

Optimising Ph.D. workflow: preparing a polished thesis in \LaTeX , R, and open-source software in less time

Dr. Andreï V. Kostyrka
University of Luxembourg

1. Brief description

The goal of this course is to showcase some free and open-source tools useful for researchers and to demonstrate how to efficiently use them to save time and effort formatting, managing bibliography, and making changes at earlier stages of workflow without redoing later steps. The focus is on writing in \LaTeX , managing bibliographies and libraries, collaborating with co-authors, organising illustrations and tables, and version control. At the end, course participants will receive feature-rich templates for an article and a full dissertation, and they will possess the required knowledge to tailor these templates to their needs.

Upon successful completion of this course, students will be able to:

- Save time by separating research, write-up, formatting, and embellishments;
- Use multiple tools to check and improve articles and theses in progress;
- Simplify and enhance plots and diagrams;
- Prepare multiple versions of research articles and accompanying presentations efficiently, without multiplying the time to generate tables and images;
- Follow publishers' guidelines and change the entire bibliography style in minutes;
- Collaborate more efficiently, track changes, resolve versioning conflicts, and make back-ups.

2. Intended audience

This course is intended for Ph.D. students in economics, management, and finance. However, all post-doctoral researchers, Ph.D. or master's students who are planning to publish at least one academic paper are encouraged to participate.

3. Text and materials

The required books, manuals, and articles for this course are listed below.

- Overleaf. (2023a). *Learn \LaTeX in 30 minutes*. and Overleaf. (2023b). *Overleaf knowledge base*.
<https://www.overleaf.com/learn>
https://www.overleaf.com/learn/latex/Learn_LaTeX_in_30_minutes
- Kottwitz, S. (2015). *\LaTeX cookbook*. Packt.
- Oetiker, T., Partl, H., Hyna, I., & Schlegl, E. (2021). *The not so short introduction to $\LaTeX 2_{\epsilon}$* .
<https://tobi.oetiker.ch/lshort/lshort.pdf>
- Roberts, A. (2022). *Getting to grips with \LaTeX* .
- Kottwitz, S. (2021). *\LaTeX beginner's guide* (2nd ed.). Packt.
- Talbot, N. L. C. (2012). *\LaTeX for complete novices*. Dickimaw Books.
- Wikibooks. (2023). *\LaTeX* . Retrieved August 30, 2023, from <https://en.wikibooks.org/wiki/LaTeX>
- Mittelbach, F., & Fischer, U. (2023). *The \LaTeX companion* (3rd ed.). Addison-Wesley.
- Lammport, L. (1994). *\LaTeX : A document preparation system: A document preparation system (2nd edition) (addison-wesley series on tools and techniques for computer t)* (2nd ed.). Addison-Wesley.
- Kopka, H., & Daly, P. W. (2003). *Guide to \LaTeX* (4th ed.). Addison-Wesley.
- Lode, C. (2023). *Even better books with \LaTeX the agile way in 2023: Streamline your writing process and connect with readers from day one*. LODE Publishing.
- Datta, D. (2017). *\LaTeX in 24 hours*. Springer International Publishing.
- Knuth, D. E. (1984). *The $T_{\epsilon}X$ book*. Addison-Wesley.
- Abrahams, P. W., Hargreaves, K. A., & Berry, K. (1990). *$T_{\epsilon}X$ for the impatient*. Addison-Wesley.

The following open-source software are used in this course:

- *Distribution*: T_EXLive or MacT_EX (<https://tug.org>);¹
- *Bibliography*: JabRef (<https://www.jabref.org>), Zotero (<https://www.zotero.org>).
- *Scripting*: The R project for statistical computing (<https://www.r-project.org>) and RStudio (<https://posit.co>);
- *File manipulation*: GNU coreutils, bash, diff, sed, grep via Cygwin (<https://www.cygwin.com>), ripgrep (<https://github.com/BurntSushi/ripgrep>);
- *Image processing*: GIMP (<https://www.gimp.org>), ImageMagick (<https://imagemagick.org>), pngquant (<https://pngquant.org>), jpegoptim (<https://github.com/tjko/jpegoptim>), optipng (<https://optipng.sourceforge.net>);
- *Comparison*: WinMerge (<https://winmerge.org>), Meld (<https://meldmerge.org>).

Other course material (lecture notes, additional data sets, computer exercises) will be made available on the Moodle site of this course.

