On the Structuring of LaTeX Projects

17 November 2025

Wouter ten Brinke, Bart Griepsma, *Aleksandra Ignatovič*, *Nhat*, Vadim Zaytsev

UNIVERSITY OF TWENTE.

Project Structuring

- LaTeX
- taken as a software project
- a software language cocktail
- build: pdflatex/lualatex, bibtex/biblatex
- entry: main.tex, paper.tex, …
- hierarchy, abstraction, modularity, encapsulation, coupling, cohesion…

On the Structuring of LaTeX Projects

Wouter ten Brinke[1], Bart Griepsma[1], Aleksandra Ignatovi [1], Nhat[2] and Vadim Zaytsev[2]

[1] Technical Computer Science, University of Twente, Enschede, The Netherlands
[2] Formal Methods & Tools (FMT), University of Twente, Enschede, The Netherlands

Abstract

In academia, LaTeX is a powerful typesetting system widely used for producing scienti c documents such as research papers, theses and reports. It allows authors signi cant freedom and control over the structure and styling of their documents. However, this exibility often leads to inconsistent internal project structures and coding styles, which can hinder maintainability and collaboration among co-authors.

In this paper, we investigate various existing traditions in structuring one's LaTeX projects. By analysing **29** academic users through interviews and surveys, we uncover prevalent practices and attitudes towards standardisation. Additionally, we mine **215** LaTeX repositories from GitHub to identify structural and stylistic patterns using feature extraction and clustering techniques. Finally, we introduce FlexiTeX, a system that allows users to maintain their preferred project structures while collaborating on shared content. FlexiTeX achieves this by parsing documents into an abstract tree representation and applying con gurable transformation rules. Our preliminary ndings suggest that while no universal standard exists, there is space for tool support in enhancing collaboration and maintainability in LaTeX projects.

1. Introduction

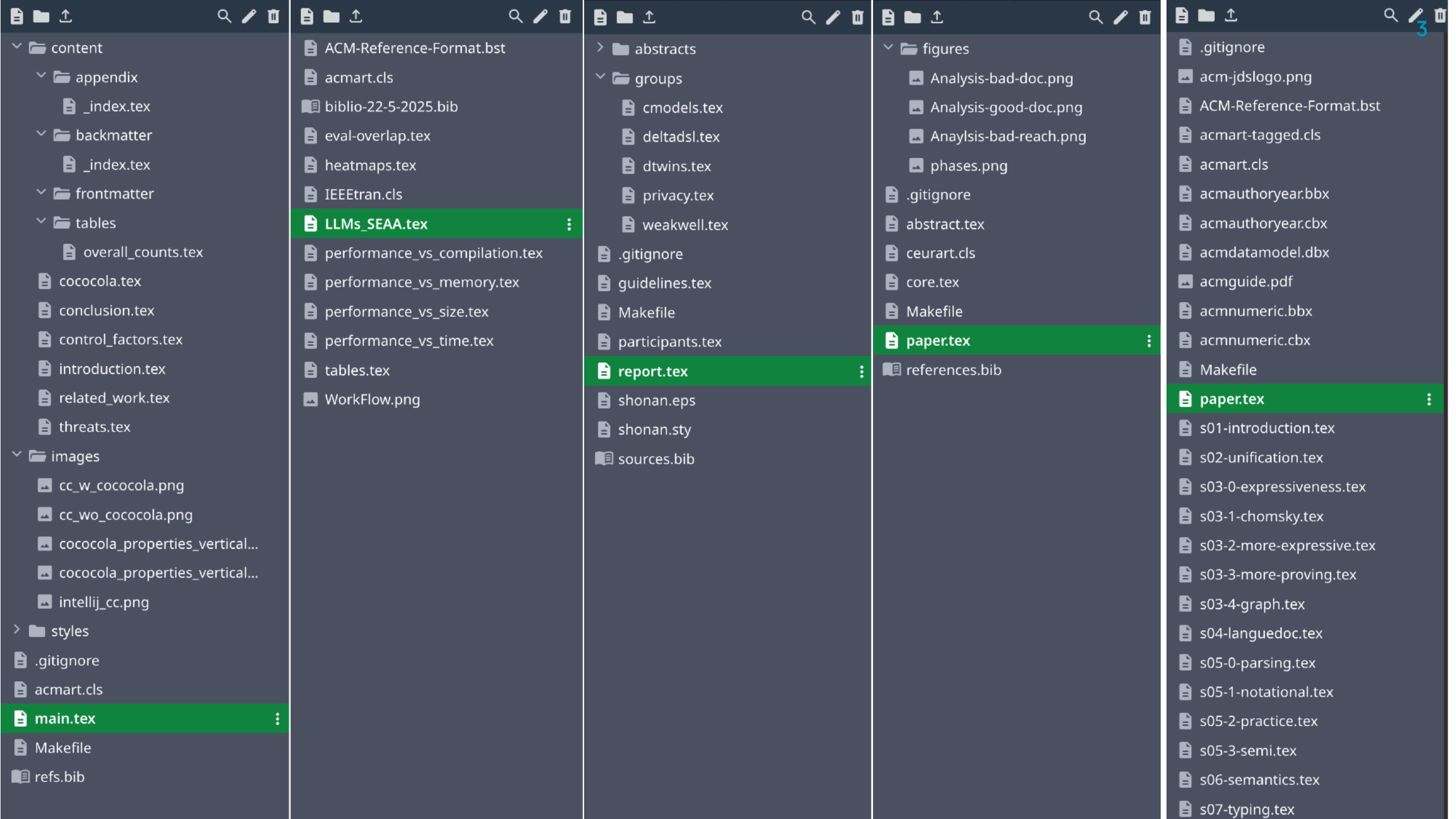
LaTeX [1] is a widely used typesetting system, particularly in academia, for producing high-quality scienti c documents. Its strengths lie in its ability to handle complex formatting, mathematical notations, as well as bibliographies. LaTeX allows authors signi cant freedom in how they structure and organise their projects, and does not enforce any standards for folder layout, le naming conventions, coding styles, etc. Publishers often make use of their own *document classes* which impose some constraints on de ning meta-information (authors' names, emails, title, subtitle, a liations) and using certain packages, as well as *bibliography styles* which dictates which elds of BibTeX entries are used and how. A very occasional journal might employ a submission system that also limits font usage or requires all content to t in one LaTeX le. Such unabashed exibility can lead to inconsistent practices, making it challenging for collaborators to work together e ectively, if they are used to drastically di erent folder structures or content clustering. Inconsistencies can also hinder maintainability, as authors may struggle in the future (when working on a resubmission, a camera ready version or an extended version of the same paper) to understand or modify documents that do not follow a clear and standardised format.

Despite its widespread use, there is currently no universally accepted standard for organising LaTeX projects. Authors often develop their own conventions for le structure, naming, and coding styles. These practices are often informal, ad hoc, and can vary widely across individuals and disciplines. This lack of standardisation leads to challenges in collaborative environments, where multiple authors may have di erent expectations and practices. In academia, such challenges are particularly pronounced, as scienti c documents often involve multiple contributors(as is often the case for research papers) and require long-term maintenance activities (common for books and PhD theses). Services such as *Overleaf* aid collaboration by supporting various build con gurations and providing templates, but they do not alleviate the issues one person's neatly curated setup is another's indecipherable labyrinth to navigate.

