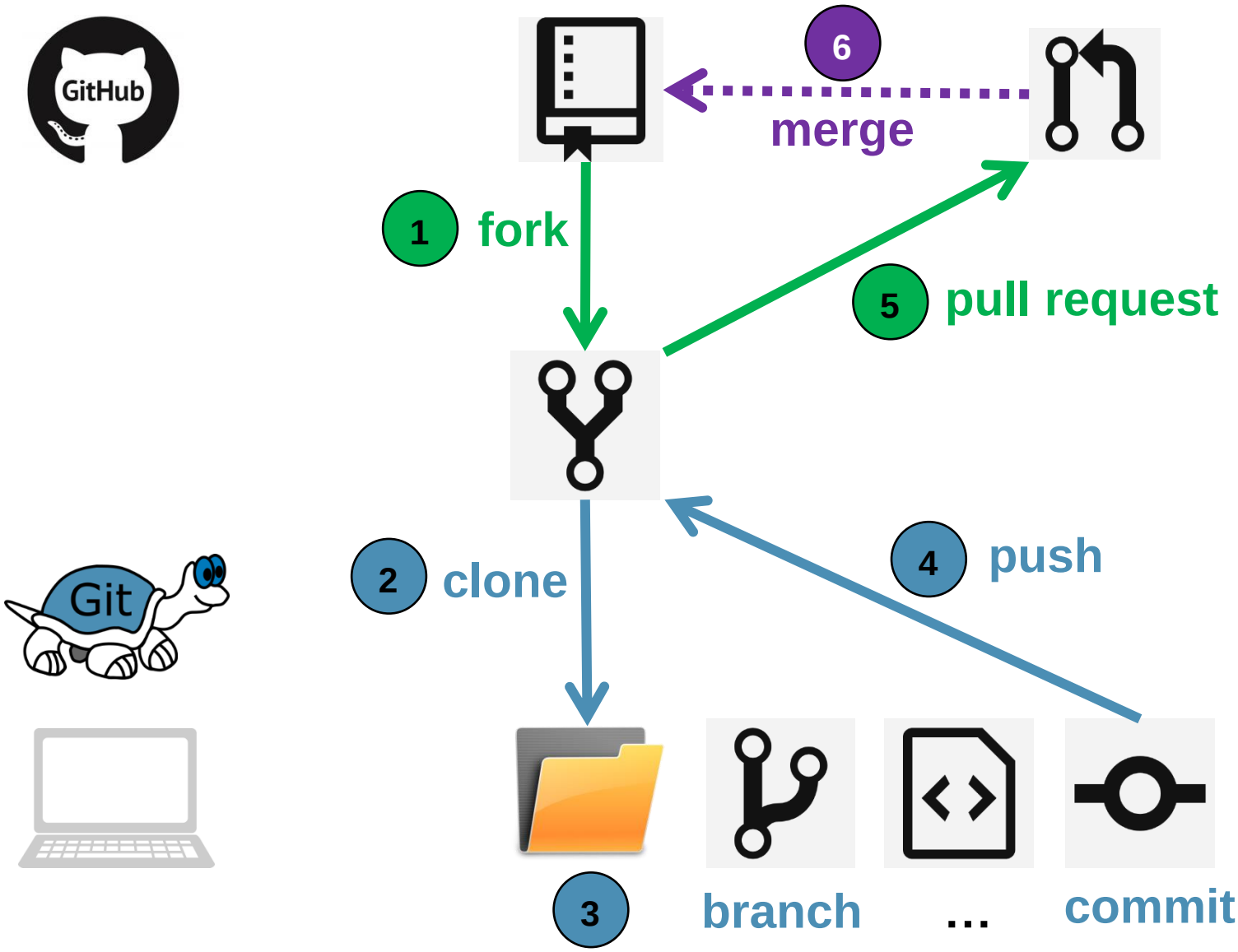
- Your own hosted cloud storage, e.g.

Distributed Revision control – The basics

- A repository contains all the files for a project
- Every change is recorded ("commit")
- Every copy of the repository contains all the data ("clone")
- Development is not necessarily linear ("branches")
- Branches can be reconciled later ("merge")
- Local changes can be sent, remote changes retrieved ("push/pull")
- **Commit early, commit often!**



