Natural Language Processing

Working with research literature

Marco Kuhlmann

Department of Computer and Information Science

Why are we having this seminar?

The ability to work with research literature is a learning outcome:

On completion of the course, you should be able to identify, assess, and make use of NLP research literature.

This is tested both in the group part and the individual part:

- Project rubric
- Guidelines for the post-project paper

This session

- How to find and cite research articles in the area of NLP

- How to assess articles in terms of their quality

- How to relate your work to the research literature

How to find and cite research articles

How to find research articles

- [ACL Anthology](#)

- [Google Scholar](#)
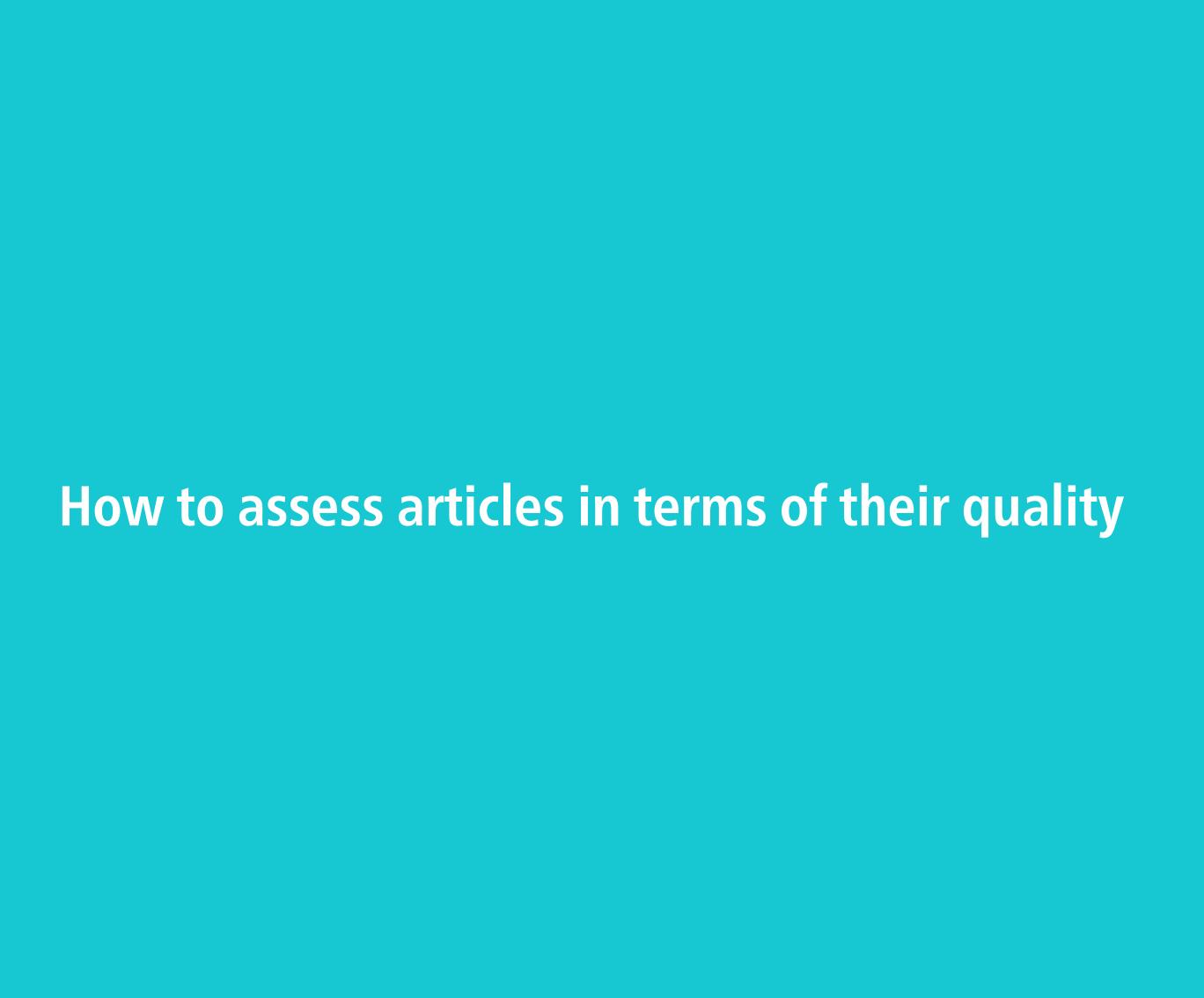
How to cite research articles

- Find the BibTeX for the article you want to cite in the <u>ACL Anthology</u> (preferably) or on <u>arXiv</u>.

