



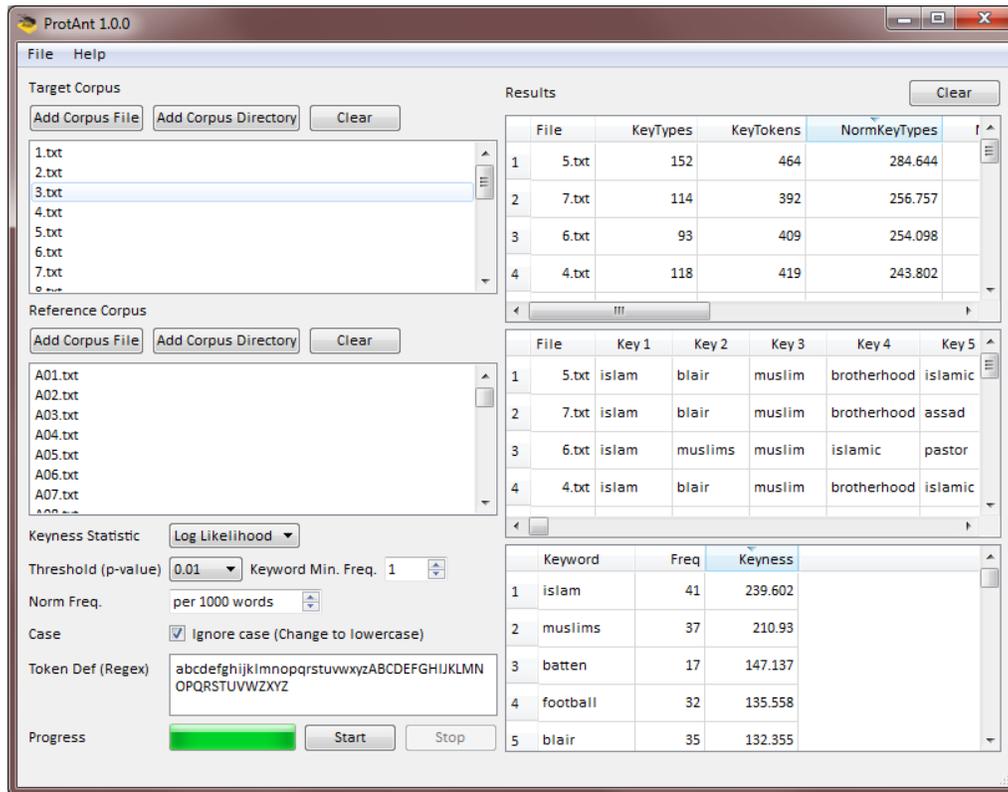
# ProtAnt (Windows)

Build 1.0.1 (Released October 21, 2015)

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 21, 2015).



## Introduction

*ProtAnt* is a freeware prototypical text detection tool developed in collaboration with Paul Baker of Lancaster University, UK. *ProtAnt* takes a corpus of texts (**UTF-8 encoded**) and compares them against a reference corpus (**UTF-8 encoded**) to find keywords that represent the characteristic features of the corpus as a whole. Then, *ProtAnt* looks at each individual file and counts how many keywords are in them. The corpus files are then ranked by the number of keywords they contain normalized in various ways. *ProtAnt* 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). *ProtAnt* 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 *ProtAnt* icon to launch the program.

### Macintosh OS X

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

## Finding prototypical texts

**Step 1:** Select the target files of the corpus you want to analyze. You can do this in four ways:

- a) Click on the File->Open Target File(s) menu option and select the files you want to analyze;
- b) Click on the File->Open Target Dir menu option and select a directory of files you want to analyze;
- c) Click on the "Add Corpus Files" button below the "Target Corpus" label and select the files you want to analyze;
- d) Drag and drop files directly onto the *ProtAnt* application.

Note 1: If you click on the File->Close Target Files menu option or click the "Clear" button below the "Target Corpus" label, the corpus files will be removed from the list.