Open Source Workflow with GitHub



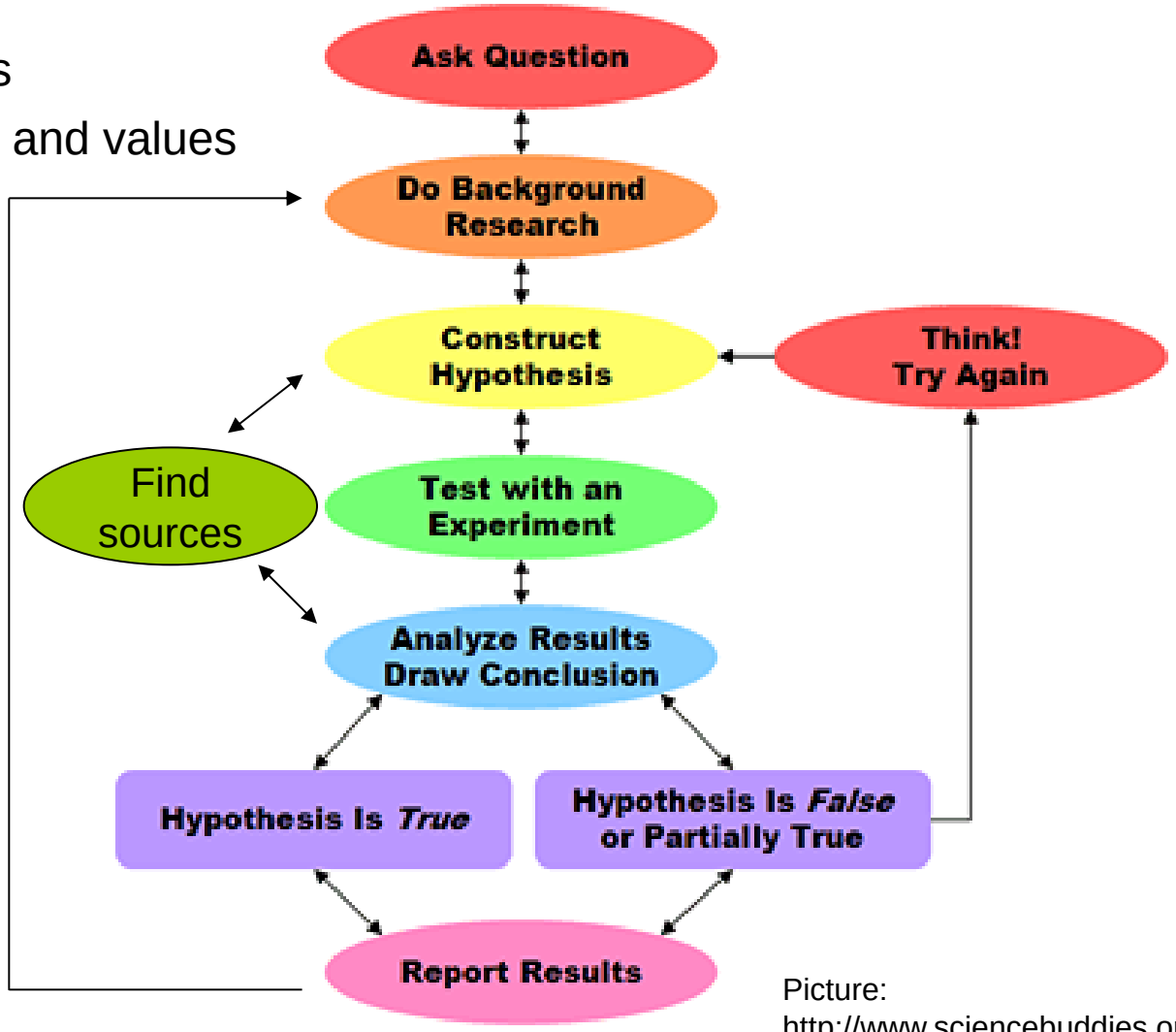
The scientific process

Overview

- Preliminaries
- Audience, scope and role of the work
- From topic to research question
- Supporting a claim: reasons and evidence

What is research?

- General: gathering information to answer a question
  - Curiosity
  - Preparing decisions
  - Questioning beliefs and values
- Scientific method



Picture:
http://www.sciencebuddies.org/

Targeting your audience

The control of cardiac irregularity by calcium blockers can best be explained through an understanding of the calcium activation of muscle groups. The regulatory proteins actin, myosin, tropomyosin, and troponin make up the sarcomere, the basic unit of muscle contraction.

Booth, Colomb, Williams: The Craft of Research. University of Chicago, 2008, 3rd ed., p. 16

Cardiac irregularity occurs when the heart muscle contracts uncontrollably. When a muscle contracts, it uses calcium, so we can control cardiac irregularity with drugs called calcium blockers. To understand how they work, it is first necessary to understand how calcium influences muscle contraction. The basic unit of muscle contraction is the sarcomere. It consists of four proteins that regulate contraction: they are actin, myosin, tropomyosin, and troponin.

Target audience: Average person?

„Hello IT....
Have you tried turning it off
and on again? … Uff, okay,
the button on the side, is it glowing? …
Yeah, you need to turn it on.“

Yesterday‘s Jam. *The IT Crowd*, Season 1, Episode 1, written by Graham Linehan, produced by Ash Atalla. Channel 4, 2006

Target audience: Expert with same background?

„Hello IT.... Yuhuh... Have you tried forcing an unexpected reboot? … You see the driver hooks a function by patching the system core table so it's not safe to unload it unless another thread is about to jump in there and do its stuff. And you don't want to end up in the middle of invalid memory.“

Yesterday‘s Jam. *The IT Crowd*, Season 1, Episode 1, written by Graham Linehan, produced by Ash Atalla. Channel 4, 2006

Audience

- Type of readers
  - Professionals
  - Well informed general readers
- Expectations
  - Entertainment
  - New facts
  - Help with understanding
  - Help with a practical problem
- Readers' background
  - Topic knowledge
  - Special interest
  - Recognized problems
  - New problems
  - Potential for controversy

