

Legal NER for Indian Court Preamble using spaCy

Aditya Pandey and Shaily Goyal and Vedant Arvind Kumbhare
University of Colorado Boulder

Abstract

The primary goal is to train a model to do Named Entity recognition on Indian Court documents, specifically the preamble section.

1 Introduction

These instructions are for authors submitting papers to *ACL conferences using L^AT_EX. They are not self-contained. All authors must follow the general instructions for *ACL proceedings,¹ and this document contains additional instructions for the L^AT_EX style files.

The templates include the L^AT_EX source of this document (`acl.tex`), the L^AT_EX style file used to format it (`acl.sty`), an ACL bibliography style (`acl_natbib.bst`), an example bibliography (`custom.bib`), and the bibliography for the ACL Anthology (`anthology.bib`).

2 Engines

To produce a PDF file, pdfL^AT_EX is strongly recommended (over original L^AT_EX plus dvips+ps2pdf or dvi2pdf). XeL^AT_EX also produces PDF files, and is especially suitable for text in non-Latin scripts.

3 Preamble

The first line of the file must be

```
\documentclass[11pt]{article}
```

To load the style file in the review version:

```
\usepackage[review]{acl}
```

For the final version, omit the `review` option:

```
\usepackage{acl}
```

To use Times Roman, put the following in the preamble:

¹<http://acl-org.github.io/ACL/PUB/formatting.html>

Command	Output	Command	Output
<code>{\ "a}</code>	ä	<code>{\ c c}</code>	ç
<code>{\ ^e}</code>	ê	<code>{\ u g}</code>	ğ
<code>{\ 'i}</code>	ì	<code>{\ l}</code>	ł
<code>{\ .I}</code>	İ	<code>{\ ~n}</code>	ñ
<code>{\ o}</code>	ø	<code>{\ H o}</code>	ő
<code>{\ 'u}</code>	ú	<code>{\ v r}</code>	ř
<code>{\ aa}</code>	å	<code>{\ ss}</code>	ß

Table 1: Example commands for accented characters, to be used in, *e.g.*, BibT_EX entries.

```
\usepackage{times}
```

(Alternatives like `txfonts` or `newtx` are also acceptable.)

Please see the L^AT_EX source of this document for comments on other packages that may be useful.

Set the title and author using `\title` and `\author`. Within the author list, format multiple authors using `\and` and `\And` and `\AND`; please see the L^AT_EX source for examples.

By default, the box containing the title and author names is set to the minimum of 5 cm. If you need more space, include the following in the preamble:

```
\setlength\titlebox{<dim>}
```

where `<dim>` is replaced with a length. Do not set this length smaller than 5 cm.

4 Document Body

4.1 Footnotes

Footnotes are inserted with the `\footnote` command.²

4.2 Tables and figures

See Table 1 for an example of a table and its caption. **Do not override the default caption sizes.**

²This is a footnote.

4.3 Hyperlinks

Users of older versions of L^AT_EX may encounter the following error during compilation:

```
\pdfendlink ended up in
different nesting level
than \pdfstartlink.
```

This happens when pdfL^AT_EX is used and a citation splits across a page boundary. The best way to fix this is to upgrade L^AT_EX to 2018-12-01 or later.

4.4 Citations

Table 2 shows the syntax supported by the style files. We encourage you to use the natbib styles. You can use the command `\citet` (cite in text) to get “author (year)” citations, like this citation to a paper by Gusfield (1997). You can use the command `\citep` (cite in parentheses) to get “(author, year)” citations (Gusfield, 1997). You can use the command `\citealp` (alternative cite without parentheses) to get “author, year” citations, which is useful for using citations within parentheses (e.g. Gusfield, 1997).

4.5 References

The L^AT_EX and BibT_EX style files provided roughly follow the American Psychological Association format. If your own bib file is named `custom.bib`, then placing the following before any appendices in your L^AT_EX file will generate the references section for you:

```
\bibliography{custom}
```

You can obtain the complete ACL Anthology as a BibT_EX file from <https://aclweb.org/anthology/anthology.bib.gz>. To include both the Anthology and your own .bib file, use the following instead of the above.

```
\bibliography{anthology, custom}
```

Please see Section 5 for information on preparing BibT_EX files.

4.6 Appendices

Use `\appendix` before any appendix section to switch the section numbering over to letters. See Appendix A for an example.

5 BibT_EX Files

Unicode cannot be used in BibT_EX entries, and some ways of typing special characters can disrupt BibT_EX’s alphabetization. The recommended way of typing special characters is shown in Table 1.

Please ensure that BibT_EX records contain DOIs or URLs when possible, and for all the ACL materials that you reference. Use the `doi` field for DOIs and the `url` field for URLs. If a BibT_EX entry has a URL or DOI field, the paper title in the references section will appear as a hyperlink to the paper, using the `hyperref` L^AT_EX package.

Acknowledgements

This document has been adapted by Steven Bethard, Ryan Cotterell and Rui Yan from the instructions for earlier ACL and NAACL proceedings, including those for ACL 2019 by Douwe Kiela and Ivan Vulić, NAACL 2019 by Stephanie Lukin and Alla Roskovskaya, ACL 2018 by Shay Cohen, Kevin Gimpel, and Wei Lu, NAACL 2018 by Margaret Mitchell and Stephanie Lukin, BibT_EX suggestions for (NA)ACL 2017/2018 from Jason Eisner, ACL 2017 by Dan Gildea and Min-Yen Kan, NAACL 2017 by Margaret Mitchell, ACL 2012 by Maggie Li and Michael White, ACL 2010 by Jing-Shin Chang and Philipp Koehn, ACL 2008 by Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui, ACL 2005 by Hwee Tou Ng and Kemal Oflazer, ACL 2002 by Eugene Charniak and Dekang Lin, and earlier ACL and EACL formats written by several people, including John Chen, Henry S. Thompson and Donald Walker. Additional elements were taken from the formatting instructions of the *International Joint Conference on Artificial Intelligence* and the *Conference on Computer Vision and Pattern Recognition*.

References

- Rie Kubota Ando and Tong Zhang. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853.
- Galen Andrew and Jianfeng Gao. 2007. Scalable training of L1-regularized log-linear models. In *Proceedings of the 24th International Conference on Machine Learning*, pages 33–40.
- Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

Output	natbib command	Old ACL-style command
(Gusfield, 1997)	\citep	\cite
Gusfield, 1997	\citealp	no equivalent
Gusfield (1997)	\citet	\newcite
(1997)	\citeyearpar	\shortcite

Table 2: Citation commands supported by the style file. The style is based on the natbib package and supports all natbib citation commands. It also supports commands defined in previous ACL style files for compatibility.

Mohammad Sadegh Rasooli and Joel R. Tetreault. 2015.
[Yara parser: A fast and accurate dependency parser.](#)
Computing Research Repository, arXiv:1503.06733.
Version 2.

A Example Appendix

This is an appendix.