BENEVOL'25: Proceedings of the 24th Belgium-Netherlands Software Evolution Workshop, 17–18 November 2025, Enschede, The Netherlands

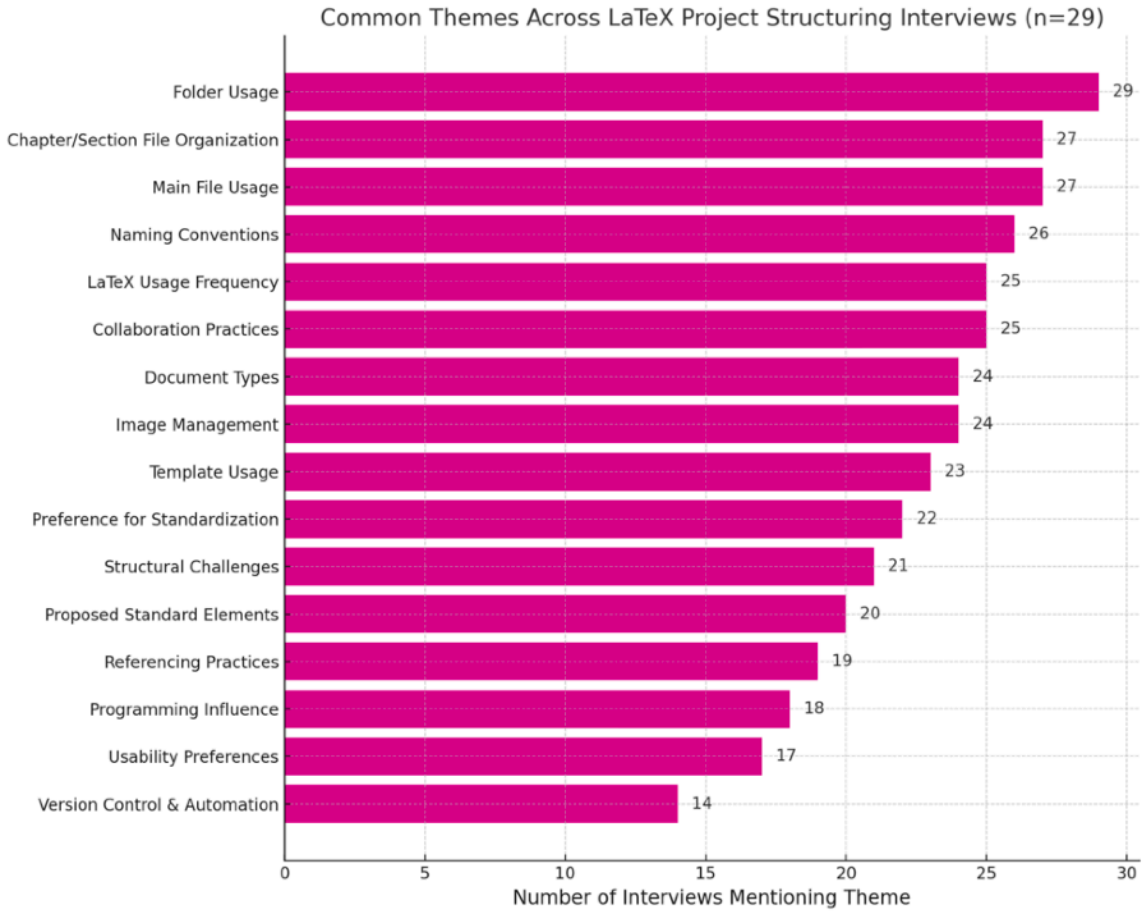
w.d.c.tenbrinke@student.utwente.nl (W. ten Brinke); b.griepsma@student.utwente.nl (B. Griepsma); a.ignatovic@student.utwente.nl (A. Ignatovi); research@nhat.run (Nhat); vadim@grammarware.net (V. Zaytsev)
https://grammarware.net/ (V. Zaytsev)
0009-0004-3110-9946 (Nhat); 0000-0001-7764-4224 (V. Zaytsev)

UNIVERSITY OF TWENTE.


content
appendix
_index.tex
backmatter
_index.tex
frontmatter
tables
overall_counts.tex
cococola.tex
conclusion.tex
control_factors.tex
introduction.tex
related_work.tex
threats.tex
images
cc_w_cococola.png
cc_wo_cococola.png
cococola_properties_vertical...
cococola_properties_vertical...
intellij_cc.png
styles
.gitignore
acmart.cls
main.tex
Makefile
refs.bib
ACM-Reference-Format.bst
acmart.cls
biblio-22-5-2025.bib
eval-overlap.tex
heatmaps.tex
IEEEtran.cls
LLMs_SEAA.tex
performance_vs_compilation.tex
performance_vs_memory.tex
performance_vs_size.tex
performance_vs_time.tex
tables.tex
WorkFlow.png
abstracts
groups
cmodels.tex
deltadsl.tex
dtwins.tex
privacy.tex
weakwell.tex
.gitignore
guidelines.tex
Makefile
participants.tex
report.tex
shonan.eps
shonan.sty
sources.bib
figures
Analysis-bad-doc.png
Analysis-good-doc.png
Anaylsis-bad-reach.png
phases.png
.gitignore
abstract.tex
ceurart.cls
core.tex
Makefile
paper.tex
references.bib
.gitignore
acm-jdslogo.png
ACM-Reference-Format.bst
acmart-tagged.cls
acmart.cls
acmauthoryear.bbx
acmauthoryear.cbx
acmdatamodel.dbx
acmguide.pdf
acmnumeric.bbx
acmnumeric.cbx
Makefile
paper.tex
s01-introduction.tex
s02-unification.tex
s03-0-expressiveness.tex
s03-1-chomsky.tex
s03-2-more-expressive.tex
s03-3-more-proving.tex
s03-4-graph.tex
s04-languedoc.tex
s05-0-parsing.tex
s05-1-notational.tex
s05-2-practice.tex
s05-3-semi.tex
s06-semantics.tex
s07-typing.tex

How do Academics Organise Projects?

