

Brez nočnih mor zaradi urejanja dokumentov

Uvod v \LaTeX za humaniste

Jakob Lenardič¹
Kristina Pahor de Maiti Tekavčič^{1,2}

¹Inštitut za novejšo zgodovino, Ljubljana

²Filozofska fakulteta UL

Delavnica JT-DH 2024
18. september 2024

Kazalnik

- 1 Predstavitev \LaTeX – osnove in raba
- 2 Zgledi edinstvene uporabe
- 3 Zgradba dokumentov \LaTeX

Predstavitev L^AT_EX – osnove in raba

Kaj je L^AT_EX?

Nekaj zgodovine in osnovnih dejstev

- L^AT_EX naslednik računalniškega programa T_EX, razvit leta 1977.
- Izgovorjava /lateh/.
- Stava in oblikovanje dokumentov s kodo; diametralno nasprotje s pristopom WYSIWYG (npr. Microsoft Word).

Dobra uvoda v slovenščini:

- »Praktičen uvod v L^AT_EX« (Slak, 2018)
- »Ne najkrajši uvod v L^AT_EX 2_ε« (Oetiker in sod., 2006)

Prednosti (in slabosti?) L^AT_EX

Prednosti

- Profesionalna oblika in tipografija.
- Dinamičnost in avtomatizacija sklicov, kazal in bibliografije.
- Ponovna uporaba lastnih že ustvarjenih dokumentov (npr. bibliografija).
- Raba obstoječih predlog pri pripravi rokopisov.
← Zelo pomembno pri mednarodnih revijah z visokim faktorjem vpliva.

Slabosti

- Relativna nedostopnost.
- Neurejenost in nekonsistentnosti obstoječih paketov.

Urejevalniki

Dve vrsti urejevalnikov:

- 1 Spletni: *Overleaf*
- 2 Namizni: *TeXstudio*, *Texmaker*, *TexWorks* itd.

Overleaf (1/3)

Dostopen tule: <https://www.overleaf.com/>.

Potrebuje uporabniški račun.

Prednosti:

- 1 Minimalno oz. skoraj nič ročnega nastavljanja.
- 2 Delna samodejna prilagoditev.
- 3 Živi v oblaku.
- 4 Vizualni urejevalnik.
- 5 Namenjen za kolaboracijo: sledenje sprememb itd.

Overleaf (2/3)

Slabosti (pri brezplačni verziji):

- 1 Omejen čas za kompilacijo.
- 2 Naprednejše funkcije niso dostopne.

Sicer pa: najboljši how-to za L^AT_EX nasploh prav od Overleaf:

<https://www.overleaf.com/learn>

Overleaf(3/3)

The screenshot displays the Overleaf online LaTeX editor interface. The top bar includes a menu, an upgrade button, the document title 'clarin2024', and icons for review, share, submit, history, layout, and chat. The left sidebar shows a file outline with 'bibliography.bib', 'CLARIN2024.sty', 'CLARIN2024.tex' (selected), and 'parlamin-citation-instr...'. The main editor area is split into two panes. The left pane shows the source code of the document, with line numbers 100 to 107 visible. The right pane shows the rendered PDF output, which includes a table of citation strategies and a caption for Table 1.

Source Code (Left Pane):

```
100 both citation practice and existing guidelines.
101 \end{abstract}
102 \raggedbottom
103
104 \section[Introduction] \label{intro}
105
106 This paper presents and discusses how the CLARIN
107 community typically cites language resource
108 technologies (LRTs), such as corpora, lexical datasets,
109 and software. We focus on how the citations adhere
110 to the \textit{Joint Declaration of Data Citation
111 Principles} \citep{martone2014data}, specifically to
112 guidelines concerning Authorship Attribution, as well
113 as other principles such as Access and Persistence. We
114 also review the guidelines offered by CLARIN B-centres
115 for data citation. On the basis of the identified
116 inconsistencies in citation practice and the gaps in
117 instructions, we draw up an explicit proposal for LRT
118 citation.
119
120 The paper is structured as follows. In Section
121 \ref{citing datasets in practice}, we present common
122 CLARIN citation practices by using the extended
123 abstracts published in last year's proceedings of the
124 annual CLARIN conference \citep{CACproceedings2023} as
125 a use case. In Section \ref{sec:citation-instructions},
126 we review the citation guidelines of those CLARIN
127 repositories that currently have B-centre
128 certification. Section \ref{ex:sec-proposal} ends the
129 paper with the proposal.
```

Rendered PDF (Right Pane):

1 Introduction

This paper presents and discusses how the CLARIN community typically cites language resource technologies (LRTs), such as corpora, lexical datasets, and software. We focus on how the citations adhere to the *Joint Declaration of Data Citation Principles* (Martone, 2014), specifically to guidelines concerning Authorship Attribution, as well as other principles such as Access and Persistence. We also review the guidelines offered by CLARIN B-centres for data citation. On the basis of the identified inconsistencies in citation practice and the gaps in instructions, we draw up an explicit proposal for LRT citation.

