



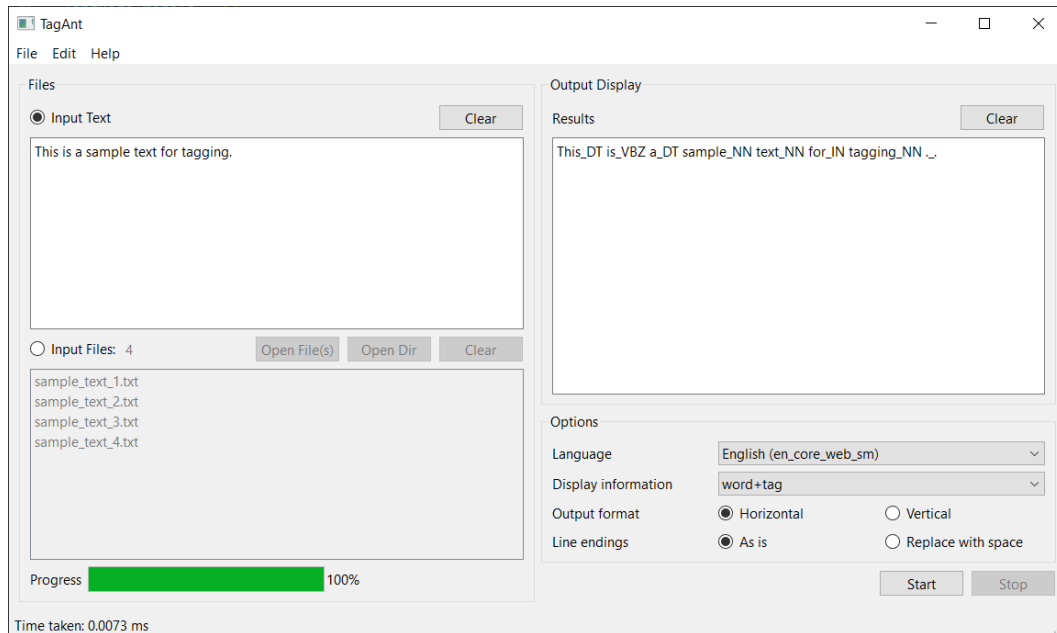
TagAnt (Windows, Macintosh, Linux)

Build 2.0.1

Laurence Anthony, Ph.D.

Center for English Language Education in Science and Engineering, School of Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan.

November 30, 2021



Introduction

TagAnt is a freeware, multi-language tagging tool built on top of the SpaCy natural language processing (NLP) framework. It takes either an input text or an input list of text files (**UTF-8 encoded**) and processes them to generate the following components for each word: WORD, PART-OF-SPEECH (POS) NAME PART-OF-SPEECH (POS) TAG, LEMMA. Options allow the choice of language, the component information to be displayed, the output format, and the handling of line-endings. *TagAnt* runs on any computer running Microsoft Windows (tested on Windows 10) and Macintosh OS computers (tested on OSX 10.15 Catalina). It is developed in Python and Qt using the *PyInstaller* compiler to generate executables for the different operating systems.

Getting Started

Windows - Installer

Double click the *TagAnt* exe file and follow the instructions to install the application into your Programs folder. You can delete the exe file when you are finished. You can start the application via the Start Menu.

Windows - Portable

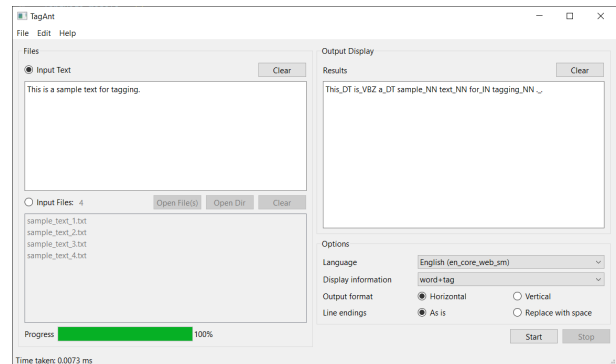
Unzip the *TagAnt* zip file into a folder of your choice. In the *TagAnt* folder, double click the *TagAnt* exe file to launch the program.

Macintosh OS X

On Macintosh systems, simply double click the *TagAnt* pkg file and follow the instructions to install the application into your Applications folder. You can delete the pkg file when you are finished.