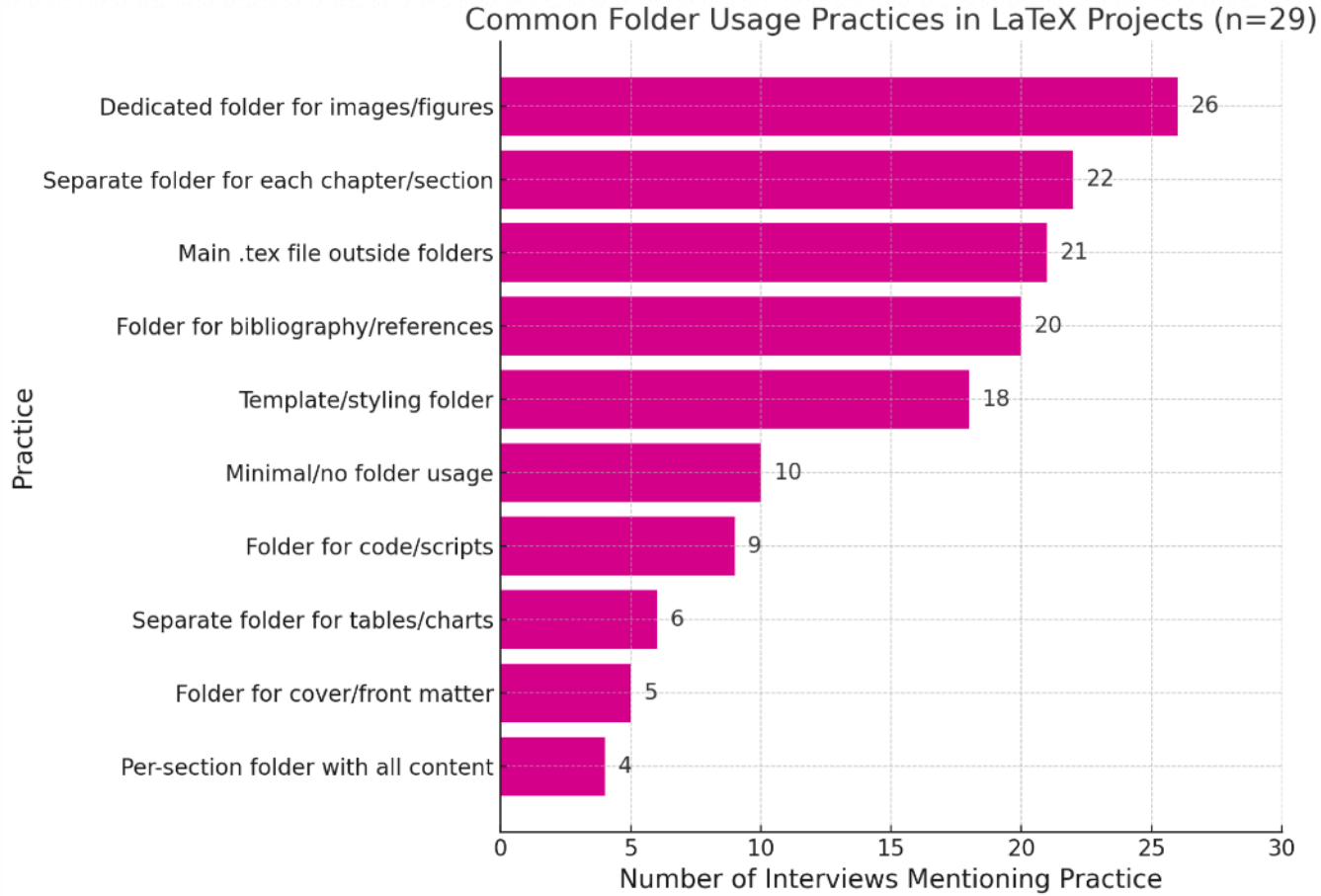
Discipline	Count
Computer Science	12
Mechanical or Electrical Engineering	8
Mathematics	5
Physics	3
Data Science	1
Total	**29**



http://purl.utwente.nl/essays/107820

How do Academics Organise Projects?

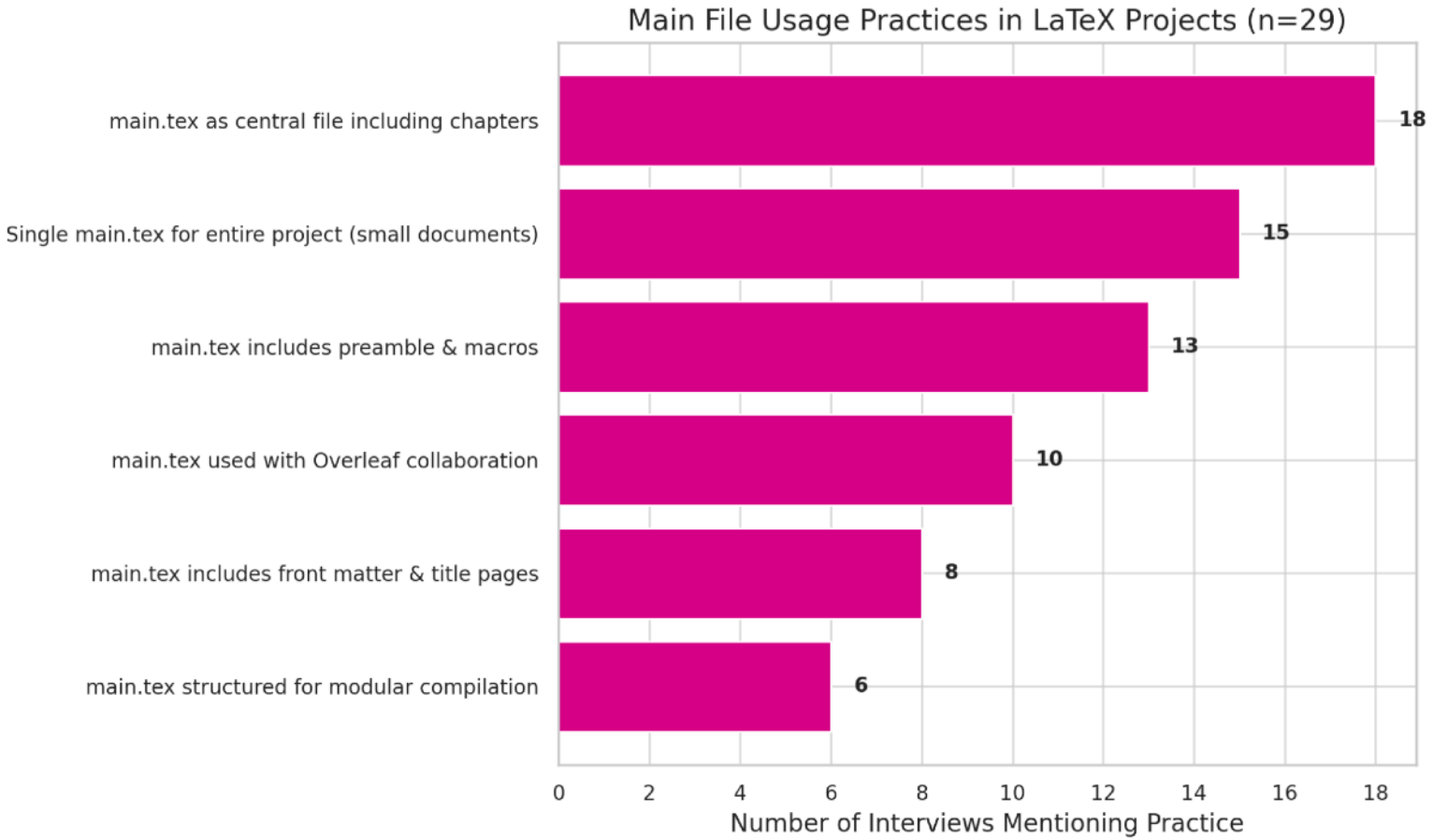
Discipline	Count
Computer Science	12
Mechanical or Electrical Engineering	8
Mathematics	5
Physics	3
Data Science	1
Total	**29**



http://purl.utwente.nl/essays/107820

How do Academics Organise Projects?