**Step 2:** Select the reference corpus files that you want to use. You can do this in four ways:

- a) Click on the File->Open Reference File(s) menu option and select the files you want to analyze;
- b) Click on the File->Open Reference Dir menu option and select a directory of files you want to analyze;
- c) Click on the "Add Corpus Files" button below the "Reference Corpus" label and select the files you want to analyze;
- d) Drag and drop files directly onto the *ProtAnt* application.

Note 1: If you click on the File->Close Reference Files menu option or click the "Clear" button below the "Reference Corpus" label, the corpus files will be removed from the list.

**Step 3:** Choose the keyness statistic, keyness threshold value (p-value), and keyword minimum frequency you would like to use to generate keywords for the target corpus.

**Step 4:** Choose the normalization function for frequencies displayed in the results window.

**Step 5:** Choose to ignore case in the target corpus (change all words to lowercase) and decide a suitable token definition based on regular expression (regex) syntax.

**Step 6:** Click "Start" to begin the segmenting process.

Note 1: The results of the prototypical text detection are shown in the top right window. The keywords appearing in each (ranked) corpus file are shown in the middle right window. The complete list of keywords is shown in the bottom right window. All columns can be sorted in either ascending or descending order by clicking on the column headers.

Note 2: The prototypical text detection analysis 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' according to the license below. 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 *ProtAnt***

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

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

For example if you download *ProtAnt 1.0.0*, which was released in 2015, you would cite/reference it as follows:  
Anthony, L. and Baker, P. (2015). *ProtAnt* (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/>

## **LICENSE for *ProtAnt***

*ProtAnt* 1.0 and any minor updates issued by AntLab Solutions (collectively 'the Software')

### TERMS GOVERNING THE USE OF THE SOFTWARE

The Software is protected by copyright and must not be used, displayed, modified, adapted, distributed, transmitted, transferred or published or otherwise reproduced in any form by any means other than strictly in accordance with the terms set out below. By installing the Software, you agree to be bound by the terms of the license. This *ProtAnt* License ("License") is made between AntLab Solutions, Tokyo, Japan as licensor, and you, as licensee, as of the date of your use of the Software. The Software is in use on a computer when it is loaded into the RAM or installed into the permanent memory of that computer, e.g., a hard disk or other storage device.

#### 1. License Material

These terms govern your use of the Software but not including subsequent versions (e.g. *ProtAnt* 2.0').

#### 2. License Grant

AntLab Solutions grants to you a personal non-exclusive non-transferable license ('the License') to use the Software in the following specific contexts.

##### a) Non-Commercial (Freeware) Use:

You may use the software for non-profit purposes on more than one computer or on a network so long as you are the sole user of the Software. (A "network" is any combination of two or more computers that are electronically linked and capable of sharing the use of a single software program.) You are not permitted to sell, lease, distribute, transfer, sublicense, or otherwise dispose of the Software, in whole or in part, for any form of actual or potential commercial gain or consideration.

##### b) Commercial Evaluation (Trial) Use:

You may evaluate (trial) the software for commercial purposes for a period of no more than fourteen (14) days from the date of download on more than one computer or on a network so long as you are the sole user of the Software.

##### c) Commercial Use

When you pay the commercial license fee established by AntLab Solutions, you may use the software for non-profit or commercial purposes on more than one computer or on a network so long as you are the sole user of

the Software. (A “network” is any combination of two or more computers that are electronically linked and capable of sharing the use of a single software program.) You will obtain a separate license for each additional user of the Software (whether or not such users are connected on a network). You are not permitted to sell, lease, distribute, transfer, sublicense, or otherwise dispose of the Software, in whole or in part, for any form of actual or potential commercial gain or consideration.

### 3. Termination

You may terminate this License at any time by uninstalling the Software and deleting it. The License will also terminate if you breach any of the terms of the License.

### 4. Proprietary Rights

The Software is licensed, not sold, to you. AntLab Solutions reserves all rights not expressly granted to you. Ownership of the Software and its associated proprietary rights, including but not limited to patent and patent applications, are retained by AntLab Solutions. The Software is protected by the copyright laws of Japan and by international treaties. Therefore, you must comply with such laws and treaties in your use of the Software. You agree not to remove any of AntLab Solutions' copyright, trademarks, and other proprietary notices from the Software.

