ACADEMIC SKILLS IN COMPUTER SCIENCE

Lecture 7: Writing Research Reports (2)

Markus Krötzsch
Knowledge-Based Systems

TU Dresden, 14th May 2019

Review: Writing

Always have a plan!

Start early!

Accept that nothing is perfect and embrace change!

Keep the reader in mind!

Use a clear structure with no surprises!

Goals for today

Learning goals of this lecture:

(1) Understand how to cite and quote properly

(2) Learn about appropriate writing style
especially regarding peculiarities of writing research reports

(3) Become aware of common issues in English writing

Citations: What can/should be cited?

What can be cited?

- Everything that can be reasonably referred to and accessed by (at least some) other readers
- Including web pages, theses, data sets (ideally with DOI), TV shows, public talks (ideally recorded at some URL), etc.
- With caution: unpublished manuscripts (clarify how this can be accessed; or use only to credit author without relying on its content)
- Not citable: personal discussions

What should be cited?

- Trustworthy research of high quality (use same guidelines as for judging research, e.g., journal article > workshop paper)
- Original sources are better than later summaries (exception: for background reading, textbooks are better than historic originals)
- Archival publications are better than web pages

Citations: stylistic details

When to use "et al."?

- "et al." abbreviates "et alia" ("and others")
- used to abbreviate lists of at least three (in some style guides: four) authors when referring to them in the main text
- almost never justified in the full references at the end of the paper

What about "ibid," "op. cit.," and "cit. loc."?

- Shortcuts invented for disciplines that frequently cite and discuss precise locations (text passages) of other works
- Extremely uncommon in computer science – don't use this without strong reason

Quotations

If you want to copy text in verbatim, citing is not enough!

Quotations

- are enclosed in quotation marks ("...") and clearly associated with a cited source (immediately before or after)
- are exactly as in the original
- with changed parts being inserted with brackets (used to clarify but not distort the meaning!)

> **Example:** As noted by Smith, "solving this problem [of efficiently sorting an array] is of vital importance in computer science, engineering, [...], and biology" [42].

Quotations should be used sparsely and are usually not needed in computer science.

Not citing verbatim quotes is plagiarism. Superficial paraphrasing is no substitute for proper citing (and is plagiarism too!).

Good Writing

Review and Outlook

We already covered some important points:

- Follow a narrative structure that is logical and appropriate
- Use structure consciously: each section, paragraph, and sentence has a purpose

Good writing has many further aspects, which mainly fall into two categories:

- Write for a reader: serve the reader (i.e., do not try to impress or intimidate)
- Respect existing conventions: standards ease readability and should not be abandoned due to ignorance or ego (grammar, spelling, layout, typography, language conventions, ...)

> Any rule can be broken, but decide carefully if an exception is really justified.

Efficient, direct language

> Say exactly what needs to be said, in the simplest possible form.

- Use the simplest words that capture your intent exactly

  > **Example:** Instead of time-efficient, say fast. Instead of numerous, say many.

- Use one word for one concept, especially for technical concepts
- Prefer shorter sentences (but avoid staccato sequences of very short sentences)
- Avoid redundant or meaningless expressions

  > **Example:** Phrases like potential risk and end result are redundant. Phrases like smart software or big data are meaningless (unless defined previously).

  Note: Explaining an important idea twice for clarity is useful repetition. Redundancy is repetition that does not add anything.

- Ask if each sentence is really needed
  ⤳ cut away the deadwood

Example: Deadwood

> "The volume of information has been rapidly increasing in the past few decades. While computer technology has played a significant role in encouraging the information growth, the latter has also had a great impact on the evolution of computer technology in processing data throughout the years. Historically, many different kinds of databases have been developed to handle information, including the early hierarchical and network models, the relational model, as well as the latest object-oriented and deductive databases. However, no matter how much these databases have improved, they still have their deficiencies. Much information is in textual format. This unstructured style of data, in contrast to the old structured record format data, cannot be managed properly by the traditional database models. Furthermore, since so much information is available, storage and indexing are not the only problems. We need to ensure that relevant information can be obtained upon querying the database."

– Example of waffle by Zobel, Writing for Computer Science, 2015

Useless phrases and padding

> Avoid wordy or redundant phrases when there is a short alternative.

Examples:

- "for the purpose of" ↦ "for"
- "first of all" ↦ "first"
- "cancel out" ↦ "cancel"
- "add together" ↦ "add"
- "it is frequently the case that" ↦ "often"
- Consider deleting qualifiers like "very", "simply", "highly", "truly", "somewhat", "quiet", "certainly", … (ask: What exactly does this add in meaning?)
- Use strong verbs: "result in performance improvements" ↦ "improve performance" or "perform an evaluation" ↦ "evaluate"

The right tone

Research reports should sound like research reports:

- No buzzwords, clichés, slang or colloquialisms
- No popular science or dumbed-down language
- No artistic or intellectual excesses (don't show off)

Rules might be broken for good reasons (like all rules)

- The boundary between "buzzword" and "technological vision" or "emerging area" can be blurry

  > **Example:** "Machine learning" could be considered a buzzword that wrongly suggests that computers behave like humans, but it has become an established term for a wide area of approaches and methods.

