



# **User Guide for Corpus Analyzing Tool**

**October 25, 2007**

**Language Technology Research Laboratory  
University of Colombo School of Computing  
Sri Lanka**

## Table of Contents

1	Introduction	3
2	Run the Application	3
3	Use the CAT	

# 1 Introduction

This document explains how to use the Corpus analyzing Tool [CAT]. This software offers the users to analyze text files and get the concordance and the unique word list. It has a special functionality for Automatic POS Tagging which was inbuilt to the system.

## 2 Running the Application

After successful installation of software and all its dependencies, follow the steps below to start the application.

### 2.1 Running on Windows

Go to,  
**Start > All programs > UCSC** and click on **CAT**.

## 3 Use the CAT

The start up screen of **CAT** appears as in Figure 3.1.

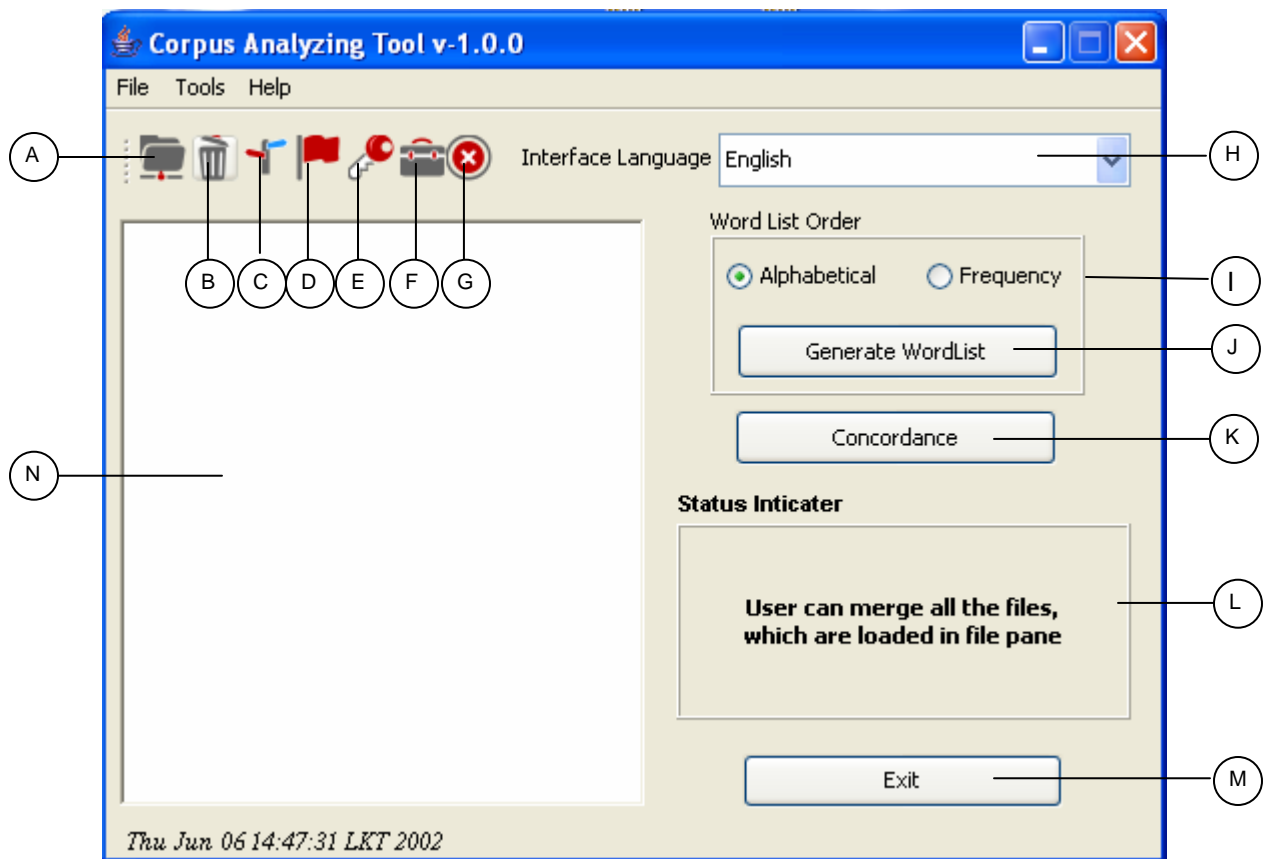


Figure 3.1: CAT Initial Screen

## PAN localization project

- A - Open files for the analysis
- B - Reset
- C - Test Encoding
- D - File Availability checking
- E - Open .cfg files
- F - Save / Create.cfg files
- G - Exit
- H - Interface Language selection option
- I - Order Option
- J - Generate the word list
- K - Generate the concordance
- L - Status Option
- M - Exit
- N - File details pane