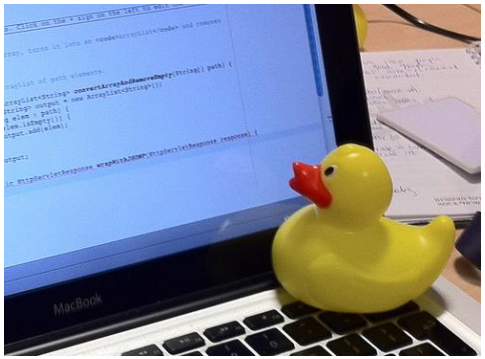
Think about your audience

- The professor is not your target audience
  - He might not need it for his research
  - But he will likely grade it based on its usefulness for others
- Think of fellow students as your target audience
  - Similar background to your own
  - Same classes and interests
  - If it is new for you, it needs to be explained to them
  - They are most likely to need your work for their own
- As always: Ask your advisor

Importance of discussing and writing

Importance of discussion

- Formulating vague ideas understandable
- How are your ideas seen with other eyes?
- Shows your weak points
- Find some wrong assumptions
- Improvements, ways to continue
- Rubber duck debugging: video explaining it, Wikipedia article



Picture: A rubber duck assisting with debugging some Java code in NetBeans / Tom Morris (2011-09-16) https://commons.wikimedia.org/wiki/File:Rubber_duck_assisting_with_debugging.jpg CC-BY-SA

Importance of writing

- To be useful, information must be shared.
- Writing permits validation of results.
- The published manuscript is the final step in research.
- It also is a measure of individual achievement.

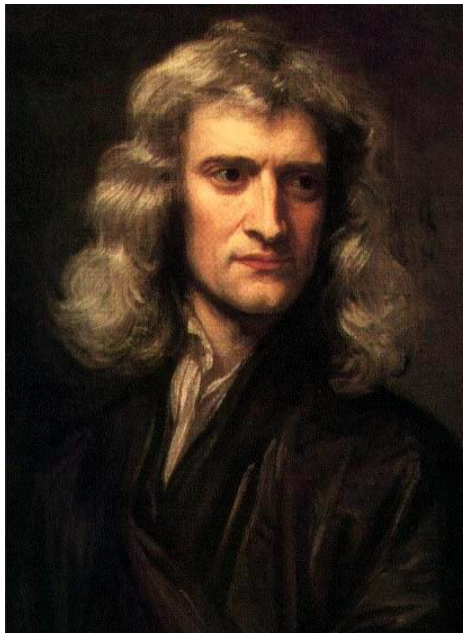
Importance of writing

- Written books can be seen

  as an ongoing conversation

  over the course of human history

- New knowledge is based on old knowledge



> If I have seen a little further it is by standing on the shoulders of Giants
>
> *Isaac Newton*

Picture:
http://en.wikipedia.org/wiki/Isaac_Newton

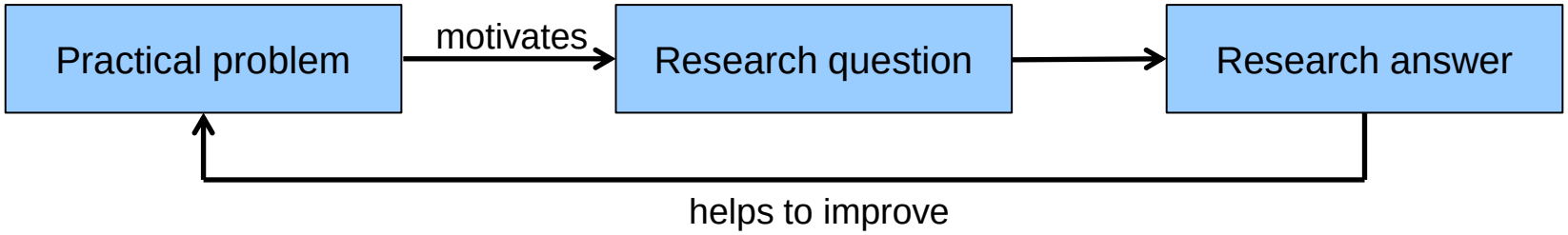
How to find a research question

From interest to topic

- Interests
  - Broad
  - Not limited to a single field or discipline
  - Often motivated by personal feelings and values
- Topic
  - Your own
  - Your advisor's
  - „Open question“ in the field
- Focused topic of **limited** scope
  - Aspects
  - Sources
  - Data

From topic to questions

- Focused topic can be expressed as a claim
- Surround the topic with questions
  - Parts and Relations
  - Place in broader context
  - Historical development
  - Characteristics
- Significance: A significant question?
  - Ask yourself “So what if I don't know it?”
  - Cost of not answering the question
  - Your motivation
  - Practical applications:



Thinking about problems

- A well defined research problem is the core of your thesis
- Make sure that you have a well defined research question in the beginning
- Common research problems
  - Application of known methods to new data
  - Comparison of different approaches to a known problem
  - Combinations of known approaches to solve a complex problem
  - ...

