

Anything2vec: generalizing word2vec via the use of meaningful contexts

Abstract

In this paper we present a generalized approach to extend the use of word2vec for non traditional NLP (Natural Language Processing). In order to exemplify the idea we use tvtropes dataset (trope names and film names only) to create a text corpus in order to give contextual information to any piece of data.

Introduction

One of the most important elements in NLP (Natural Language Processing) is context. Word2vec (Mikolov et al. 2013)

The *ICCC Proceedings* will be produced from electronic manuscripts submitted by the authors. These must be PDF (*Portable Document Format*) files formatted for 8-1/2" × 11" paper.

Please see the conference web site (accessible from <http://computationalcreativity.net/>) for submission instructions, including different possible paper types and their lengths.

Word Processing Software

As detailed below, ICCC has prepared and made available a set of L^AT_EX macros and a Microsoft Word template for use in formatting your paper. If you are using some other word processing software, please follow the format instructions given below and ensure that your final paper looks as much like this sample as possible.

Style and Format

L^AT_EX and Word templates that implement these instructions can be retrieved electronically at the conference web site.

Layout

Print manuscripts two columns to a page, in the manner in which these instructions are printed. The exact dimensions for pages are:

- left and right margins: 0.75"
- column width: 3.375"
- gap between columns: 0.25"
- top margin—first page: 1.375"

- top margin—other pages: 0.75"
- bottom margin: 1.25"
- column height—first page: 6.625"
- column height—other pages: 9"

Format of Electronic Manuscript

For the production of the electronic manuscript, you must use Adobe's *Portable Document Format* (PDF). A PDF file can be generated, for instance, on Unix systems using pdf_latex (or ps2pdf from a postscript file) or on Windows systems using Adobe's Distiller. There is also a website with free software and conversion services: <http://www.ps2pdf.com/>. For reasons of uniformity, use of Adobe's *Times Roman* font is strongly suggested. In L^AT_EX2e, this is accomplished by putting

```
\usepackage{times}
```

in the preamble.

Additionally, you must specify the American **letter** format (corresponding to 8-1/2" × 11") when formatting the paper.

Title and Author Information

Center the title on the entire width of the page in a 15-point bold font. Below it, center the author name(s) in a 12-point bold font, and then center the address(es) in a 10-point regular font. Credit to a sponsoring agency can appear on the first page as a footnote.

Blind Review ICCC may use a double blind review process where the author information must (or may) be hidden in the submissions. Please see the call for papers for instructions on possible blind reviews.

Abstract

Place the abstract at the beginning of the first column 3" from the top of the page, unless that does not leave enough room for the title and author information. Use a slightly smaller width than in the body of the paper. Head the abstract with "Abstract" centered above the body of the abstract in a 10-point bold font. The body of the abstract should be 9-point in the same font as the body of the paper.

The abstract should be a concise, one-paragraph summary describing the general thesis and conclusion of your paper.

A reader should be able to learn the purpose of the paper and the reason for its importance from the abstract. As a guide, abstracts should normally be up to 250 words long.

Text

The main body of the text immediately follows the abstract. Use 10-point type in *Times Roman* font.

Indent when starting a new paragraph, except after major headings.

Headings and Sections

When necessary, headings should be used to separate major sections of your paper. (These instructions use many headings to demonstrate their appearance; your paper should have fewer headings.)

Section Headings Print section headings centered, in 12-point bold type in the style shown in these instructions. Leave a blank space of approximately 10 points above and 4 points below section headings. Do not number sections.

Subsection Headings Print subsection headings left justified, in 12-point bold type. Leave a blank space of approximately 8 points above and 3 points below subsection headings. Do not number subsections.

Subsubsection Headings Print subsubsection headings inline in 10-point bold type. Leave a blank space of approximately 6 points above subsubsection headings. Do not number subsubsections.

Special Sections You may include an unnumbered acknowledgments section, including acknowledgments of help from colleagues, financial support, and permission to publish.

Any appendices directly follow the text and look like sections. In general, appendices should be avoided in ICCG manuscripts.

The references section is headed “References,” printed in the same style as a section heading. A sample list of references is given at the end of these instructions. Note the various examples for books, proceedings, multiple authors, etc. Use a consistent format for references, such as that provided by BiBTeX. The reference list should not include unpublished work.