4. Lecture schedule

The course will be given in a hybrid format, which means that the participants who cannot be present physically due to the limited room capacity / health issues can join the session via Webex/

#	Date	Room	Time	Webex link
1	27/11/2023	CK C14	14:00 – 17:00	Session 1
2	29/11/2023	CK A17	14:00 – 17:00	Session 2
3	04/12/2023	CK C14	14:00 – 17:00	Session 3
4	06/12/2023	CK A17	14:00 – 17:00	Session 4
5	11/12/2023	CK A17	14:00 – 17:00	Session 5

1. Showcase of file formats. Setting up the research workflow. Synchronisation, versioning, backing up. Difference highlighting. Basic principles of L^AT_EX.
2. Writing articles in L^AT_EX. Mathematical formulæ. Environments. Custom macros. Presentations with beamer. References and citations in Bib_TE_X.
3. Bibliography organisation with specialised tools. Changing bibliography styles. Error handling, debugging, and troubleshooting. L^AT_EX best practices.
4. Reproducing graphics from multiple sources. Exporting from Excel and Stata to R. Basics of graphics in R. Highly custom plots. Basics of image editing and retouching.
5. Spelling, grammar, and syntax checking. Incorporating revisions and changes. Basics of reproducible research; future-proofing.

5. Grading

Regular class participation: 10% (every class meeting yields 2%).

Students need to come to the classroom; active participation and questions from the audience are encouraged.

Assignment: 30%. Due: 18th of December 2023. *Other assignment options will be published on Moodle.* The first assignment involves creating a Bib_TE_X library from the personal collection of research articles. Use JabRef, other automation tools, Google Scholar, and publisher information to create a personal BIB resource of at least 50 articles. Check the names, make sure that capital abbreviations are not forced into lowercase by encasing the sensitive words in {curly braces}.

Final project: 60%. Due: 31st of January 2024. The participants may choose one of the following tasks:

¹The use of MiK_TE_X is highly discouraged because it is an error-prone custom distribution with little maintenance. Use only official T_EXLive of MacT_EX distributions.

- *In fair weather prepare for foul.* Assemble your existing research into a preliminary version of the Ph.D. thesis. Choose a set of formatting guidelines (e. g. APA) and arrange the formulæ, tables, figures, and bibliography accordingly. Check the \LaTeX syntax with ChkTeX , proof-read the text with LanguageTool and/or Grammarly, produce clean graphics (with R or other appropriate tool). Highlight the isolated changes in the source code at each step with diff, WinMerge, or Meld.
- *Convert your master's thesis into \LaTeX .* This option is relevant for **those who are just starting their Ph.D. career** and do not have any research articles or even drafts. Do the same as described above for your existing master's thesis. Choose a set of formatting guidelines (e. g. APA) and arrange the formulæ, tables, figures, and bibliography accordingly. Check the \LaTeX syntax with ChkTeX , proof-read the text with LanguageTool and/or Grammarly, produce clean graphics (with R or other appropriate tool). Highlight the isolated changes in the source code at each step with diff, WinMerge, or Meld.
- *Convert an MS Word article to \LaTeX .* Take a DOC/DOCX file of a research article (no shorter than 7000 words) with formulæ, graphics, tables, and bibliography, extract the images (recreate them if they do not look professional), create a Bib \LaTeX resource, and produce a beautiful \LaTeX document.
- *Create accompanying material for an article.* Prepare a beautiful A0-format \LaTeX poster and a neat job-market presentation in beamer (no less than 25 slides) for your current article that may be used in the future for presentations at FDEF.
- *Digitise a book and typeset it in \LaTeX .* Take any book that is in public domain, contains formulæ and/or illustrations, and has not been converted to text / digitally typeset (at least 50 pages), scan it (if there is no scanned version), perform optical character recognition via tesseract / leptonica, proof-read the contents, typeset it beautifully, and design a pretty cover.

Hint: **choose the task that is the most relevant for your research** or the one that can be later reused in other projects. You may reuse your existing material in the assignment.

6. Technical requirements for the final project

1. No restrictions on the compiler (pdf \LaTeX or Lua \LaTeX) or CTAN packages, but the submissions must be compilable with full \TeX Live 2023 or on Overleaf **without errors**
2. The amount of warnings and overfull / underfull messages must be brought to a minimum (ideally zero)
3. File formats:
 - .tex, .bib, .R, .py, .rb for the documents / data bases / content generators
 - .png / .pdf / TikZ for graphics
 - .bat or .sh for toolchains
 - .docx and .odt are **not accepted**; .jpg graphics are strongly discouraged

7. Contact information

- Office: Campus Kirchberg, room G214
- Email: andrei.kostyrka@uni.lu
- Course homepage: <https://moodle.uni.lu/course/view.php?id=7455>
- Office hours: By appointment