### 5. Distribution

Except as may be expressly allowed in Section 2, or as otherwise agreed to in a written agreement signed by both you and AntLab Solutions, you will not distribute the Software, either in whole or in part, in any form or medium.

### 6. Transfer and Use Restrictions

You may not sell, license, sub-license, lend, lease, rent, share, assign, transmit, telecommunicate, export, distribute or otherwise transfer the Software to others, except as expressly permitted in this License Agreement or in another agreement with AntLab Solutions. You may not modify, reverse engineer, decompile, decrypt, extract, or otherwise disassemble the Software.

### 7. Warranties

ANTLAB SOLUTIONS MAKES NO WARRANTIES WHATSOEVER REGARDING THE SOFTWARE AND IN PARTICULAR, DOES NOT WARRANT THAT THE SOFTWARE WILL FUNCTION IN ACCORDANCE WITH THE ACCOMPANYING DOCUMENTATION IN EVERY COMBINATION OF HARDWARE PLATFORM OR SOFTWARE ENVIRONMENT OR CONFIGURATION, OR BE COMPATIBLE WITH EVERY COMPUTER SYSTEM. IF THE SOFTWARE IS DEFECTIVE FOR ANY REASON, YOU WILL ASSUME THE ENTIRE COST OF ALL NECESSARY REPAIRS OR REPLACEMENTS.

### 8. Disclaimer

ANTLAB SOLUTIONS DOES NOT WARRANT THAT THE SOFTWARE OR SERVICE IS FREE FROM BUGS, DEFECTS, ERRORS OR OMISSIONS. THE SOFTWARE OR SERVICE IS PROVIDED ON AN “AS IS” BASIS AND ANTLAB SOLUTIONS MAKES NO OTHER WARRANTIES OR CONDITIONS, EXPRESS OR IMPLIED, WITH RESPECT TO THE SOFTWARE INCLUDING WITHOUT LIMITATION THE IMPLIED WARRANTIES OR CONDITIONS OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

### 9. Limitation of Liability

ANTLAB SOLUTIONS WILL HAVE NO LIABILITY OR OBLIGATION FOR ANY DAMAGES OR REMEDIES, INCLUDING, WITHOUT LIMITATION, THE COST OF SUBSTITUTE GOODS, LOST DATA, LOST PROFITS, LOST REVENUES OR ANY OTHER DIRECT, INDIRECT, INCIDENTAL, SPECIAL, GENERAL, PUNITIVE OR CONSEQUENTIAL DAMAGES, ARISING OUT OF THIS LICENSE OR THE USE OR INABILITY TO USE THE SOFTWARE OR SERVICE. IN NO EVENT WILL ANTLAB SOLUTIONS'S TOTAL AGGREGATE LIABILITY (WHETHER IN CONTRACT (INCLUDING FUNDAMENTAL BREACH), WARRANTY, TORT (INCLUDING NEGLIGENCE), PRODUCT LIABILITY, INTELLECTUAL PROPERTY

INFRINGEMENT OR OTHER LEGAL THEORY) WITH REGARD TO THE SOFTWARE AND/OR THIS LICENSE EXCEED THE LICENSE FEE PAID BY YOU TO ANTLAB SOLUTIONS. FURTHER, ANTLAB SOLUTIONS WILL NOT BE LIABLE FOR ANY DELAY OR FAILURE TO PERFORM ITS OBLIGATIONS UNDER THIS LICENSE AS A RESULT OF ANY CAUSES OR CONDITIONS BEYOND ANTLAB SOLUTIONS' REASONABLE CONTROL

#### 10. Jurisdiction

These terms will be governed by Japanese law and the Japanese courts shall have jurisdiction.

## **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**

### 1.0.1

This is minor update

Bug fix: When a general Unicode character class token definition was supplied (e.g. `\p{L}` for letters), non-English texts were not being processed correctly. This is now fixed.

### 1.0.0

This is the first version of the program

Copyright 2015 Laurence Anthony. All rights reserved.