Citations

Citations within the text should include the author’s last name and the year of publication, for example (Boden 1992). Append lowercase letters to the year in cases of ambiguity. Treat multiple authors as in the following examples: (Lyu, Rockmore, and Farid 2004) (for more than two authors) and (Veale and Hao 2007) (for two authors). If the author portion of a citation is obvious, omit it, e.g., Woods (1981). Collapse multiple citations as follows: (Asuncion and Newman ; Ruch 2007; Kantrowitz 1990). *See below for L^AT_EX commands for producing citations.*

Using L^AT_EX and BiBTeX to Create Your References At the end of your paper, you can include your reference list by using the following commands (which will insert a heading **References** automatically):

```
\bibliographystyle{iccc.sty}
\bibliography{iccc.bib}
```

The list of files in the bibliography command should be the names of your BiBTeX source files (that is, the .bib files referenced in your paper).

The iccc.sty file includes a set of definitions for use in formatting references with BiBTeX. These definitions make the bibliography style fairly close to the one specified previously. To use these definitions, you also need the BiBTeX style file iccc.bst available in the author kit on the ICCG web site.

The following commands are available for your use in citing references:

\cite: Cites the given reference(s) with a full citation. This appears as “(Author Year)” for one reference, or “(Author Year; Author Year)” for multiple references.

\shortcite: Cites the given reference(s) with just the year. This appears as “(Year)” for one reference, or “(Year; Year)” for multiple references.

\citeauthor: Cites the given reference(s) with just the author name(s) and no parentheses.

\citeyear: Cites the given reference(s) with just the date(s) and no parentheses.

Warning: The iccc.sty file is incompatible with the hyperref package. If you use it, your references will be garbled. *Do not use hyperref!*

Footnotes

Avoid footnotes as much as possible; they interrupt the flow of the text. Place footnotes at the bottom of the page in 9-point font. Refer to them with superscript numbers.¹ Separate them from the text by a short line.²

Illustrations

Place all illustrations (figures, drawings, tables, and photographs) throughout the paper at the places where they are first discussed (at the bottom or top of the page), rather than at the end of the paper. If placed at the bottom or top of a page, illustrations may run across both columns.

Illustrations must be rendered electronically or scanned and placed directly in your document. In most cases, it is best to render all illustrations in black and white; however, since the proceedings are produced and distributed electronically, if color is important for communicating your message, it may be included. Line weights should be 1/2-point or thicker.

Number illustrations sequentially. Use references of the following form: Figure 1, Table 2, etc. Place illustration numbers and captions under illustrations. Leave a margin

¹This is how your footnotes should appear.

²Note the line separating these footnotes from the text.

of 1/4-inch around the area covered by the illustration and caption. Use 9-point type for captions, labels, and other text in illustrations.

Acknowledgments

The preparation of these instructions and the L^AT_EX and Word files was facilitated by borrowing from similar documents used for AAAI and IJCAI proceedings.

References

- Asuncion, A., and Newman, D. UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences. <http://www.ics.uci.edu/mlearn/MLRepository.html>.
- Boden, M. 1992. *The Creative Mind*. London: Abacus.
- Kantrowitz, M. 1990. Natural language text generation in the OZ interactive fiction project. Technical Report CMU-CS-90-158, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
- Lyu, S.; Rockmore, D.; and Farid, H. 2004. A digital technique for art authentication. *Proceedings of the National Academy of Sciences* 101(49):17006–17010.
- Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space. *CoRR* abs/1301.3781.
- Ruch, W., ed. 2007. *The Sense of Humor: Explorations of a Personality Characteristic*. Mouton Select. Berlin: Mouton de Gruyter.
- Veale, T., and Hao, Y. 2007. Comprehending and generating apt metaphors: A web-driven, case-based approach to figurative language. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence (AAAI-07)*, 1471–1476. Vancouver, British Columbia: AAAI Press.
- Woods, W. 1981. Procedural semantics as a theory of meaning. In Joshi, A.; Webber, B.; and Sag, I., eds., *Elements of Discourse Understanding*. Cambridge, UK: Cambridge University Press. 300–334.