Chapter G:I

I. Scientific Toolbox

- Literature Research
- Oral Presentations
- Scientific Writing
Scientific Writing
Content of a paper

- Most of the above hints on talks still hold
  - Science is storytelling
    Seminar: No scientific break-through expected, rather summarize and discuss.
  - Science needs to be understood

- Papers are more complete
  - Tell the whole story, avoid gaps in argumentation
  - But: Include only relevant content
    Don't expect too much prior knowledge.
  - But: No details on knowledge that can be presupposed

- Papers should be sound
  - Need to be precise more than in talks
  - Use logical arguments, from broad context to deep details
  - Formalize concepts if needed / helpful
Scientific Writing
Structure of a paper

- High-level structure
  - Title, author information, abstract
  - Introduction
  - Usually 2–5 sections
    - Related work, approach, experiments, etc.
  - Conclusion
  - References
    - ... and sometimes appendices

- Section structure
  - Often numbered subsections (2.1, 2.2, ...)
  - If any, subsubsections unnumbered
  - Always have text introducing (sub)sections

- Section headings
  - Stick to the standard: “Introduction” is first, “Conclusion” is last, etc.
  - Short misleading headings worse than longer clear ones
Abstract

- A concise high-level summary of the paper
- Usually 5–10 sentences

- One “approach”
  - Motivation and context (1 sentence)
  - Problem and why not solved (1–2 sentences)
  - Question addressed in the paper (1 sentence)
  - Approach with some details (2–3 sentences)
  - Evaluation, results, conclusion (1–3 sentences)

- Or in other words
  - What is the problem? Why is it a problem?
  - What is the solution? Why is it a solution to the problem?
Scientific Writing

Sections

- Introduction
  - The abstract in more detail
  - Tell the whole story, from context to conclusion
  - Still high-level
  - Understandable for computer scientists

- Content sections
  - The introduction in more detail
  - Elaborate on related work, concepts, models, data, approaches, experiments, and results
  - More technical, for researchers from the area

- Conclusion
  - The introduction in less detail
  - Summarize story in retrospective, give outlook
  - Semi-technical
Scientific Writing
Style

- Write clearly, unambiguously, and concise
- Don’t make things complex (common misunderstanding)

- Some guidelines
  - Use impersonal or “we” form
  - Avoid pronouns with unclear references
  - Use explicit discourse markers, such as “because”
  - Blurring is non-scientific, such as “It could be . . . ”
  - English sentences are short, one statement per sentence
  - Again: Avoid grammar and spelling errors
  - Highly recommended: Writing for Computer Science by Justin Zobel
Scientific Writing
Tables, figures, terms, and footnotes

- Tables and figures
  - In papers, just number increasingly
    Figure 1, 2, … Table 1, 2, … (NOT: Figure 2.1, 2.2, …)
  - Tables: Horizontal lines suffice
  - No included font larger than article font
  - Explain in text and in caption

- Technical terms
  - Introduce where needed, don’t overformalize
  - Use well-defined terms, AIA & AUA
    Always introduce acronyms & avoid unnecessary acronyms.
  - Don’t use synonyms for terms
    Reader is misled to check whether intentional differences exist.

- Footnotes
  - Only for secondary information
  - Reduced readability, should be an exception
  - Don’t cite literature using footnotes
Citation

- In-text reference to a bibliographic source
- Different styles
  Acronyms [ACW17], ACL style (Ajjour et al., 2017), numbers [42], . . .

What to cite

- Any reuse, paraphrase, summary, or translation of content
- Rule of thumb: Always clarify what is from you and what from others
  Also have to cite yourself if you use your own sources.
- Better one citation too much than too few

How to cite

- Direct reuse. Put in quotes (shorten with [. . .]), give source
  Unit segmentation is “[. . .] the splitting of a text into argumentative segments” [ACW17].
- Other citations. Give source close-by
  As Ajjour et al. point out, segmentation is the first task of an argument mining pipeline [ACW17].
- Large text portions. Give source once in the beginning
  In the following paragraph, we summarize the segmentation approach of Ajjour et al. [ACW17].
Scientific Writing

References

- Bibliographical information at the end of the paper
- Exactly those references cited in the text
- Information should be complete and homogenous

- Needed meta-information
  - All literature. Author, year, title
  - Conferences/Workshops. Proceedings, pages
  - Journals. Journal name, issue, number, pages
  - Books. Edition if any, publisher
  - Only online. Give URL with access date

- Bibtex
  - LaTeX handles references automatically using bibtex

See part on organizing literature above.
To sell another’s ideas or expressions as one’s own
See en.wikipedia.org/wiki/Plagiarism

On purpose or due to lack of giving sources

Plagiarism is not(!) a trivial offense
In some countries considered as crime.

Proper citing avoids all plagiarism issues

Consequences
– Major cases lead to the denial of being published, graded, or worse
– Minor cases can still negatively affect a grade or review outcomes

Webis network [www.webis.de]
– We do research on text reuse detection
– See publications, shared tasks, or the tool picapica [www.picapica.org]