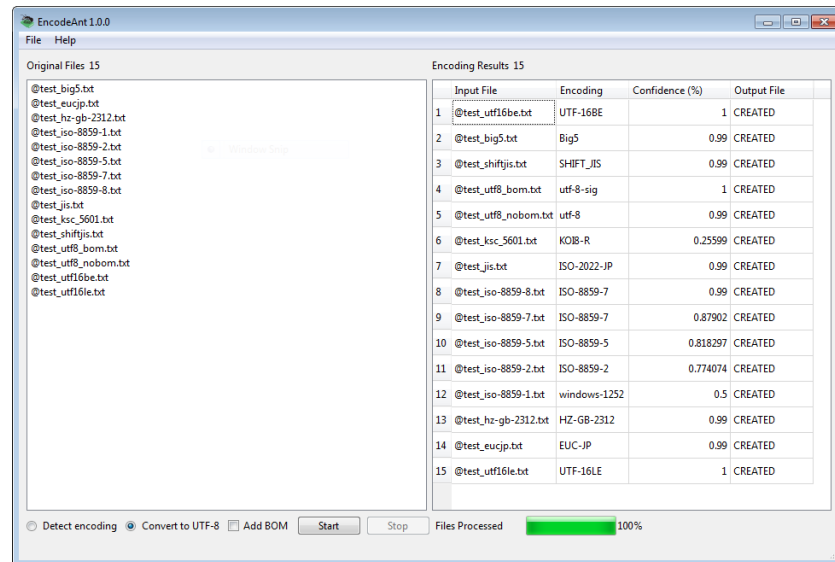# EncodeAnt (Windows)

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 19, 2014).

Introduction

EncodeAnt is a freeware character encoding auto-detection and conversion tool. *EncodeAnt* takes either an input text or an input list of text files (e.g. .txt) and attempts to auto-detect the character encoding that the files use. It also has an option to auto-convert the character encoding of the files to UTF-8, which is a standard used in most corpus research. The converted files are saved in a separate folder leaving the original files untouched.

EncodeAnt runs on any computer running Microsoft Windows (tested on Win 98/Me/2000/NT, XP, Vista, Win 7, Win 8). 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 *EncodeAnt* icon to launch the program.

Step 1: Select the files you want to analyze. You can do this in three 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) Drag and drop files directly onto the *EncodeAnt* application.

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

      Note 2: If you click on the File->Close Files menu option, the files will removed from the list.

Step 2: Choose the "Detect encoding" to detect the encodings of your files or the "Convert to UTF-8" radiobox to convert the encodings of your files to the standard UTF-8 encoding.

Step 3: Click the "Add BOM" checkbox if you want to add a UTF-8 BOM (Byte Order Mark) to the beginning of your files. This is generally not recommended but it is useful on Microsoft Windows system that use the BOM to open the file correctly in programs like Notepad.

Step 4: Click "Start" to begin the detection/conversion process.

    Note 1: If the "Convert to UTF-8" option is selected, the converted files will be saved where the original file was stored under a sub-folder called "utf8".

    Note 2: The detection/conversion 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

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 *EncodeAnt*

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

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

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

Anthony, L. (2014). *EncodeAnt* (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

EncodeAnt 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

None

REVISION HISTORY

1.0.0
This is the first version of the program