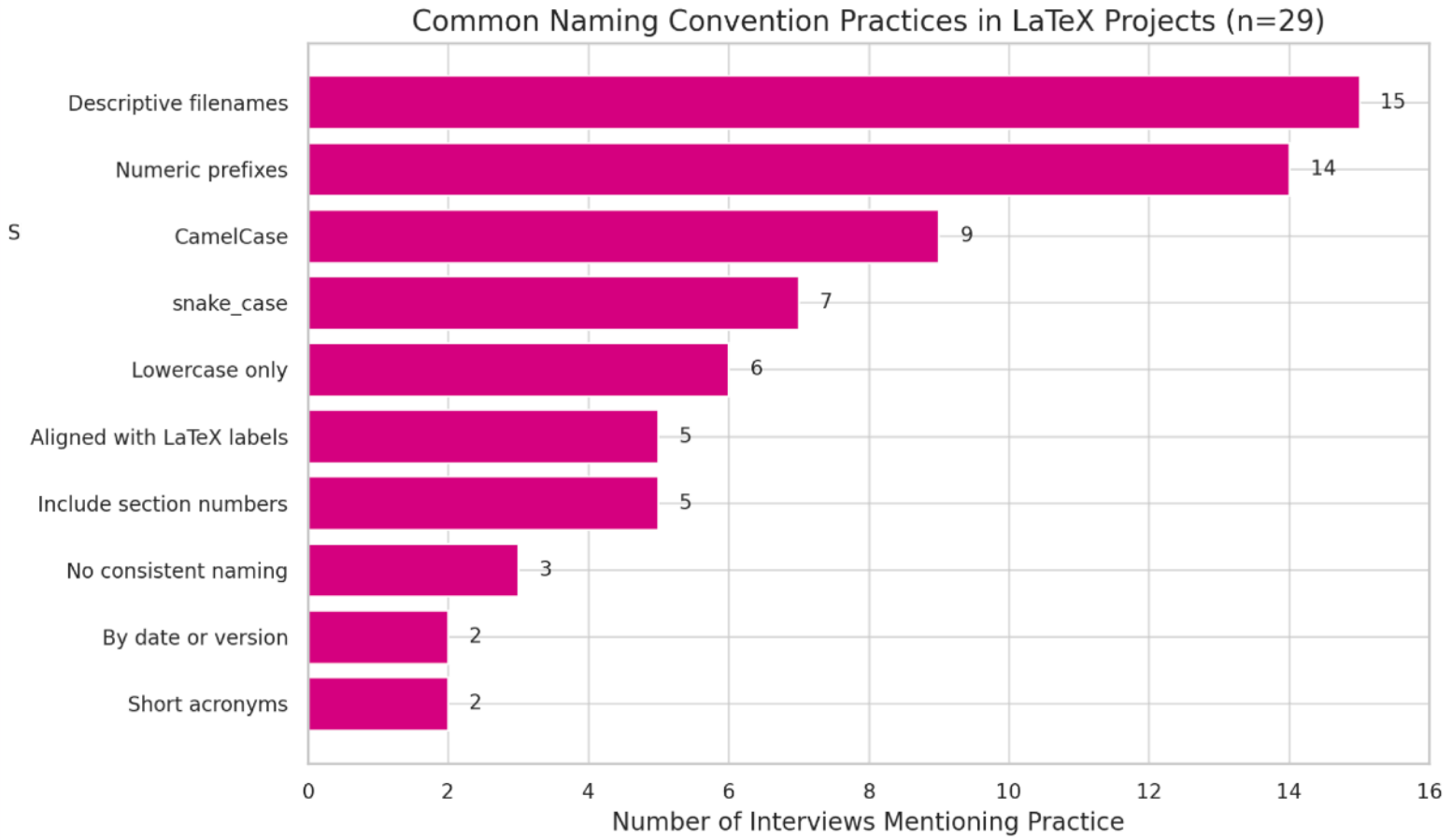
Discipline	Count
Computer Science	12
Mechanical or Electrical Engineering	8
Mathematics	5
Physics	3
Data Science	1
Total	**29**



http://purl.utwente.nl/essays/107820

How do Academics Organise Projects?

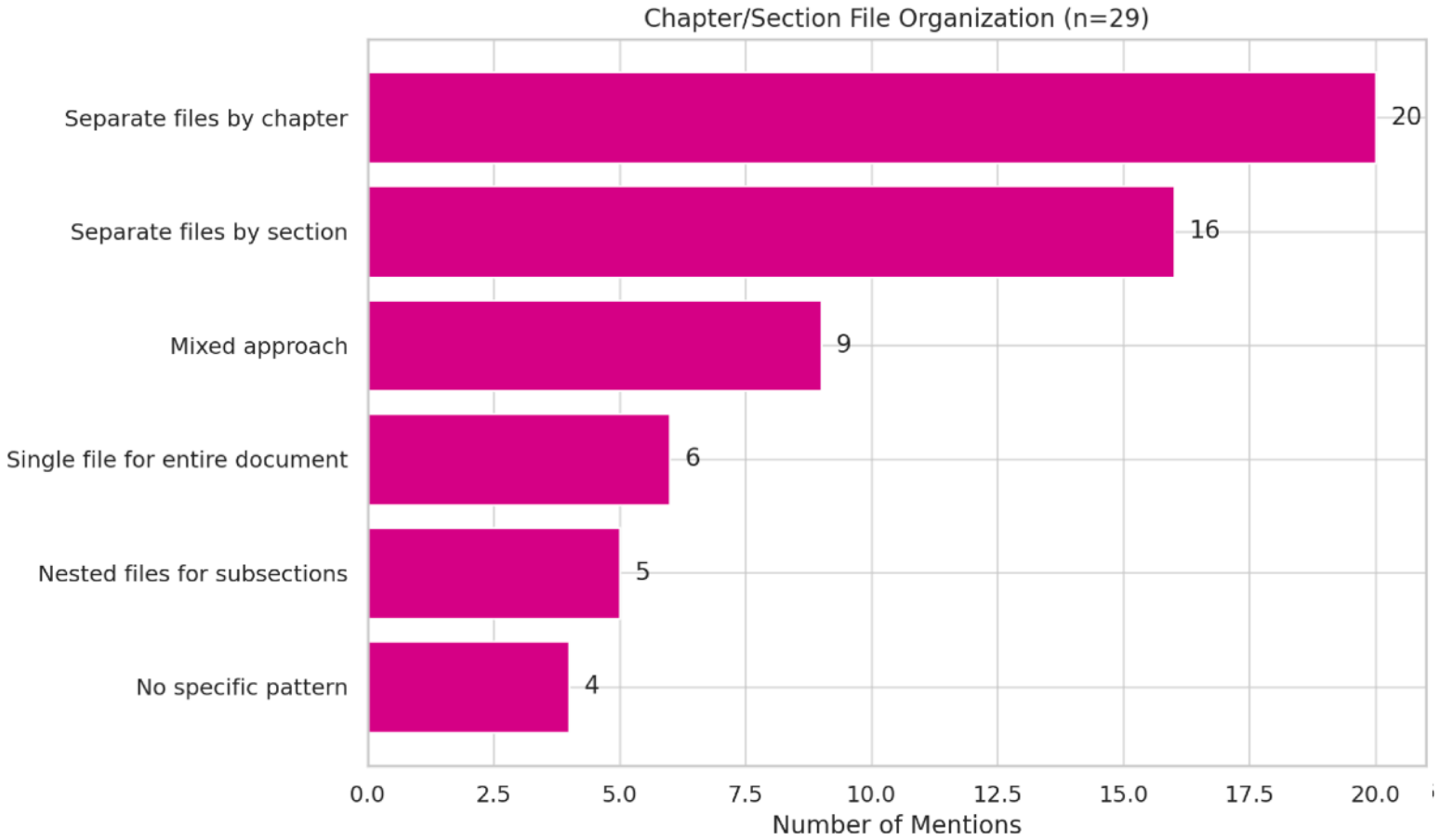
Discipline	Count
Computer Science	12
Mechanical or Electrical Engineering	8
Mathematics	5
Physics	3
Data Science	1
Total	**29**



http://purl.utwente.nl/essays/107820

How do Academics Organise Projects?

