

Seminar Natural Language Processing (NLP) — Part 4

Basics of Scientific Literature

Henning Wachsmuth

<https://ai.uni-hannover.de>



Scientific literature

▪ Scientific literature

- **Narrow sense.** Academic papers that present original research by scientists
- **Broader sense.** Any document presenting scientific content

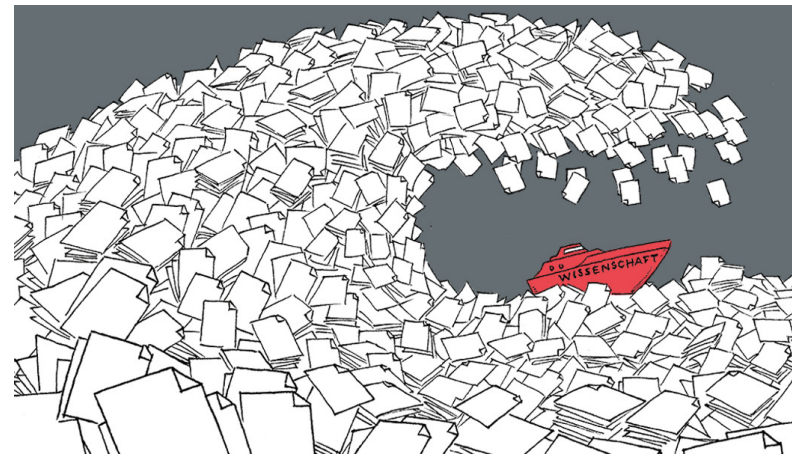
Both senses are covered here, even though the focus is on the narrow sense.

▪ Flood of publications

- No one can read all scientific literature
- >200 million papers overall
- >7 million papers published every year
- ~70% never cited by other papers

▪ Learning goals

- Why and how to make use of scientific literature
- What literature exists and how to assess its quality
- How to acquire and organize literature



<https://13s.de>, Jonas Hauss

Doing literature research

▪ Literature research

- Fundamental task in science
- Time-intensive and tedious — but necessary
- Often, the first task to be done



▪ Goals for presenting science (focus of seminar)

- Obtain all information relevant to the scope of a given problem
- Obtain background information
- Obtain evidence for your or others' claims

... and similar reasons

▪ Goals for doing science

- Find alternative approaches or perspectives
- Find out if your approach to a problem is new
- Don't reinvent the wheel

You are rarely the first to work on a problem. If you are, what does that tell you?

Selecting literature

▪ Types of literature

1. **Books.** Theory, basics, approved techniques
2. **Journal articles.** Completed research lines
3. **Conference papers.** State-of-the-art research
In NLP/CS, major publication type
4. **Workshop papers.** New ideas, ongoing research
5. **Preprints.** Papers inofficially published on free web platforms (e.g., arxiv.org)
6. **Conference/Online tutorials.** Easy access to basics and techniques
7. **Popular science magazines.** Public communication of research lines
8. **Other websites.** Anything



<https://pixabay.com>

▪ What types to use

- Generally, literature should be peer-reviewed (mostly holds for Type 1–4)
- Rule of thumb: **books > journals > conferences > workshops**
> tutorials > magazines > websites **preprints?**
- **But, for example: top conferences > average journals**

The symbol > stands for “preferred over” here

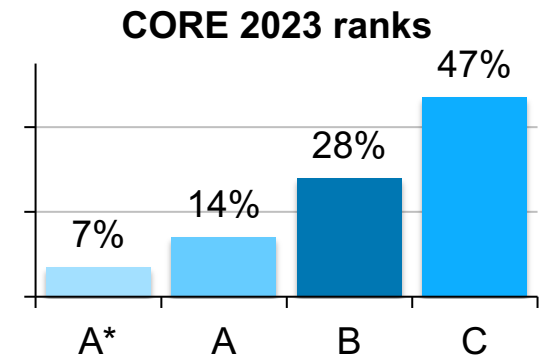
Assessing quality of literature

▪ Conference rankings

- Best is A*/A; B still reasonable; unranked doubtful

No ranking has complete coverage, though

- A well-reputed ranking is CORE core.edu.au/conference-portal



▪ Journal rankings

- Partitioned into quartiles Q1–Q4 (Q1 best, Q2 still reasonable)
- Journals also have impact factors derived from citations (reasonable as of 4+)
- A good place to check Journals is Scimago www.scimagojr.com

▪ Number of citations

- Indicator of importance, rather for *relative* comparisons within a topic

Notice: Newer papers naturally tend to have fewer citations

- A good resource for citation numbers is Google Scholar scholar.google.de

▪ Disclaimer

- Good and bad research appears at all places
- Often, only reading helps

Reading and finding literature

▪ **Reading a paper efficiently**

1. Read abstract, introduction, and conclusion
2. Look at figures and tables
3. Decide whether worth reading everything
4. Read goal-driven

Specify questions to be answered during reading



<https://pixabay.com>

▪ **Exemplary questions to be answered**

- Is the given paper really relevant to the topic of interest?
- Does the given paper provide evidence for the fact to be backed up?
- How do the methods in two given papers on the same task differ?