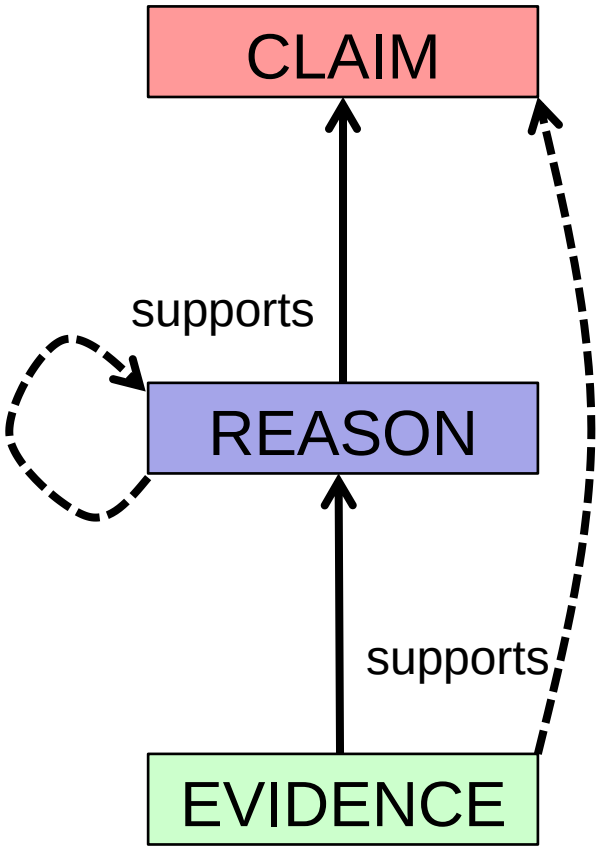
Example

- Interest
  - I am interested in artificial intelligence
- Topic
  - Using neural networks to forecast economic time-series data
- Focused topic
  - Economic forecasting using time-lagged feedforward neural networks with the objective of forecasting aggregate business sales
  - Focus on data: Standard and Poor's (S&P) 500 index and interest rates
  - Focus on tools: Mathworks' Matlab and NeuroDimension's NeuroSolutions
- Claim
  - Current stock market prices were correlated to past stock prices and it is possible with the tools to make significant predictions in the (historical) data
- Significance
  - If true, then it might also be possible to forecast future economic data.

Structure your research

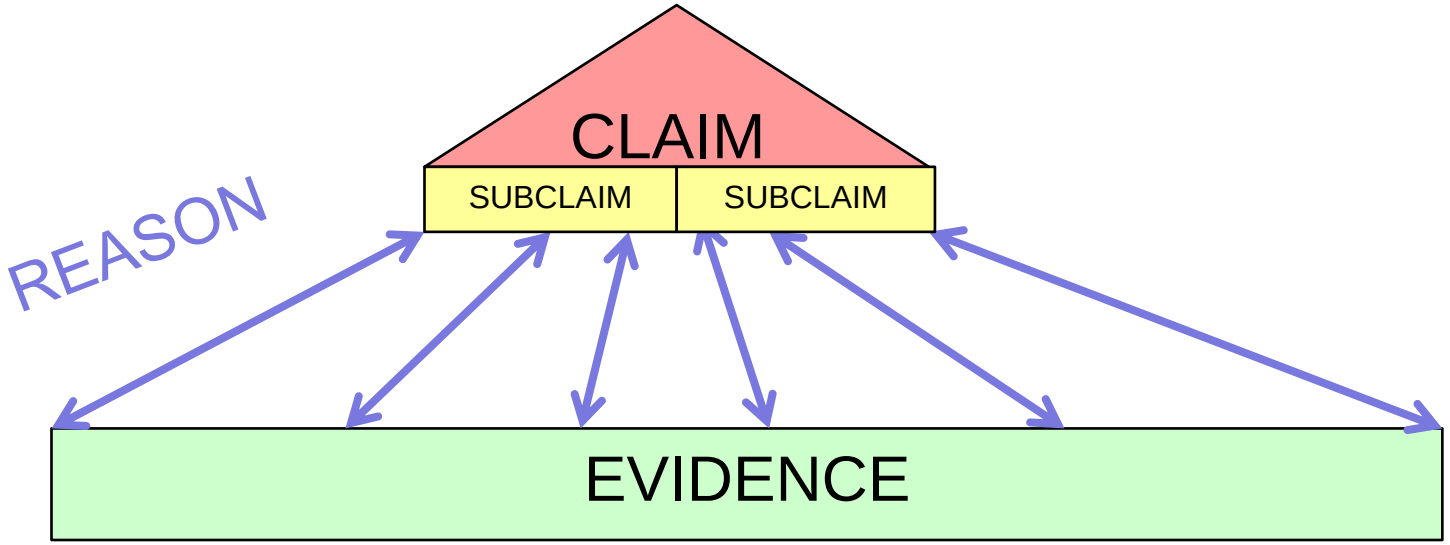
Elements of a research argument

- Claim
  - Your proposition
  - “Retrieval results can be improved by indexing with a thesaurus”
- Reason
  - Supports a claim (or a reason)
  - “because thesauri solve the synonym problem”
- Evidence
  - Supports a reason (or claim)
  - Is based on an observation or data
  - “in the experimental data set, retrieval recall improved by 200% while precision was reduced by only 5%”



Claim Reason Evidence

- The claim is the main point of a report
- Make specific claims
- Generic claims are hard to support
  - In the natural sciences, specific claims are preferred
  - Acknowledge limiting conditions (all data vs. some data)
- Connect claim and evidence through reason



Good Reasoning (Example 1)

[W]e argue that the company-internal processing of the market information provided by salespeople represents a critical resource that allows for the development of successful new products

via new product advantages and the adoption of new products by salespeople.

Data pertaining to 219 new product projects and 269 companies from various industries provide empirical evidence

that the intensity of sales force integration in the context of new product development significantly affects new product success beyond the effect of marketing integration.

Hildesheim, A. (2012) Internal Knowledge Exploitation – The Role of Sales Force Integration in New Product Development. Mannheim [Dissertation].