- Metaphors and analogies can help readers to understand complex abstract ideas. This is not "dumbing down" language.

  > **Example:** Calling certain acyclic graphs "trees" helps human intuition.

The right words

Select words that are specific and familiar.

- Short>long: many>numerous, new>novel, use>utilise, begin>initiate, etc.
- Familiar>exotic: quick>expeditious>apace, easy>elementary, hence>thence
- However, do not simplify away the meaning

  Example: "Our algorithm uses features of the input data to get better results" is vague and ambiguous. Better: "Our algorithm analyses the node-degree distribution to reduce memory consumption."

- Be careful with words you got from a dictionary:
  – Re-translate unknown words to your language to check if they have additional, unsuitable meanings
  – Use web search to find examples of how a word is used
  – Formulate your text in English rather than trying to translate from another language word by word

Careful claims

Researchers use cautious, precise language for claims and conclusions.

Wrong: "Our experiments prove that our algorithm is better than existing algorithms."

Right: "In our experiments on several real-world data sets, the algorithm ran about 30% – 65% faster than existing solutions. This suggests that our approach can compete with the state of the art in practical cases."

Typical problems:
- presenting experimental evidence as definitive facts that can never change
- making claims too general (over-simplification makes claims wrong)
- using superlatives (best, perfect, ideal, . . .), especially when referring to own work

Cautious phrasing does not weaken results but rather improves their credibility

Motivational claims

Claims that are used for motivation (and not as the actual contribution) also need to be supported.

Example: "In the near future, most travelers will start their multi-modal journey through a seamlessly connected smart city with intelligent mobility services at home." [RGWK2018] and [RGWK2017]

Especially in introductions, one must also resist the urge to exaggerate.

Example: "Cognitive modeling is a method that has taken psychological and cognitive research by storm." [RBR2018]

If a claim is so obvious that no support seems needed, or so general that it is impossible to find a reference to support it, one should ask if it makes sense to state it.

Example: "Massive amounts of data, stored in disparate data sources, have to be integrated and matched to support data analyses that can be highly beneficial to businesses, governments, and academia." [KGV2018]

Careful claims: weasel words

Unfounded claims do not get better through empty phrases that pretend support.

Avoid weasel words: (a.k.a. "references to anonymous authority")
- "it is often claimed that" (where? by whom?)
- "experts agree that" (who?)
- "a growing body of evidence suggests" (citations?)
- "it is widely acknowledged" (by whom?) or "it is known that" (from where?)
- "often", "many", "typically", . . . (may need justification)
- "it is clear/obvious" (is it really?)
- . . .

Guideline: cut out false suggestions of support; provide real support (citations, justification, results); weaken claims for which no support can be found

Clearly and obviously

> Do not wrongly suggest that insights or conclusions are obvious.

- Typical phrases: "obvious", "clear", "easy to see", "immediate", "of course"
- If it really is, then the reader will usually not need to be told
- Wrong uses: to safe the effort of explaining properly (laziness), to hide lack of support (weasel), as empty phrase added to every immediate claim (in a reflex)
- Acceptable use: to indicate the (relative) difficulty of drawing some conclusion that is not explained in detail

> **Example:** "The first part of this claim is easy to see using arguments that are similar to the ones in the proof of Lemma 42."

Guideline: try to avoid such phrases – doing so usually improves your writing and may expose hidden problems

---

Active and passive / we and I

> Active voice is mostly preferable. "We" is common even for a single author.

Active voice is more direct and easier to read
- Bad: "Experiments are conducted to validate the approach."
- Good: "We conduct experiments to validate our approach."
- Active is not always the obvious choice though: "In recent years, large data centres have been built to host modern computing clusters."

Referring to the authors:
- It is most common to use "we" for the authors, even if there is only one
- "I" is rare; used to emphasise personal views (e.g., acknowledgements in a thesis)
- Third person often sounds artificial, but can be used as well
  - "A preliminary study was conducted by some of the authors [42]." might be better than "... by some of us [42]."
  - Acknowledgements in papers often use third person, e.g., "The author was supported by grant number ABC."

Simplifying sentences further

Discussed so far: cut deadwood, simplify words, avoid redundancy & empty phrases, prefer active over passive

Other methods:
- Split long sentences that are several sentences joined by "and" (or similar)
- Consider using bulleted or numbered lists for sequences of important points
- Unravel complicated nested sentence constructions
- Avoid parenthetical remarks that distract from the main point
  - often you can just drop them
  - alternatively give the information in main text (earlier or later)
  - possibly use a footnote (but sparsely)

Example

> "Sorting algorithms, which are essential for many processing tasks – in particular in data analysis –, can benefit from the use of tries (generalised prefix trees) and other efficient data structures that enable faster traversal of records, and can thus be improved in performance, regarding both time and memory."

This convoluted sentence can be split easily:

> "Sorting algorithms are essential for many processing tasks, especially in data analysis. They can benefit from efficient data structures, such as tries, that enable faster traversal of records. Using such data structures leads to sorting algorithms that are faster and need less memory."