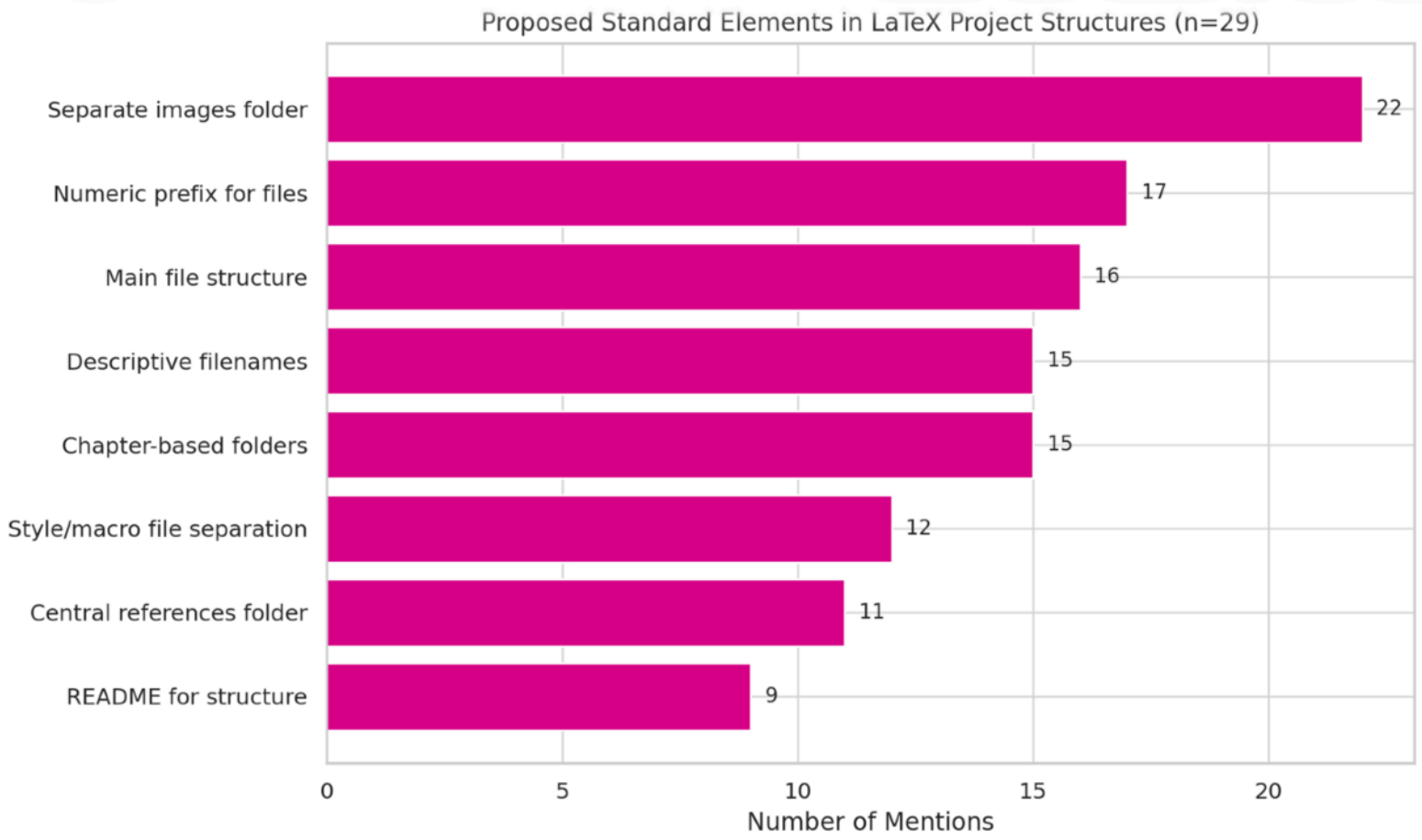
Discipline	Count
Computer Science	12
Mechanical or Electrical Engineering	8
Mathematics	5
Physics	3
Data Science	1
Total	**29**



http://purl.utwente.nl/essays/107820

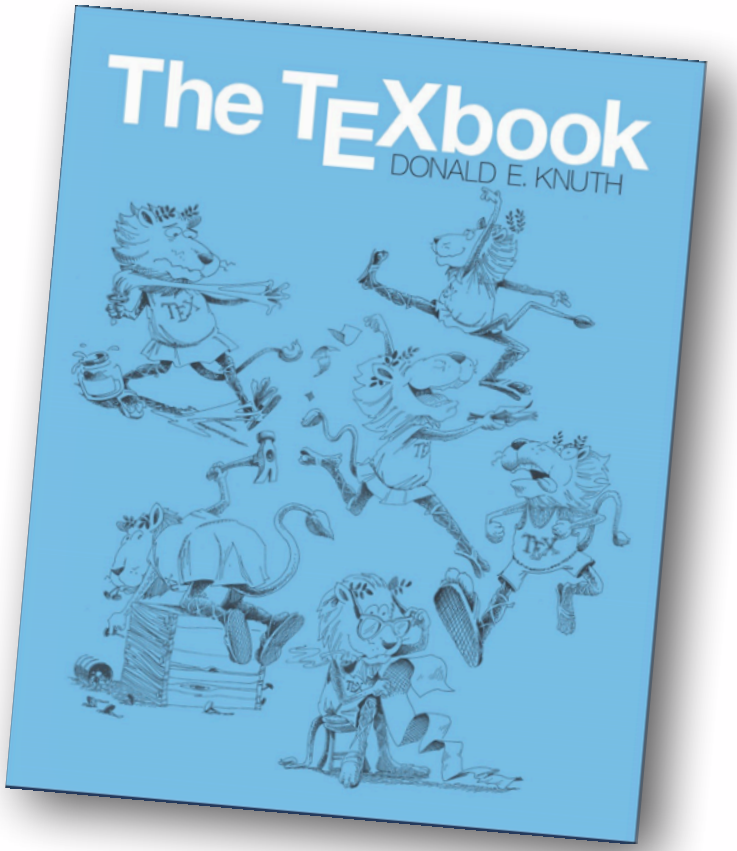
How do Academics Organise Projects?

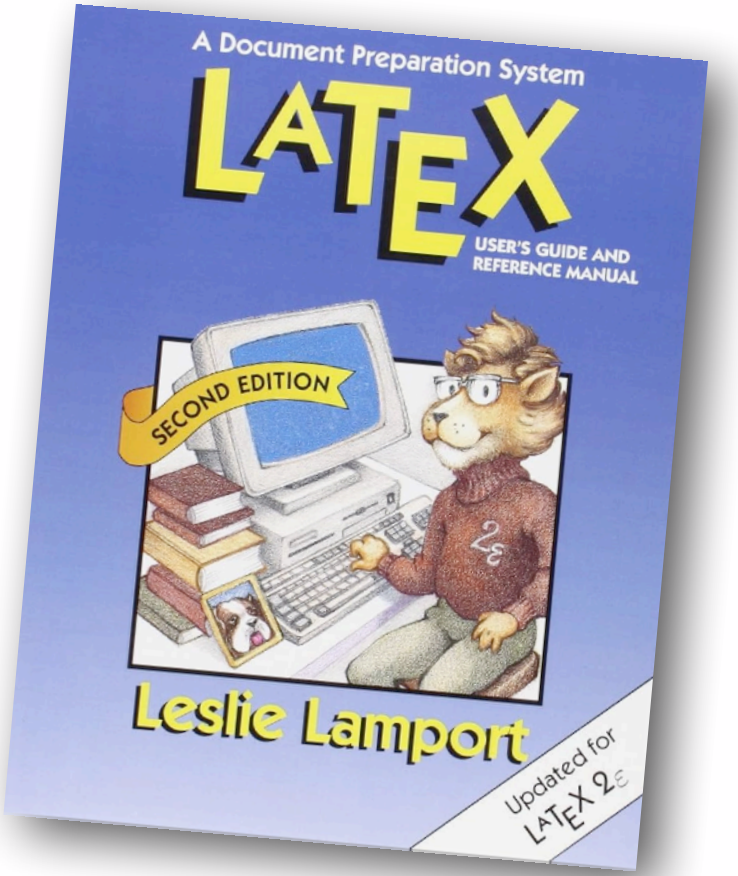
Discipline	Count
Computer Science	12
Mechanical or Electrical Engineering	8
Mathematics	5
Physics	3
Data Science	1
Total	**29**