- Add the BibTeX entry at the end of the `custom.bib` file inside the <u>Overleaf</u> project for your paper.

- Cite the article using `\citep` (Cooley and Tukey, 1965) or `\citet`: Cooley and Tukey (1965).

  For more alternatives, see the examples in Table 2 of the instructions.

Tasks 1

Screen the sample articles to find the following:

- Examples of incomplete entries or inconsistencies across entries in the lists of references of the sample articles.

- Examples of arXiv preprints that are also available as regular articles in the ACL Anthology.

- Examples of in-parenthesis (`\citep`) and in-text (`\citet`) citations. Did the authors use them correctly?

How to assess articles in terms of their quality

Peer-reviewed research

- **Peer review** is the quality assessment of research by other researchers in the field (peers).

- In NLP, peer review is typically **double-blind**: the reviewers do not know the authors, the authors do not know their reviewers

- Example: <u>OpenReview</u>

Pre-prints

- A **pre-print** is a version of a research article published before or concurrently with formal peer-review.

- The most well-known pre-print server is <u>arXiv</u>.

- If an article has been accepted to a conference or journal, cite the peer-reviewed version, not the pre-print.

Rankings of publication channels

- Different professional organisations (often research councils) provide lists with rankings of publication channels.

- Two well-known examples:

  - CORE Conference Rankings (conferences)

  - Norwegian List (journals)

- Ranking criteria include factors such as quality of the peer-review process, citation impact, and acceptance rates.

Top-ranked conferences in the area of NLP

- ACL, Meeting of the Association for Computational Linguistics

- EMNLP, Empirical Methods in Natural Language Processing

- EACL, Conference of the European Chapter of the ACL

- NAACL, Conference of the North American Chapter of the ACL

Level-2 journals within NLP

- <u>Computational Linguistics</u>

- <u>Natural Language Engineering</u>
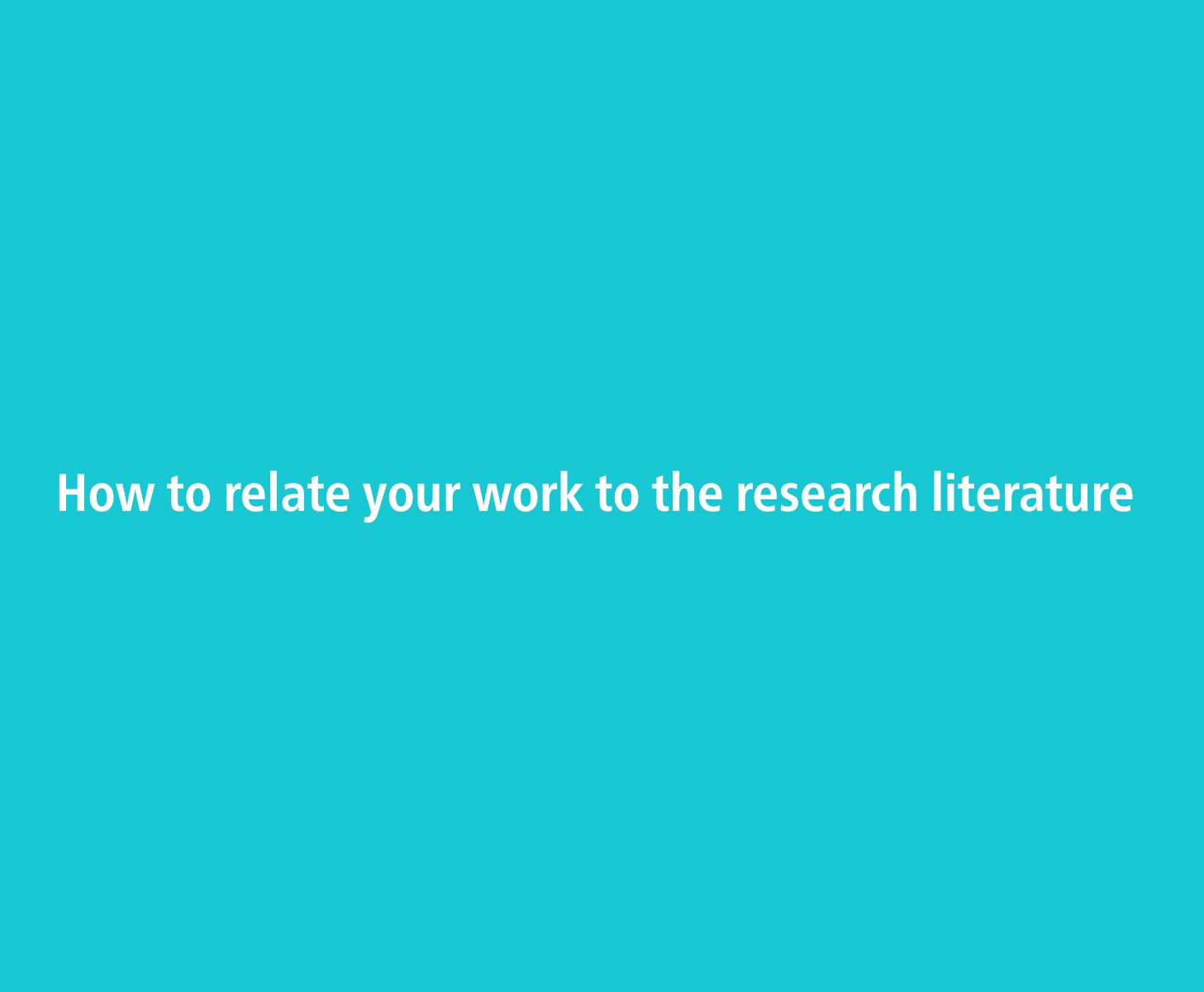
Citation indexes