▪ **Finding the next paper**

- Follow promising references at the end of a given paper
- Find promising papers citing the given paper
- Learn to identify the best search terms

Rule of thumb: As specific as possible, but as abstract as needed

Acquiring literature

▪ Obtaining papers

- Many papers freely available online
- Others might be free from a university network
- If neither, maybe your advisors can help



<https://pixabay.com>

▪ Important sources

- [ACL Anthology](http://aclweb.org/anthology) for most core NLP papers aclweb.org/anthology
- [ACM Digital Library](http://dl.acm.org) for many important computer science papers dl.acm.org
- [DBLP](http://dblp.dagstuhl.de) for any literature related to computer science dblp.dagstuhl.de
- [Google Scholar](http://scholar.google.de) for any scientific literature scholar.google.de

... along with general web search

▪ Accessing books

- Check whether available in the library (physically or digitally)
- Some accessible online, for example, on [Google Books](http://books.google.de) books.google.de

Purchasing books can make sense when of continuous importance to you

Organizing literature

▪ Literature organization

- Maintain overview, start from the beginning
- “Extra” effort will pay off

▪ Create logical folder structure

- Build your own view of the field
- Logically subdivide topics, but don't over-engineer

For instance, `./literature/computational-sociolinguistics/social-bias/` — but maybe not deeper

▪ Rename all literature consistently

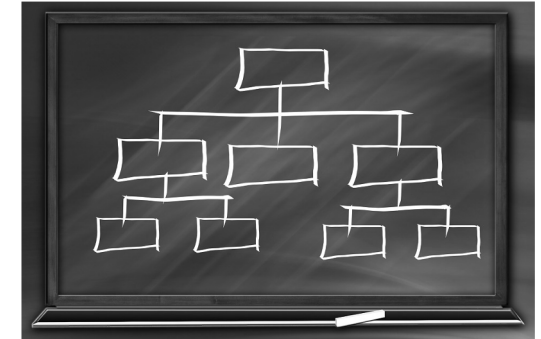
- Simplifies browsing and finding
- We use `<1stauthor><2digityear>-<full-title-lower-case-no-special-chars>.pdf`

For example: `ajjour17-unit-segmentation-of-argumentative-texts.pdf`

▪ Organizing meta-information

- Bibliographical information needed when citing literature
- Store bibtex of literature whenever available

Learn more on en.wikipedia.org/wiki/BibTeX; many pages provide bibtex entries



tools for literature organization exist
e.g., Citavi and Zotero

Beyond scientific literature

▪ Open Research Knowledge Graph (ORKG)

- Most scientific literature is largely unstructured (text, images, etc.)
- TIB/L3S/LUH investigate how to make scientific knowledge more structured



Overview

Why ORKG?

Team

History

Media Coverage

Publications

Workshops

Outreach

ScienceGRAPH

SciNEXT

Training

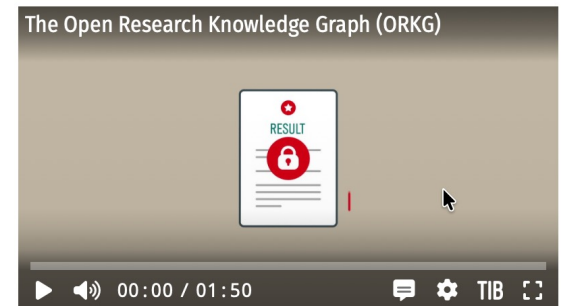
Research is a fundamental pillar of societal progress. Yet, scientific communities face great difficulties in sharing their findings. With approximately 2.5 million newly published scientific articles per year, it is impossible to keep track of all relevant knowledge. Even in small fields, researchers often find themselves drowning in a publication flood, contributing to major scientific crises such as the reproducibility crisis, the deficiency of peer-review and ultimately the loss of knowledge.

The underlying problem is that we never updated our methods of scholarly communication to exploit the possibilities of digitalization. This is where the **Open Research Knowledge Graph** comes into play!

The ORKG makes scientific knowledge human- and machine-actionable and thus enables completely new ways of machine assistance. This will help researchers find relevant contributions to their field and create state-of-the-art comparisons and reviews. With the ORKG, scientists can explore knowledge in entirely new ways and share results even across different disciplines.

Get started or **get involved** now to tackle current and future scientific challenges. Let's get it done. Collaboratively and openly.

Want to learn more? The lecture "**Towards Knowledge Graph based Representation, Augmentation and Exploration of Scholarly Communications**" by our lead Sören provides an overview of the ORKG initiative.



<https://orkg.org>

Takeaways

▪ **Scientific literature**

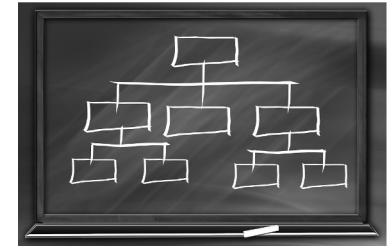
- Academic papers, books, tutorials, and more
- From basics and techniques to state-of-the-art research
- Literature research is a fundamental part of scientific work



<https://flickr.com>

▪ **Aspects covered here**

- Types and quality of scientific literature
- Efficient reading of papers
- Acquisition and organization of literature



<https://pixabay.com>

▪ **Literature in the seminar**

- First read the paper that we provide
- Then find further literature where needed or helpful
- For your presentation, focus on academic papers



<https://pixabay.com>