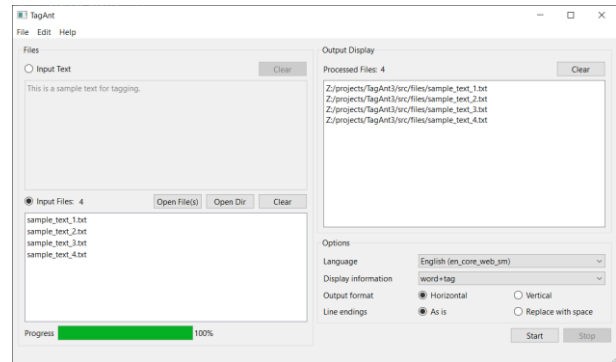
Tagging Input Text

- Step 1:** Under "Files", select the "Input Text" radiobox on the left of the main window.
- Step 2:** Enter the text that you want to process.
- Step 3:** Under "Options" choose the settings to use (see the next section for further details)
- Step 4:** Click "Start" to begin the tagging process. The tagging process can be stopped at any time by clicking the "Stop" button.



Tagging Input Files

- Step 1:** Under "Files", select the "Input Files" radiobox on the left of the main window.
- Step 2:** Enter the text that you want to process.
- Step 3:** Under "Options" choose the settings to use:
- Step 4:** Click "Start" to begin the tagging process.
The tagging process can be stopped at any time by clicking the "Stop" button.



Options

The following options are available during processing.

Language:

Select the language to be processed. The following languages are available:

Catalan, Chinese, Danish, Dutch, English, French, German, Greek, Italian, Japanese, Lithuanian, Macedonian, Multi-language, Norwegian Bokmål, Polish, Portuguese, Romanian, Russian, Spanish

Display information:

Select the word components to be shown. Combinations of the following are possible:

WORD, POS name, POS tag, LEMMA

Output format

Select the output format to be shown. The "Horizontal" (Default) option, will show the output matching the input on a line-by-line basis. The "Vertical" option, will show the components for every word of the input on separate lines. In both cases, the word components are combined with underscores (e.g. dog_NN)

Line endings:

Select how line endings should be processed. The "As is" (Default) option leave the line endings as is, without no changes. The "Replace with space" option, will replace all line endings with spaces, rendering the complete file as a single line file.

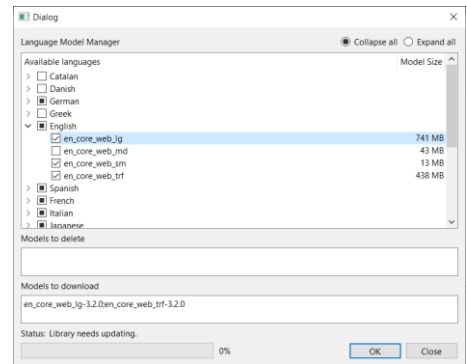
Additional Features

The output display can be selected and copied via the Edit menu. The following shortcuts are also available as is standard on the operating system:

Windows:	CTRL-A ⇨ Select All	CTRL-C ⇨ Copy	CTRL-V ⇨ Paste
Macintosh:	CMD-A ⇨ Select All	CMD -C ⇨ Copy	CMD -V ⇨ Paste

Tagging Input Text

TagAnt comes with several pre-installed SpaCy language models, including those for English, French, German, Italian, Japanese, and Chinese. If you want to tag your data using other language models, you can use the Language Model Manager to install them. Access the Language Model Manager via the File Menu. Select language models to add/remove and then click OK to update the language model library.



NOTES

Comments/Suggestions/Bug Fixes

All new editions and bug fixes are listed in the revision history below. However, if you find a bug in the program, or have any suggestions for improving the program, please let me know and I will try to address the issues in a future version.

This software is available as 'freeware' (see Legal Matter below), but it is important for my funding to hear about any successes that people have with the software. Therefore, if you find the software useful, please send me an e-mail briefly describing how it is being used.

CITING/REFERENCING *TagAnt*

Use the following method to cite/reference *TagAnt* according to the APA style guide:

Anthony, L. (YEAR OF RELEASE). *TagAnt* (Version VERSION NUMBER) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>

For example, if you download *TagAnt 2.0.0*, which was released in 2021, you cite/reference it as follows:

Anthony, L. (2021). *TagAnt* (Version 2.0.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>

Note that the APA instructions are not entirely clear about citing software, and it is debatable whether the "Available from ..." statement is needed. See here for more details:

<http://owl.english.purdue.edu/owl/resource/560/10/>

KNOWN ISSUES

- 1) If a very large file is copied to the "Input Text" box and tagged, once the results are generated, copying the results might use up all the available system memory and cause the program to crash. To avoid this problem, large files should always be tagged using the "Input File" option.