- The impact of a paper and its authors can be compared using different measures.

- **H-index:** An author with h-index k has written at least k papers, each of which has been cited at least k times.

- **I10-index (Google Scholar):** Number of publications of the author with at least 10 citations.

Tasks 2

Look up the following people on Google Scholar:

- Marco Kuhlmann

- Joakim Nivre

- Chris Manning

What are their h-indexes? Looking at these profiles, do you see any potential problems with this way of measuring impact?

How to relate your work to the research literature

How to relate your work to the literature

- Implement a method described in the literature.

  as the baseline of your project, as a method you want to evaluate

- Use previous work to set the stage for your work.

  Syntactic parsing is an important part of many NLP pipelines (Smith, 2016).

- Use previous work to show that your contribution is important.

  Smith (2016) calls for language-specific feature engineering, but, to the best of our knowledge, no work on this exists for Bulgarian.

How to relate your work to the literature

- Compare your results to the results of previous work.

  Our accuracies are at the same level as those of Smith (2016).

- Compare your method to that of previous work.

  Our method extends that of Smith (2016) by using attention even in the decoder part of the system.

- Point to related work for future work.

  In the future, it would be interesting to extend the evaluation of our system to the benchmark presented by Smith (2016).

Tasks 3

Skim through the sample articles. For each citation, describe how the authors make use of the cited work.

Project rubric – Research literature

- Your presentation shows awareness and understanding of the research literature, and it is clear how your work compares to that work. (30)

- Your presentation features a detailed and enlightening discussion of the research literature (e.g., comparison of results and/or method). (60)

Project rubric – Soundness

- Reasonable work. The method is not bad, the analysis is appropriate, and the main claims are probably correct. (30)

- The approach is very apt, and the claims are convincingly supported by proper experiments, theoretical analysis or <u>related work</u>. (60)

Post-project paper – Description

- Clear and precise account of the project idea, its results and your role in it. Several details and examples. (30)

- Includes experience from <u>additional reading of scientific literature</u>. Focus on aspects significant for your learning. (60)

  You must have read articles that were not already part of the presentation <u>and</u> include this experience into your paper.

Post-project paper – Examination

- Considers how specific concepts and skills emerged in the project and explains them so that an outsider can understand them. (30)

- Develops an enhanced understanding of concepts and skills from the course in light of the project and any additional reading. (60)

  It must be clear that the additional reading was instrumental for you developing an enhanced understanding of the course content.

Post-project paper – Articulation

- Clear what you learned, how exactly you learned it, and how the learning is relevant to the learning objective. (30)

- Places the learning from the project in a broader context by explaining its relevance for the research community. (60)

  Use references to support your argument for why your project is relevant.

Post-project paper – Formal aspects

- Each paragraph has a central idea. Few errors in grammar and punctuation. Mostly academic language. Mostly correct use of terminology. (20)

- Paragraphs have a good flow. Virtually no errors in grammar and punctuation. Academic language throughout. Precise use of terminology. <u>Correctly formatted references.</u> (40)

  complete entries, consistent entries, correct use of citations

This session

- How to find and cite research articles in the area of NLP

- How to assess articles in terms of their quality

- How to relate your work to the research literature