### Important points for Usage of CAT

1. Fresh users need to select the text files from different locations to analysis with using *Open Files Button (A)* on the tool bar or go to *File → Open .txt files*.  
User need to save the file details by press the *Save File Details Button (F)* on the tool bar. (The user settings will be saved as .cfg file, in the given location)  
If the user already has saved file details, then click on the *Open File Details Button (E)*, on the tool bar or go to *File → Open .cfg files* to open the cfg file.

**NOTE: Keep in mind that all the files are needed same language for analysis. The CAT does not support Multilanguage text files.**

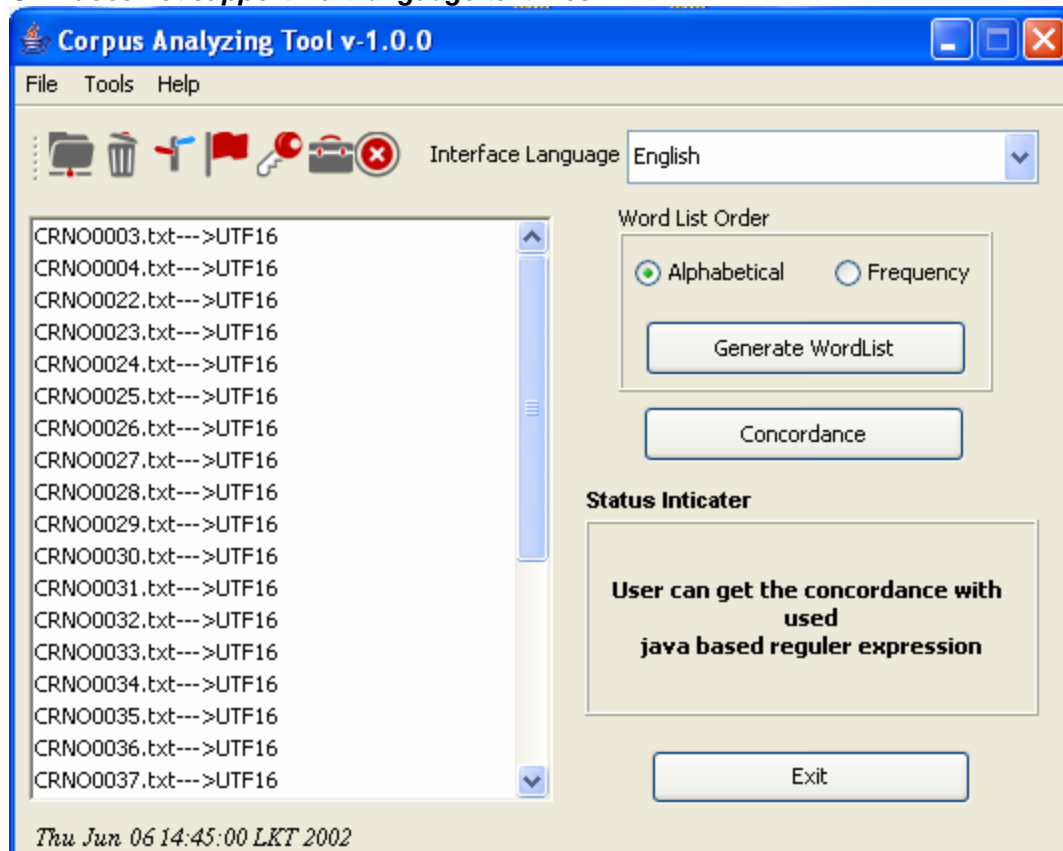


Figure 3.2: Show the File encoding



6. **To get the Concordance**, click the button *Concordance (K)* after that the Concordance setting window will appear.  
User need to given the searching word, left and right side number of words, and regular expression option. Here the tool support only java based regular expressions.

To search the word with using the regular expression then select the Regular expression based search option.

The following Figure 3.4 shows the Concordance setting window.