The paper is structured as follows. In Section 2, we present common CLARIN citation practices by using the extended abstracts published in last year's proceedings of the annual CLARIN conference (Lindén et al., 2023a) as a use case. In Section 3, we review the citation guidelines of those CLARIN repositories that currently have B-centre certification. Section 4 ends the paper with the proposal.

2 Citing LRTs in practice

The results of the survey of citing LRTs in the abstracts published in (Lindén et al., 2023a) are given in Table 1. What we have looked at is how any kind of LRT, be it a language corpus, lexical resource, tool, or language model, is cited in the abstracts. The table shows that there are overall 189 citations across the 37 extended abstracts published in the proceedings, while citations are grouped together under 5 major citing strategies.

Citation strategy	#	%
LRT citation	20	11%
Paper about LRT	74	39%
Paper & LRT citation	7	4%
URL in text	76	40%
Paper & URL in text	12	6%
Total	189	100%

Table 1: Citing practices in CLARIN Conference Proceedings 2023 (Lindén et al., 2023a)

The first strategy is what we label 'LRT citation', which makes up 20 (or 11%) of the citation cases. This refers to those cases where the actual LRT is cited, with full authorship attribution and a URL to the repository where the LRT or its metadata are available. Several examples of this strategy are found in the paper by Xu et al. (2023), who for instance cite the *Lexical pronunciation dictionary for language technology* by providing the following reference: (Nikulašdóttir, 2021). Because this sort of citation ensures the proper findability, identifiability and authorship attribution of the LRTs, it is – along with the

¹See the full results in [this Google Spreadsheet](#).

Namizna raba – osnove

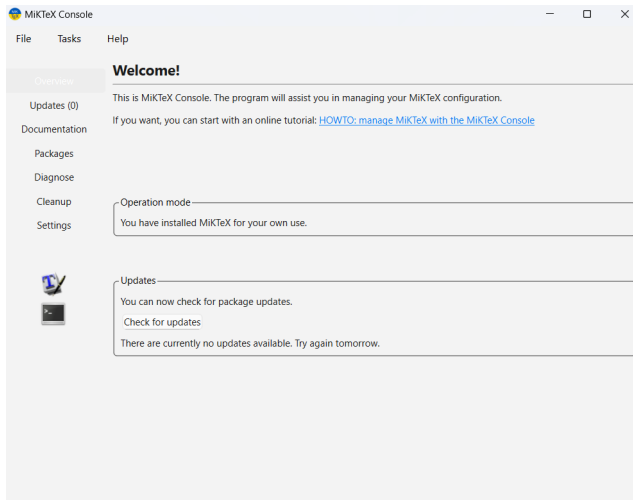
Potrebni dve reči:

- 1 Namestitev neke distribucije \LaTeX .
- 2 Namestitev enega izmed urejevalnikov.

Več distribucij:

npr. **TeX Live**, **MacTex**, **MiKTeX** ← zastonjsko in/ali
odprtokodno!

Namizna raba – zgled konzole MiKTeX



Namizna raba – \TeX studio

<https://www.texstudio.org/>

The screenshot shows the TeXstudio interface. The left pane displays the LaTeX source code for a document titled 'bibliography.bib'. The code defines a frame containing a list of items under the heading 'Prednosti' (Pros) and a list of items under the heading 'Slabosti' (Cons). The right pane shows the rendered output of the document. The title 'Prednosti (in slabosti?) \LaTeX ' is displayed at the top. Below it, the section 'Prednosti' is followed by a bulleted list of pros, and the section 'Slabosti' is followed by a bulleted list of cons. The status bar at the bottom indicates the document is in 'Ready' state, using 'sl_Slovenian' and 'UTF-8' encoding.

```
\begin{frame}(Prednosti (in slabosti?) \LaTeX{})  
  Prednosti  
  \begin{itemize}  
    \item Profesionalna oblika in tipografija.  
    \item Dinamičnost in avtomatizacija sklicov, kazal in bibliografije.  
    \item Ponovna uporaba lastnih že ustvarjenih dokumentov (npr. bibliografija).  
    \item Raba obstoječih predlog pri pripravi rokopisov. \(\leftarrow\) Zelo pomembno pri mednarodnih revijah z visokim faktorjem vpliva.  
  \end{itemize}  
  Slabosti  
  \begin{itemize}  
    \item Relativna nedostopnost.  
    \item Neurejenost in nekonsistentnosti obstoječih paketov.  
  \end{itemize}  
\end{frame}
```