Good Reasoning (Example 2)

Like a conventional computer, [a differentiable neural computer (DNC)] can use its memory to represent and manipulate complex data structures, but, like a neural network, it can learn to do so from data. […]

When trained with reinforcement learning, a DNC can complete a moving blocks puzzle in which changing goals are specified by sequences of symbols.

Taken together, our results demonstrate that DNCs have the capacity to solve complex, structured tasks that are inaccessible to neural networks without external read–write memory.

Graves et al. in: Nature 538, 471–476 (27 October 2016)

How to argue scientifically

- Be open-minded
  - Start your argument with a state where everything is possible
  - Go your way, but consider possible objections and alternatives
  - Explain why the objections or alternatives do not apply
- Facts are key
  - A claim does not become true just because you want it this way
  - Deal with possibility that some of your initial claims might be wrong
- On the shoulders of giants
  - Logical foundations of reasoning
  - Refer to well-established definitions
  - Common practices (e.g. one would **not** use a survey to prove correctness of an algorithm)
- Short and to the point
  - If you can’t summarize your argument, maybe your claim is too broad?

Good/Bad Scientific Practice

Bad Reasoning

Darwin's theory of evolution is wrong.

Because parts of humans could not have evolved.

There are numerous irreducibly complex systems in nature.

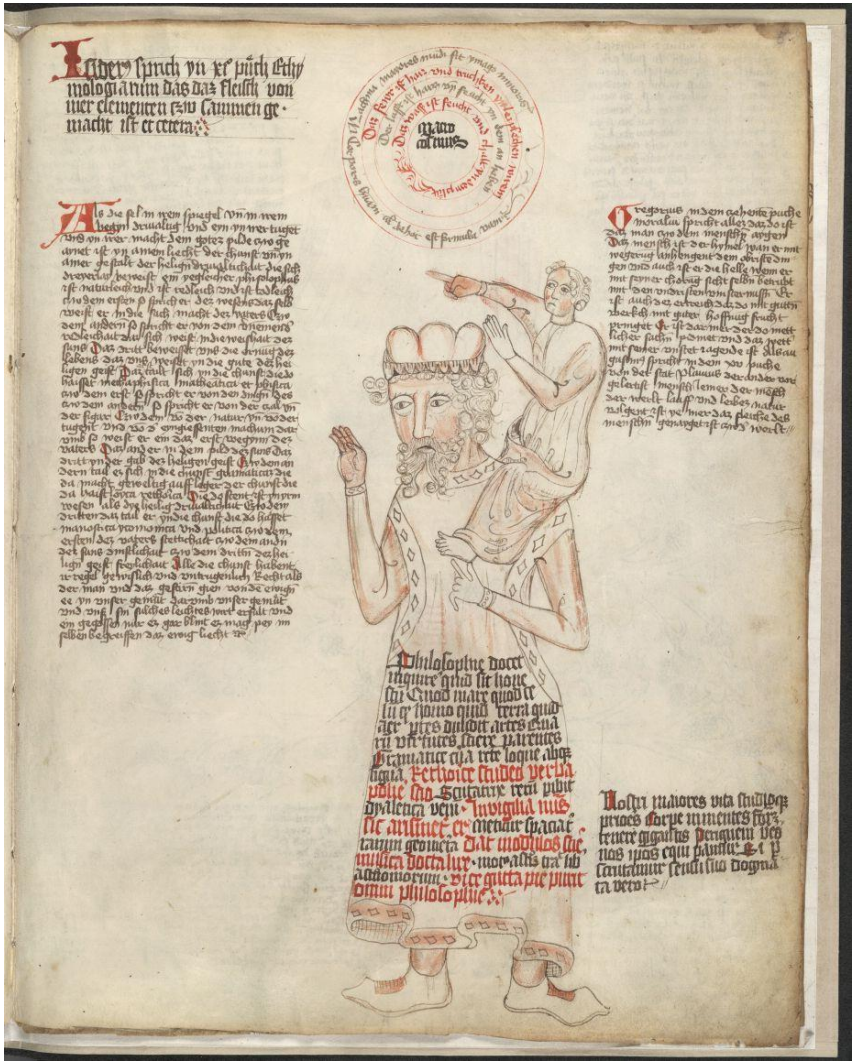
For example the human eye is an irreducibly complex system.

Irreducibly complex biological systems cannot be produced directly by slight, successive modifications of a precursor system.

What are actually "irreducibly complex biological systems" at all?

Michael Behe, Darwin's Black Box, S. 39ff

Standing on the shoulders of giants



Always **state on who's shoulders** you're standing!

Plagiarism is the one thing you absolutely must **avoid**!

Encyclopedic manuscript containing allegorical and medical drawings
South Germany, ca. 1410 Rosenwald 4 (image 15)
http://lcweb2.loc.gov/cgi-bin/ampage?collId=rbc3&fileName=rbc0001_2006rosen0004page.db&recNum=14

Plagiarism

Using someone else's work **without attribution**

→ End of career!
→ No degree!
→ Expulsion from university!

Graphic by „user8“ on http://de.guttenplag.wikia.com/wiki/Datei:Thumb_xxl.png retrieved on 2015-10-16

Copy and paste w/o attribution

Doe (2008), p. 18:

80% of respondents were tempted to procrastinate by using Facebook, hence we predict a similar pattern for Twitter usage.