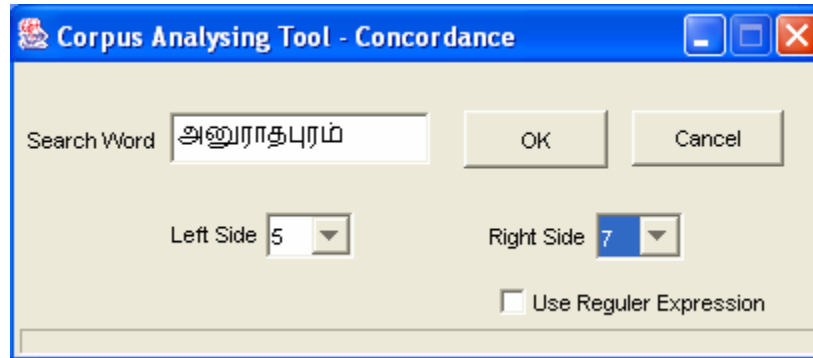


Figure 3.4: Concordance Setting Window

7. **To view the Concordance** then click on the OK button, then the Concordance output will appear on next window.  
The sample Concordance out put as show the following figure 3.5.

The sample Concordance out put as show the following figure 3.5.

[illegible]

Figure 3.5: Sample Concordance Output

8. **To use the CAT for highlight** a particular word/phrase in a text file. User click on the particular text file on the View file pane (N), then a special highlighting window will appear. The following figure 3.6 shows the highlighting window properties.

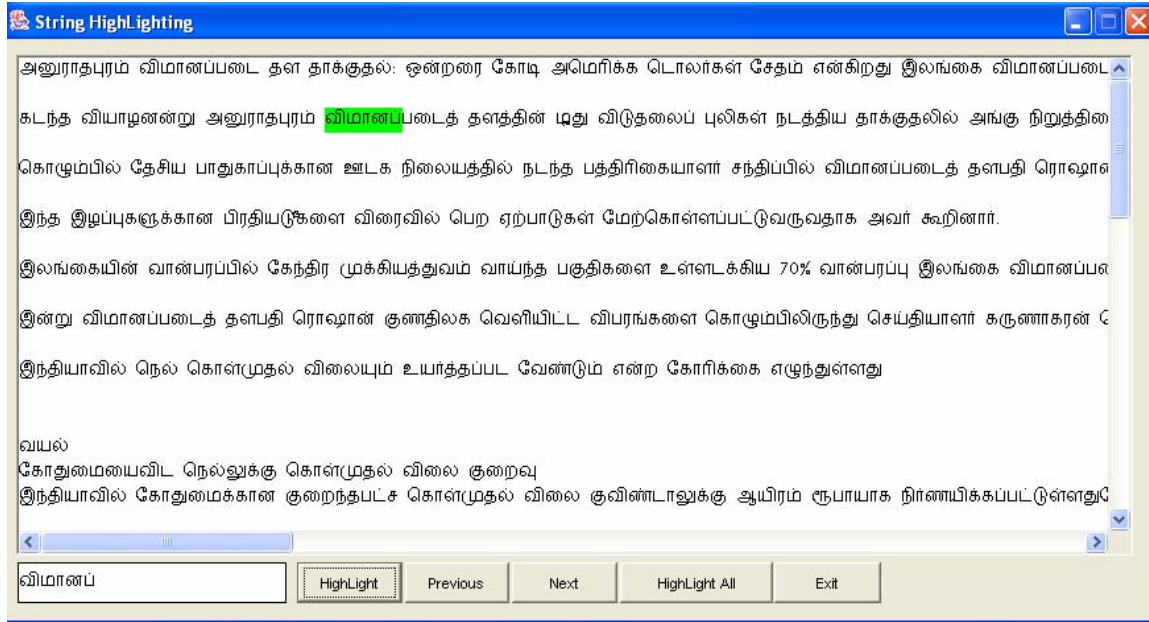


Figure 3.6: Text Highlighting Screen

9. **To merge the files**, to be choose the *Tools* → *Merge Files*. The CAT to be merge files which are available on the File details pane (N) and generate a single file given folder location.
10. **To need conform the loaded files locations** are correct then press the *Availability check* (D) button on the tool bar.
11. **To be exit CAT** then press on the Exit button (M) or choose the *File* → *Exit*.