Prednosti (in slabosti?) \LaTeX

Prednosti

- Profesionalna oblika in tipografija.
- Dinamičnost in avtomatizacija sklicov, kazal in bibliografije.
- Ponovna uporaba lastnih že ustvarjenih dokumentov (npr. bibliografija).
- Raba obstoječih predlog pri pripravi rokopisov.
← Zelo pomembno pri mednarodnih revijah z visokim faktorjem vpliva.

Slabosti

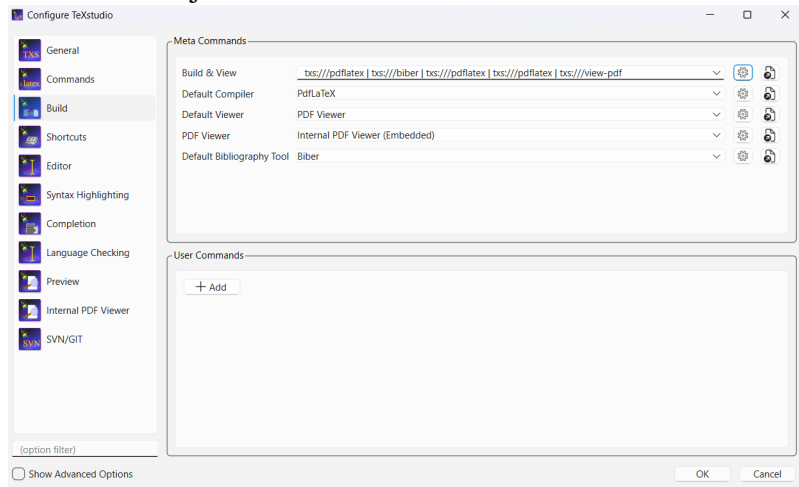
- Relativna nedostopnost.
- Neurejenost in nekonsistentnosti obstoječih paketov.

Podsistemi urejevalnika (1/2)

- Trije (štirje?) različni kompajlerji:
 - ① pdfLaTeX
 - ② XeLaTeX
 - ③ LuaLaTeX
 - ④ ?LaTeX
- Bibliografski sistemi:
 - ① natbib
 - ② bibtex/biblatex
- pdfViewer
- Cel kup drugih reči: npr. spellchecker.

Podsistemi urejevalnika (2/2)

Potrebno nekaj ročne nastavitve



Zgledi edinstvene uporabe

Hiperpovezave

- Bibliografija **eksaktno** upošteva obstoječe standarde (npr. APA, CMOS, MLA).
- Dinamične povezave znotraj teksta.

Tudi v tejle predstavitvi.

Npr., slika vmesnika Overleaf je bila na slajdu 9,
slika TeXstudio pa na slajdu 12.

Jezikoslovje – glosirani zgledi

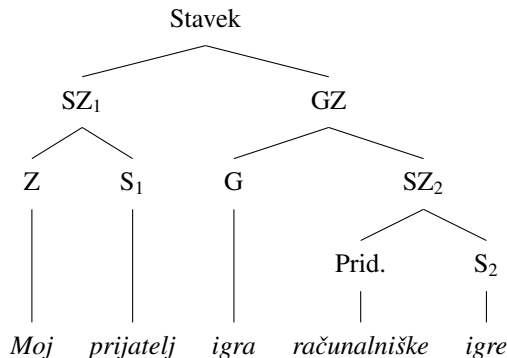
Najlažji način za sledeče sezname:

- (1) a. Jóni líkuðu þessir sokkar. [Icelandic]
Jon.DAT like.PL these socks.NOM
‘John likes these socks.’
- b. Janu so všeč tile štumfi. [Slovenian]
Jan.DAT AUX.3PL like these socks.NOM
‘Jan likes these socks.’

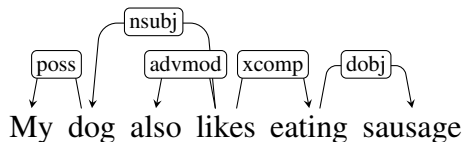
(Zgled (1a) od Bobaljik, 2008, str. 298)

Jezikoslovje – besednozvezna skladnja

Paket **FOREST**, ki ga je razvil Sašo Živanović s FF.