You:

80% of respondents were tempted to procrastinate by using Facebook, hence we predict a similar pattern for Twitter usage.

Plagiarism

Attribute everything!

Doe (2008), p. 18:

80% of respondents were tempted to procrastinate by using Facebook, hence we predict a similar pattern for Twitter usage.

Plagiarism

You:

Based on the findings by Doe, that "80% of respondents were tempted to procrastinate by using Facebook" (Doe 2008, p.18), we strongly expect the same correlation in using Twitter.

Translation without attribution is still plagiarism

Doe (2008), p. 18

80% of respondents were tempted to procrastinate by using Facebook, hence we predict a similar pattern for Twitter usage.

You:

80 Prozent der von uns Befragten kamen in Versuchung, auf Facebook zu prokrastinieren.

Plagiarism

Rewrite, reorder, redefine – and it is still plagiarism

Kleiner, Lott (2006)

For all $\rho \geq 0$, put

$$D(\rho) := \sup\{\bar{R}_k\,(x, t_k) \,|\, k \geq 1, x \in B(x_k, \rho) \subset (M_k, \bar{g}_k(t_k))\},$$

and let ρ_0 be the supremum of the ρ's for which $D(\rho) < \infty$.

Cao, Zhu (2006):

For each $\rho \geq 0$, set

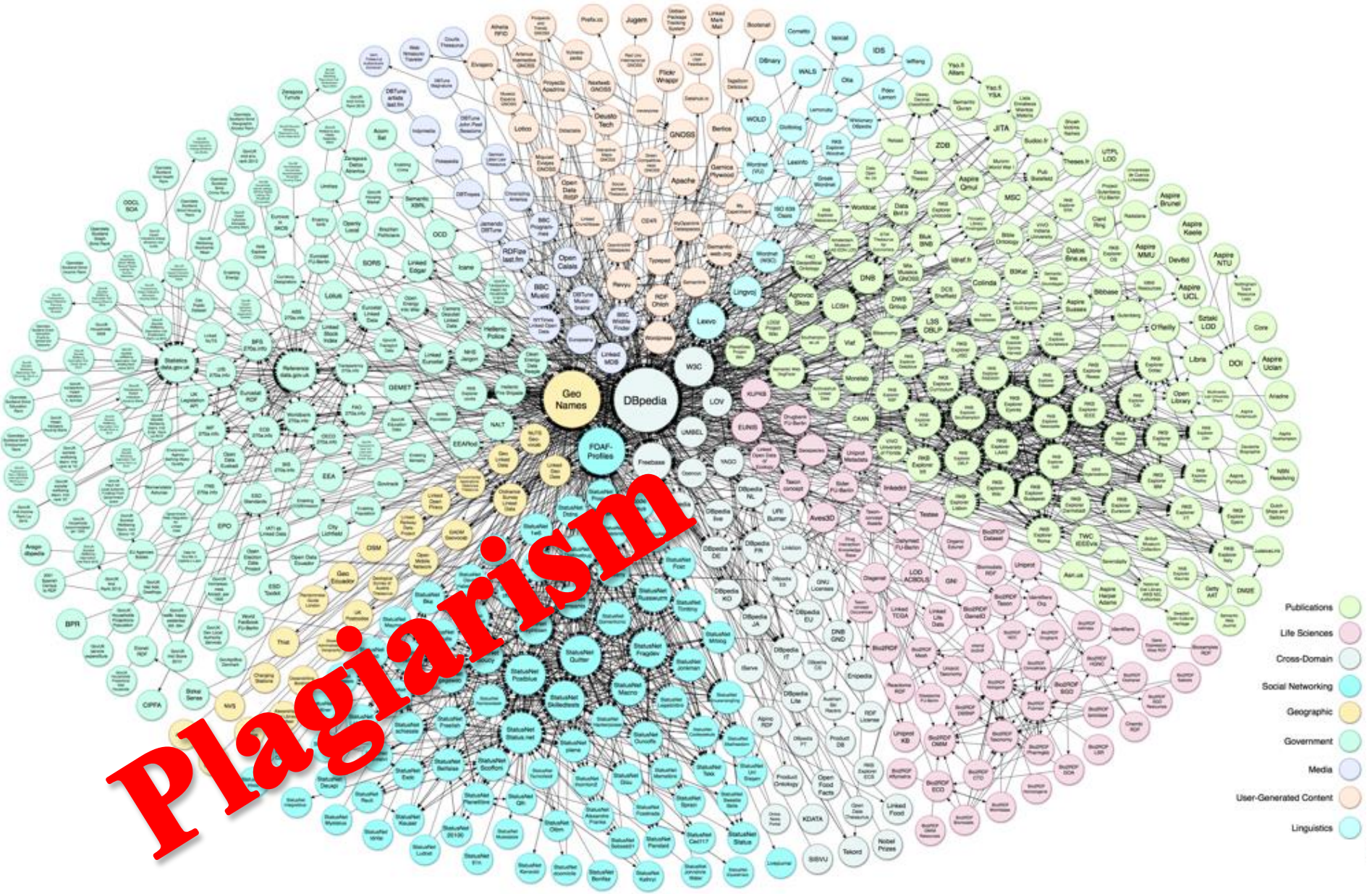
$$M(\rho) = \sup\{R(x, 0) \,|\, k \geq 1, x \in M_k \text{ with } d_0(x, x_k) \leq \rho\}$$