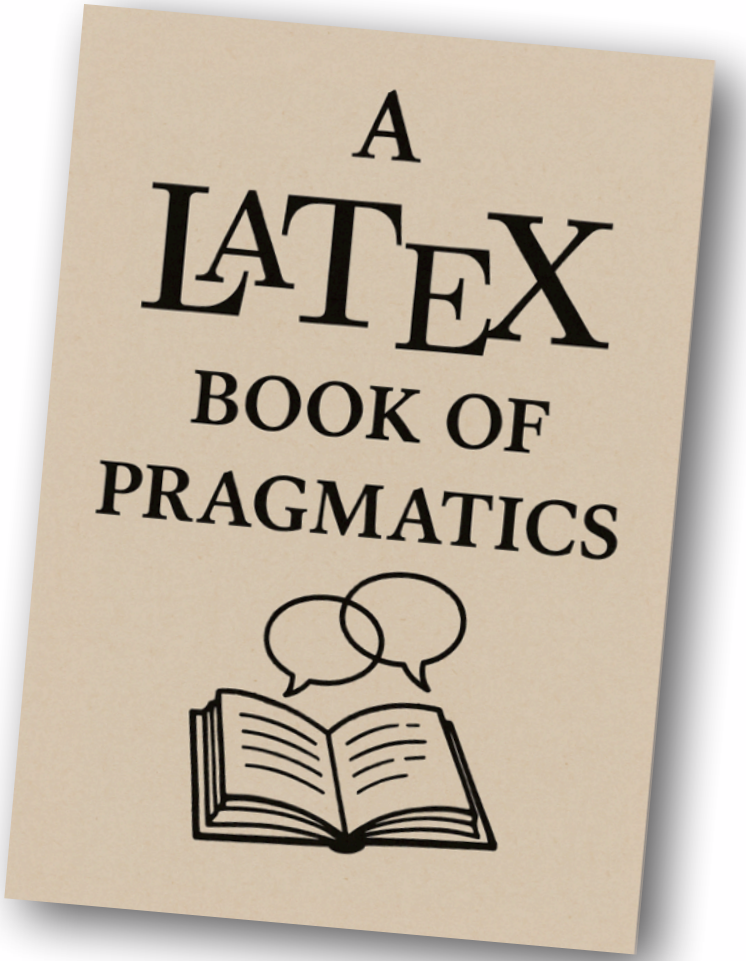


http://purl.utwente.nl/essays/107820

How about a Book?



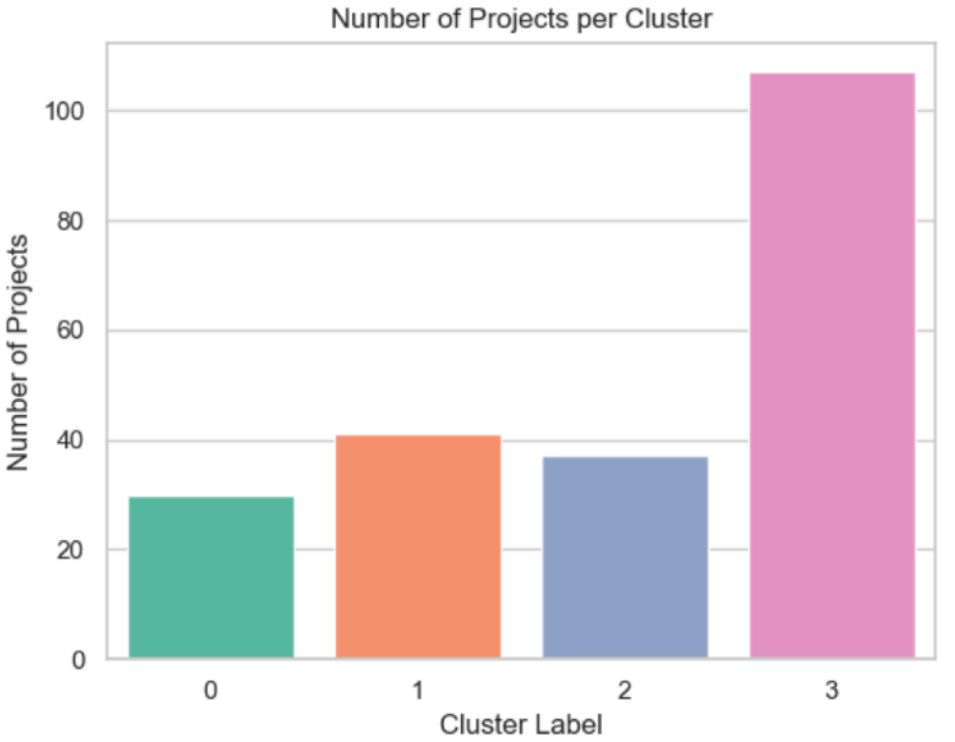




Usage in the Wild

- Cluster 0:
  - dozens of .tex, deep folders
- Cluster 1:
  - ~21 .tex, ~4 folders
  - inclusion, Makefiles
- Cluster 2:
  - 4- .tex, shallow, ~2kLOC, no \input
- Cluster 3:
  - ~19 .tex, ~3 folders, no Makefile

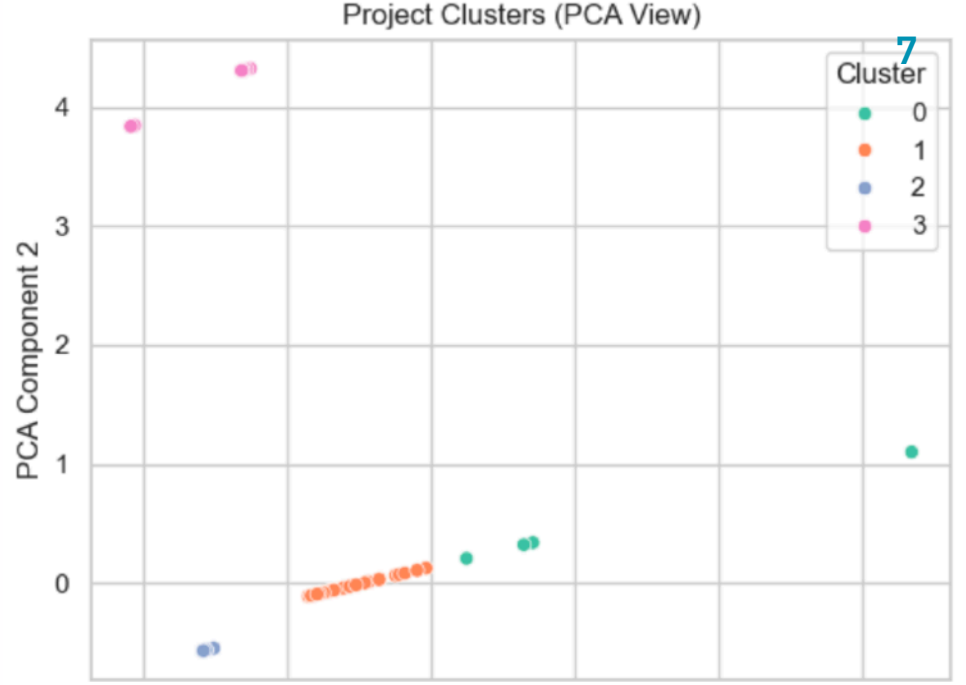
http://purl.utwente.nl/essays/107264
https://github.com/Bart0TW/LaTeX_academic_dataset

