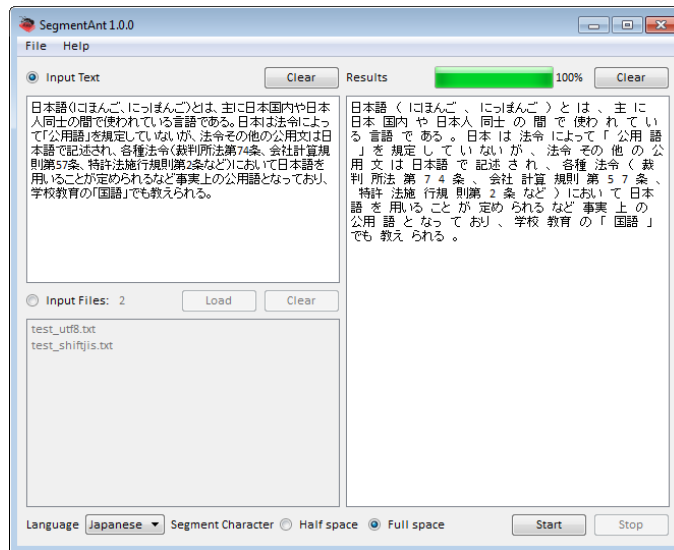# SegmentAnt (Windows)

Build 1.0.0

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.

Help file version: 001 (October 18, 2014).

Introduction

SegmentAnt is a freeware Japanese and Chinese segmenting tool based on various tagging tools including the TinySegmenter engine (developed by Taku Kudo <taku@chasen.org> and implemented in Python by Masato Hagiwara < http://lilyx.net/pages/tinysegmenterp.html>) and the smallseg engine that is available at https://code.google.com/p/smallseg/. *SegmentAnt* takes either an input text or an input list of text files (***UTF-8 encoded***) and splits the texts into 'tokens" separated by either half-width or full-width spaces. *SegmentAnt* runs on any computer running Microsoft Windows (tested on Win 98/Me/2000/NT, XP, Vista, Win 7, Win 8) and Macintosh OS X computers (tested up to OS X 10.9 Mavericks). It is developed in Python and Qt using the *PyInstaller* compiler to generate executables for the different operating systems.

Getting Started (No installation necessary)

Windows

On Windows systems, simply double click the *SegmentAnt* icon to launch the program.

Macintosh OS X

On Macintosh systems, simply double click the *SegmentAnt* zip file. The zip file will unzip the *SegmentAnt* application. Then, you can drag the *SegmentAnt* application to your application folder, your desktop, or anywhere else you like. Throw away the zip file when you are finished.

Segmenting Input Text

Step 1: Select the "Input Text" radiobox on the left of the main window.

Step 2: Select the language of the files using the combobox next to the "Language" label.

Step 3: Choose either the "Half space" or "Full space" radiobox for the segment character of the output results.

Step 4: Click "Start" to begin the segmenting process.

Note 1: The segmenting process can be stopped at any time by clicking the "Stop" button.

Segmenting Input Files

Step 1: Select the "Input Files" radiobox on the left of the main window.

Step 2: Select the files you want to segment. You can do this in four ways:

   a) Click on the File->Open File(s) menu option and select the files you want to segment;

   b) Click on the File->Open Dir menu option and select a directory of files you want to segment;

   c) Click on the "Load" button next to the "Input Files" label and select the files you want to segment;

   d) Drag and drop files directly onto the *SegmentAnt* application.

   Note 1: The number of selected files is shown next to the "Input Files" label.

   Note 2: If you click on the File->Close Files menu option or click the "Close" button next to the "Input Files" label, the input files will removed from the list.

Step 3: Select the language of the files using the combobox next to the "Language" label.

Step 4: Choose either the "Half space" or "Full space" radiobox for the segment character of the output results.

Step 5: Click "Start" to begin the segmenting process.

   Note 1: A segmented version of the original file will be saved in the same folder as the original file with "_segmented" appended to the end of the filename.

   Note 2: The segmenting process can be stopped at any time by clicking the "Stop" button.

Additional Features

The output display can be selected, copied, and pasted 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

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 *SegmentAnt*

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

Anthony, L. (YEAR OF RELEASE). *SegmentAnt* (Version VERSION NUMBER) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

For example if you download *SegmentAnt 1.0.0*, which was released in 2014, you would cite/reference it as follows:

Anthony, L. (2014). *SegmentAnt* (Version 1.0.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http:// www.laurenceanthony.net/

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

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

LEGAL MATTER

SegmentAnt can be used freely for individual use for non-profit research purposes, and freely distributed on the condition that this read me file is attached in an unaltered state. If the software is planned to be used in a group environment, you are required to inform me how the software is to be used, and I will then determine if you can have permission to use it. The software comes on an 'as is' basis, and the author will accept no liability for any damage that may result from using the software.

KNOWN ISSUES

If a very large file is copied to the "Input Text" box and segmented, 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 segmented using the "Input File" option.

REVISION HISTORY

1.0.0
This is the first version of the program