and

$$\rho_0 = \sup\{\rho \geq 0 \mid M(\rho) < +\infty\}.$$

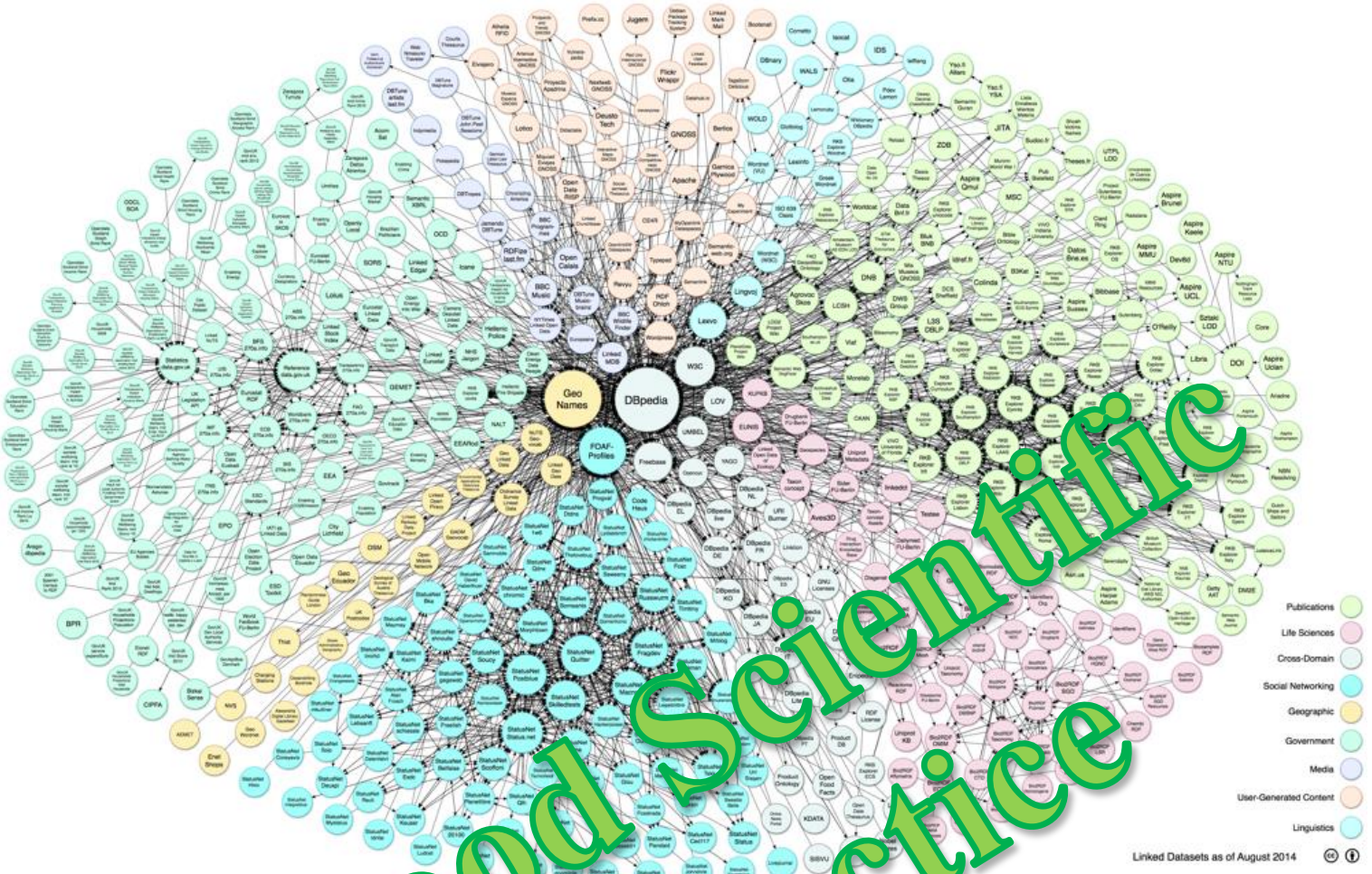
Plagiarism

Be even more careful about pictures/tables



Plagiarism

Attribute close to the quote



Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak: Linking Open Data cloud diagram 2014. http://lod-cloud.net/

Better be too thorough than too sloppy

The Big Bang theory is the prevailing cosmological model for the universe [1] from the earliest known periods through its subsequent large-scale evolution. [2][3][4] The model describes how the universe expanded from a very high-density and high-temperature state, [5][6] and offers a comprehensive explanation for a broad range of phenomena, including the abundance of light elements, the cosmic microwave background (CMB), large scale structure and Hubble's law. [7]

2 sentences

[1] Overbye, Dennis (20 February 2017). "Cosmos Controversy: The Universe Is Expanding, but How Fast?". New York Times. Retrieved 21 February 2017.

[2] Silk, Joseph (2009). Horizons of Cosmology. Templeton Press. p. 208.

[3] Singh, Simon (2005). Big Bang: The Origin of the Universe. Harper Perennial. p. 560.

[4] Wollack, Edward J. (10 December 2010). "Cosmology: The Study of the Universe". Universe 101: Big Bang Theory. NASA. Archived from the original on 14 May 2011. Retrieved 2017-04-15. "The second section discusses the [illegible] of the Big Bang theory that make it so compelling as the likely valid description of our universe."