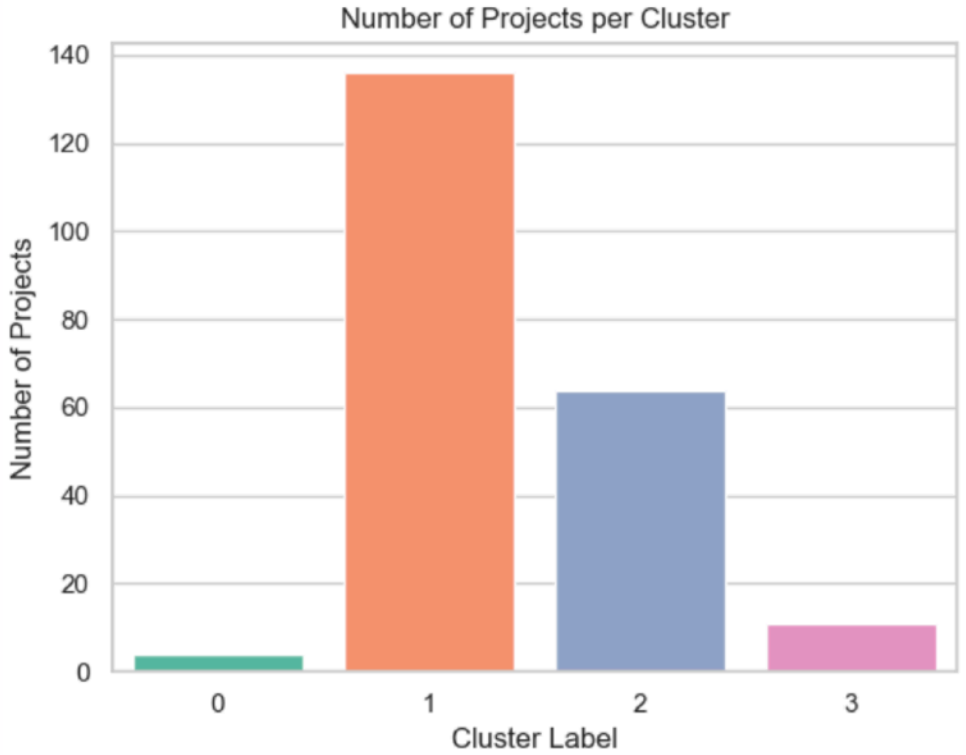


Usage in the Wild

- Cluster 0:
  - ~965 macros/commands!
- Cluster 1:
  - ~52 macros/commands
  - most prominent
- Cluster 2:
  - 3-4 commands w/o arguments
- Cluster 3:
  - ~21 macros/commands redefined

