Natural Language Processing

Working with research literature

Marco Kuhlmann
Department of Computer and Information Science

LINKÖPING UNIVERSITY

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 learning outcome is tested in the project.

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 ACL Anthology (preferably) or on arXiv.

- Add the BibTeX entry at the end of the `custom.bib` file inside the Overleaf 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: OpenReview

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 arXiv.

- 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

- Computational Linguistics
- Natural Language Engineering

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, what are the advantages and disadvantages with this way of measuring impact?

How to relate your work to the research literature

How to relate your work to the literature

- 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.

- Implement and evaluate a method described in the literature.

  Replication study

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.

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