[5] "First Second of the Big Bang". How The Universe Works 3. 2014. Discovery Science.

[6] "Big-bang model". Encyclopædia Britannica. Retrieved February 2015.

[7] Wright, E. L. (9 May 2009). "What is the evidence for the Big Bang?". Frequently Asked Questions in Cosmology. UCLA, Division of Astronomy and Astrophysics. Retrieved 16 October 2009.

7 references

Source: https://en.wikipedia.org/wiki/Big_Bang

Epilog and outlook

Learning from examples

- Some student thesis are online:

  https://madoc.bib.uni-mannheim.de/view/types/thesis.html

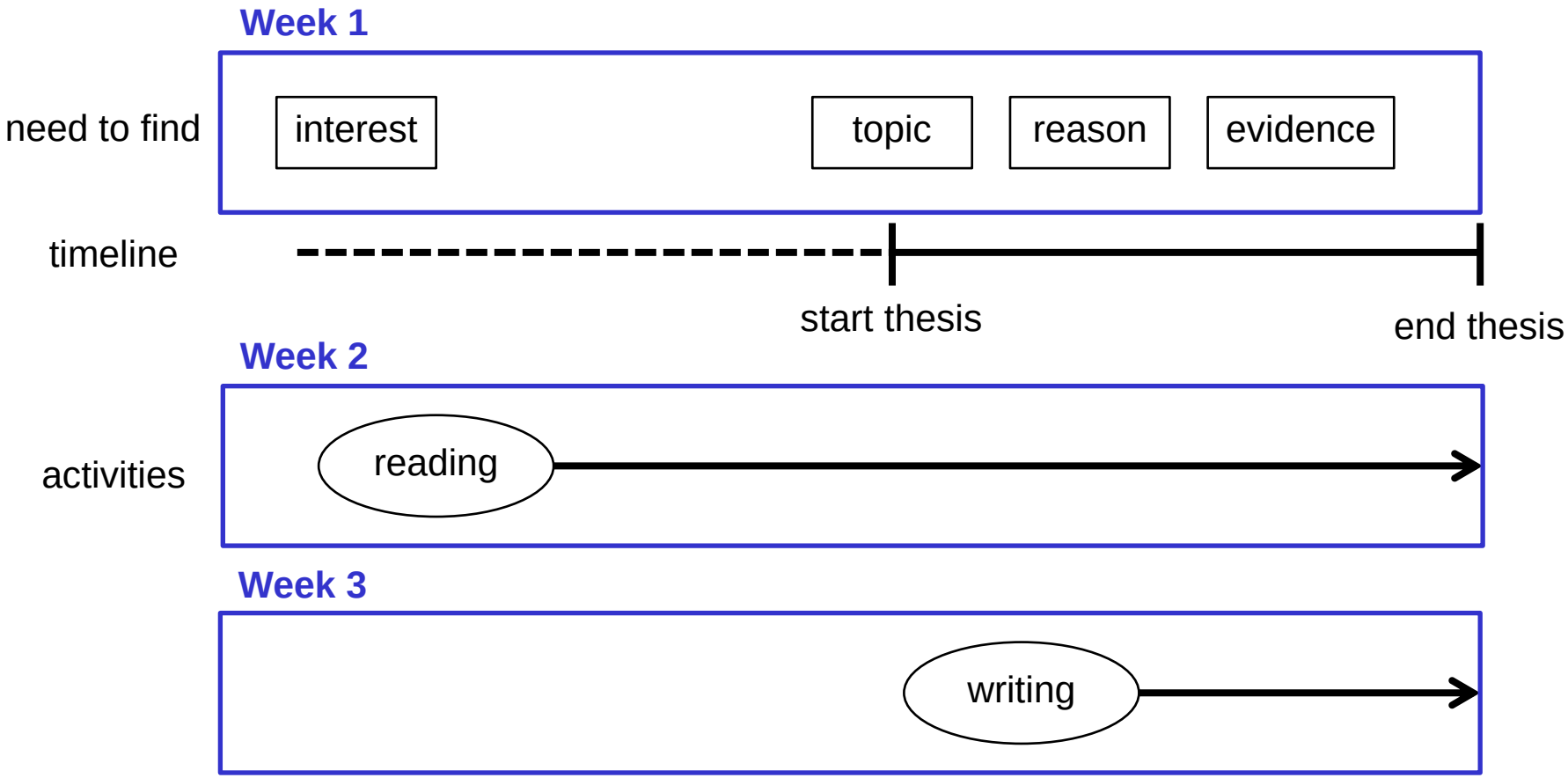
- They may serve you as some inspiration...
- Read the abstracts and introductions to journal articles and try to spot their argumentative structure

Preparing

- Know your target audience
- Know your main research question and the (expected) answer
- Sketch your argument structure
  - Main claim
  - Sub claims
  - Supporting reasons and evidence
- Consider objections and alternatives

→ You can now certainly start drafting

Overview


Week 1
need to find
interest
topic
reason
evidence
timeline
start thesis
end thesis
Week 2
activities
reading
Week 3
writing

Outlook for Week 2 and 3

- Reference management systems
- Sources
- Bibliographic research
- Tips / Strategies for search
- Writing and drafting
- Citations and bibliography

- Tips for writing
- LaTeX
- Exam