Jezikoslovje – odvisnostna skladnja



Matematika

V formuli (2) je opredeljen test χ^2 .

V formuli (3) pa nekaj logike.

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (2)$$

$$\llbracket Ni \text{ vsak pameten} \rrbracket^g = \neg \forall x P(x) \equiv \exists x \neg P(x) \quad (3)$$

Zgradba dokumentov \LaTeX

Smerokaz

Na kratko predstavimo:

- ① Ukaze
- ② Okolja
- ③ Zgradbo dokumentov
 - ① Vrste datotek
 - ② Preambula
 - ③ Jedro
- ④ Bibliografija

Ukazi

Sledeča skladnja ukazov:

`\funkcija[Par1, ..., Parn]{Arg1, ..., Argn}`

Na primer, ukaz

To je `\textit{ležeči}` tekst.

vrne

To je *ležeči* tekst.

Okolja

Sledeča skladnja okolij:

```
\begin{okolje} ... \end{okolje}
```

Na primer, okolje

```
\begin{equation}
```

$$\frac{dy}{dx} = f(x)$$

```
\end{equation}
```

vrne:

$$\frac{dy}{dx} = f(x) \tag{4}$$

Osnovna zgradba dokumenta

V dveh delih

- Preambula »preamble«
- Jedro dokumenta med `\begin{document}`
...`\end{document}`

Vrste datotek

- Glavni načeloma dve: **.tex** in **.bib**
- Cel kup pomožnih datotek, ki se generirajo ob kompilaciji.
- Datoteke, ki pridejo s predlogami, npr. **.sty**.

Preambula

Zgled

```
1 \documentclass[12pt]{article}
2 \usepackage[T1]{fontenc}
3 \usepackage{geometry}
4 \geometry{
5     a4paper,
6     left=25mm,
7     right=25mm,
8 }
9 \usepackage[backend=biber, style=apa]{biblatex}
10 \addbibresource{bibliography.bib}
11 \usepackage{hyperref}
12
13 \begin{document}
14 ...
15 \end{document}
```

Jedro dokumenta

Zgled

```
\begin{document}  
\section{Uvod} \label{sec:intro}
```

V pričujočem prispevku bomo predstavili korpusnojezikoslovno raziskavo na podlagi korpusa `\textit{ParlaMint 4.1}` `\parencite{erjavecetal2023}`.

Članek ima sledečo zgradbo. V razdelku `\ref{sec:metod}` predstavimo metodologijo. V razdelku ...

```
\section{Metodologija} \label{sec:metod}
```

Tekst tekst ...

```
\printbibliography  
\end{document}
```

Bibliografija

Vedno enaka oblika, ne glede na bibliografski sistem.

Zgled vsebine v .bib fajlu za članek (Erjavec in sod., 2023)

```
@article{erjavecetal2023,  
  title = {The ParlaMint corpora of parliamentary proceedings},  
  author = {Tomaž Erjavec and others},  
  year = {2023},  
  journal = {Language Resources and Evaluation},  
  volume = {57},  
  pages = {415--448},  
  doi = {10.1007/s10579-021-09574-0},  
}
```

Še nekaj reči

- Matematično okolje znotraj \$ \$:. Na primer,

`$\forall x \exists y$`

producira

$\forall x \exists y$

- Marsikateri znaki, npr.

`%, _, &, {, }`

so operatorji.

Uide se jim z rabo `\`, npr. `\&`.

- Vse kar sledi `%`, kompajler ignorira.
- Prelom vrstic z `\\`.

Literatura I



Bobaljik, J. D. (2008). Where's phi? Agreement as a post-syntactic operation. V D. Harbour, D. Adger & S. Béjar (Ur.), *Phi-Theory: Phi features across interfaces and modules*.



Erjavec, T., Ogrodniczuk, M., Osenova, P., Ljubešić, N., Simov, K., Pančur, A., Rudolf, M., Kopp, M., Barkarson, S., Steingrímsson, S., in sod. (2023). The ParlaMint corpora of parliamentary proceedings. *Language resources and evaluation*, 57(1), 415–448.
<https://doi.org/10.1007/s10579-021-09574-0>



Oetiker, T., Partl, H., Hyna, I., & Schlegl, E. (2006). Ne najkrajši uvod v L^AT_EX 2_ε [Prevod: Bor Plestenjak].
<https://users.fmf.uni-lj.si/plestenjak/vaje/latex/lshort.pdf>

Literatura II



Slak, J. (2018). Praktičen uvod v L^AT_EX.

https://e6.ijs.si/~jslak/files/prakticen_uvod_v_latex.pdf