https://github.com/Bart0TW/LaTeX_academic_dataset

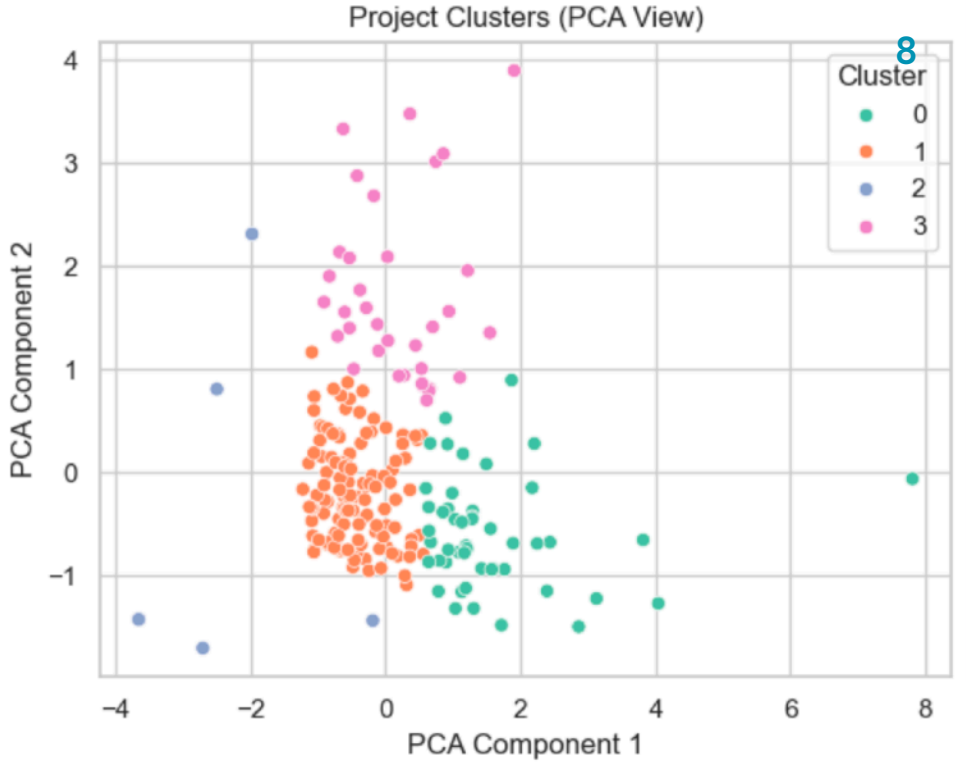


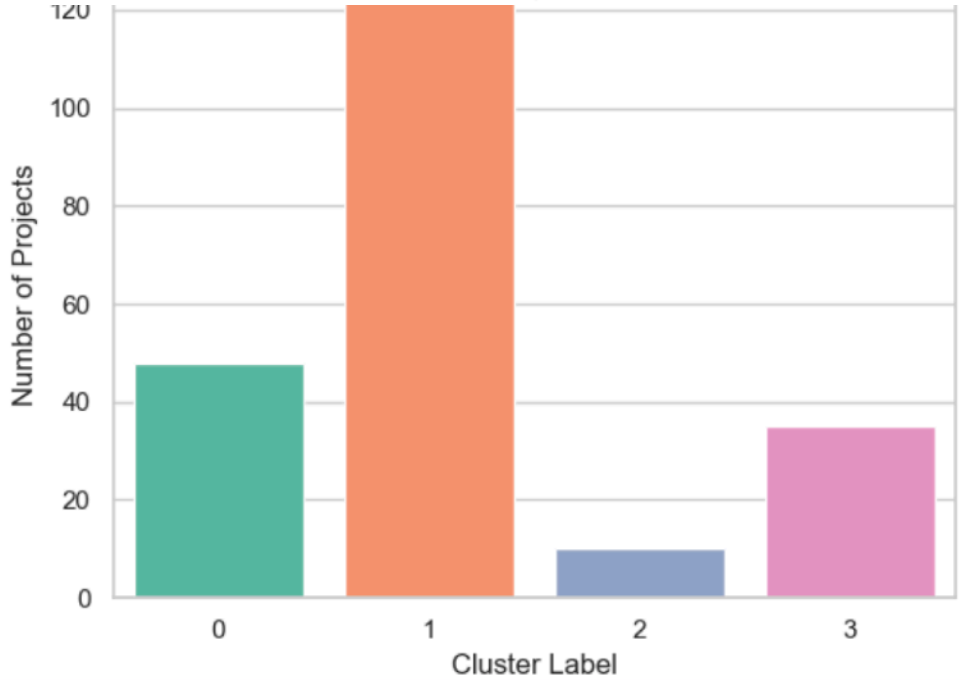


Usage in the Wild

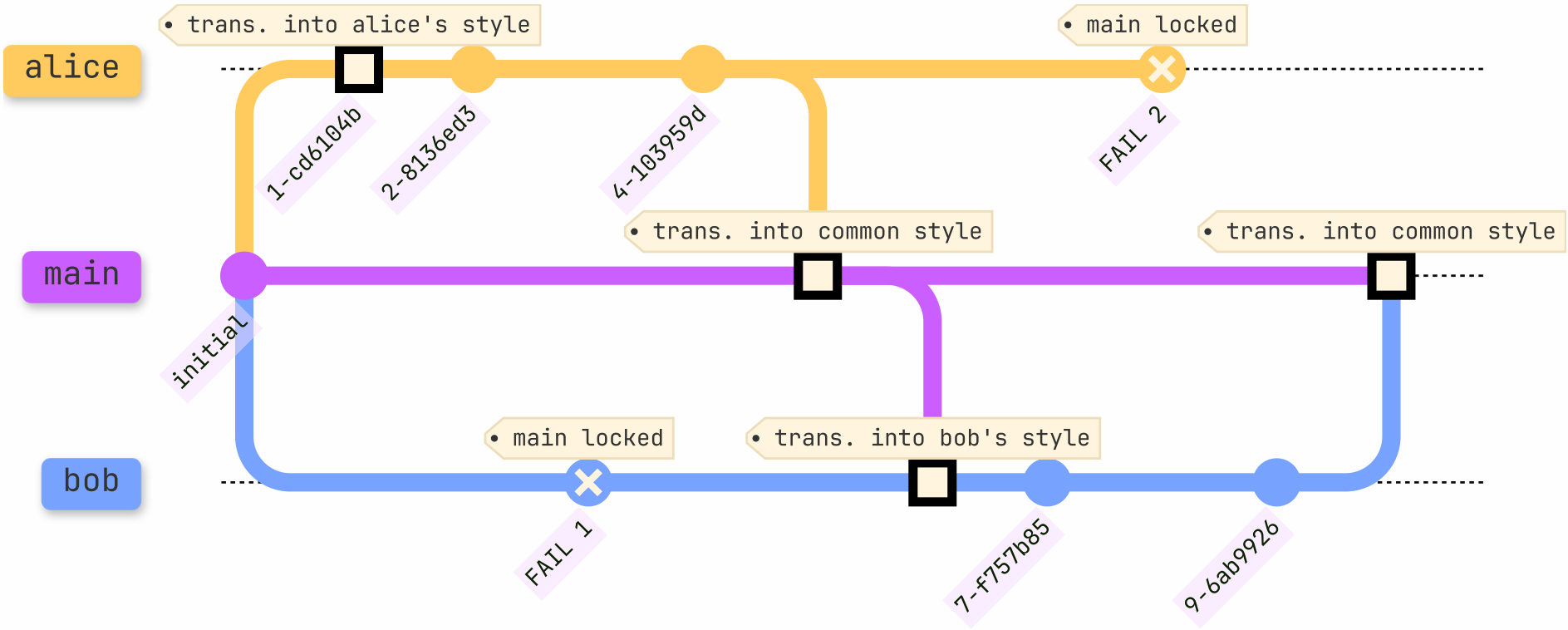
- Cluster 0:
  - lines up to ~2600 chars
- Cluster 1:
  - ~48 chars/line, ~655 max
  - fewer comments
- Cluster 2:
  - shorter lines, tab-indented
- Cluster 3:
  - ~21% of lines have comments

https://github.com/Bart0TW/LaTeX_academic_dataset



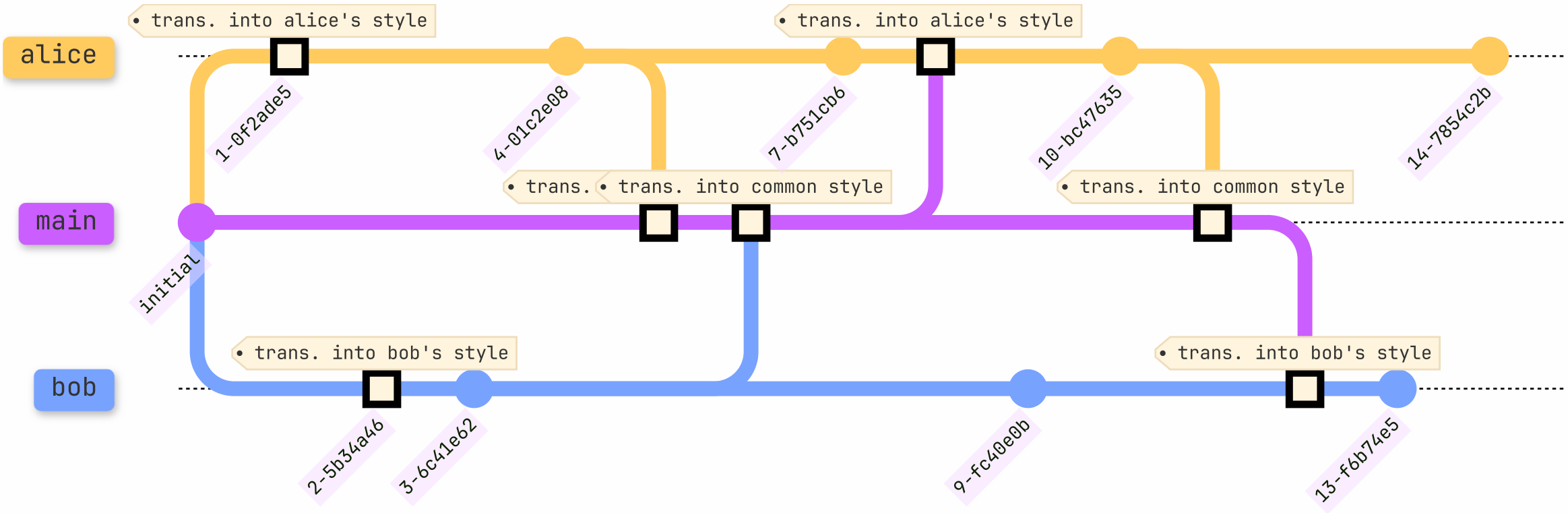


FLEXITEX (turn-based)



http://purl.utwente.nl/essays/107262

FLEXITEX (diff-based)



http://purl.utwente.nl/essays/107262

Conclusion

- LaTeX projects are like software projects
- LaTeX users are determined
- we interviewed ~30 people (12 from CS)
- we analysed 215 projects
  - https://github.com/Bart0TW/LaTeX_academic_dataset
- we release the FlexiTeX tool (MIT)
  - https://github.com/wtb04/FlexiTeX