Note: Splitting sentences is just a first step towards good phrasing. One should now ask if these sentences convey the key points (e.g., why do we single out data analysis and tries?), and which supporting citations to include for the claims. Finally, one should try to make them sound less clumsy.

Adding structure with paragraphs

Every paragraph should have a single, clear purpose.

Guidelines:

- Think of one headline per paragraph
- The first sentence introduces the main topic; the others elaborate or exemplify
- The last sentence is more prominent and should be picked with care
- Paragraphs should be linked to form a logical narrative
- Nevertheless, consecutive paragraphs are more strongly separated than sentences within one paragraph

Example: Anaphors (terms that depend on prior context, such as "it" and "he") may need to have their antecedents (the terms they refer to, such as "the algorithm" or "Alan Turing") re-introduced to be clear.

Clarity

It requires effort to write text that transports a clear, unambiguous message.

Readers often misunderstand your writing in unexpected ways

- When you turn your ideas into text, you will always recognise the intended meaning in the words
- Others have only the written words, and might interpret them very differently
- Even accomplished writers cannot foresee all misunderstandings that may arise
- In technical writing, one cannot rely on common sense to resolve vague language

Try to understand the perspective of your readers; ask friends or colleagues for feedback

Neutral and non-violent language

Mind cultural sensitivities; avoid being obscure or offensive to some readers.

- No profanity, of course
- No sexism, racism, or other offensive attitude should be voiced or implied
  It can be hard to guess what offends some people. Use common sense. Some authors even suggest the ungrammatical use of the pronoun "their" to avoid mentioning gender in examples.
- Some cultural specificities, e.g., in examples, might be hard to understand globally

Example: When submitted to an international journal, a manuscript about a Chinese online platform needs to add additional explanations to make screenshots understandable to English readers.

- Apparently harmless details might cause problems in other cultures

Example: The W3C OWL working group once was asked to replace the widely used "Wine ontology" with an example that does not mention alcohol. (This request was not followed at the time.)

- Works of others are to be discussed respectfully, even if critical points are raised

English

The English language

English writing brings several particular complications

- There is no official "correct" English, but many styles of writing
- Pronunciation is largely unrelated to spelling, but does affect grammar

> **Example:** One can write "a historical fact" or "an historical fact", depending on one's preferred pronunciation (but there is no such choice in "a history of formal languages"). Authors in logic programming encounter this a lot when talking about "a(n) Herbrand base".

- Many English texts are written by foreigners with varying degrees of proficiency

Nevertheless, language issues are often easy to resolve and it is a sign of carelessness to submit texts that are full of trivial spelling and grammar errors

AE vs. BE

The predominant versions of English are British English (BE) and American English (AE), which mostly lead to different spelling but also in some vocabulary:

BE	AE
neighbour, colour	neighbor, color
centre	center
sceptical	skeptical
fulfil, traveller	fulfill, traveler
analyse, realisation	analyze, realization[1]
program (software), programme (plan/schedule)	program (always)
football, lift, chips, crisps	soccer, elevator, fries, chips

Both variants are common in research papers. Select one and stick with it.

(and use a spellchecker while you write; do another full pass before you submit)

[1] Some British style guides also recommend the -ize forms.

Allow

The correct use of "allow" is a mystery to many foreign speakers.

There are three correct forms:

(1) "To allow somebody to do something" – to enable somebody to do something

> **Example:** "Lemma 4 allows us to prove our main result."

(2) "To allow something'" – to permit something

> **Example:** "Our implementation allows further optimisations."

(3) "To allow for something" – to take something into account

> **Example:** "Our algorithm allows for several input formats."

However, the form "allow to do" does not exist in English

> **Example:** "Using faster algorithms allows to speed up software" is wrong, but one can say "Using faster algorithms allows speeding up software."

Summary

References must be citable and should be trustworthy and of high quality

Literal quotation is rare; unquoted paraphrasing is plagiarism

Language in research reports should be plain, specific, and factual

There are many pitfalls in English writing that even regular speakers are not aware of

> **What's next?**
> - Typesetting with LaTeX
> - Persuasive writing and argumentation
> - Oral presentations

References

[RGWK2018] Rocznik, Goffart, Wiesche, Krcmar: An Implementation and Evaluation of User-Centered Requirements for Smart In-house Mobility Services. Proc. of the 41st German Conference on AI (KI'18), Springer, 2018.

[RGWK2017] Rocznik, Goffart, Wiesche, Krcmar: Towards Identifying User-Centered Requirements for Smart In-House Mobility Services. Künstliche Intelligenz, Volume 31, Springer, 2017.

[RBR2018] Riesterer, Brand, Ragni: The Predictive Power of Heuristic Portfolios in Human Syllogistic Reasoning. Proc. of the 41st German Conference on AI (KI'18), Springer, 2018.

[KGV2918] Karapiperis, Gkoulalas-Divanis, Verykios: Summarization Algorithms for Record Linkage. Proc. of the 21th International Conference on Extending Database Technology (EDBT'18). OpenProceedings.org, 2018.