

ConcGramCore v.1.0.2 download and setup

Created on 18 April 2018

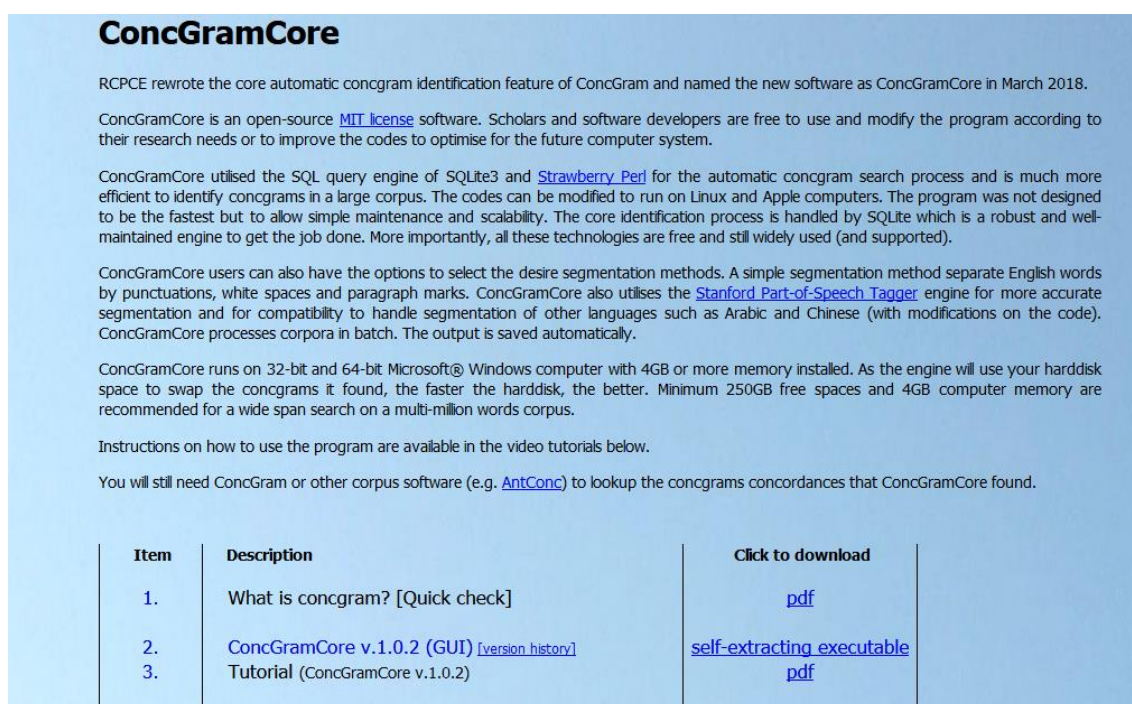
Pre-requisition

Microsoft Windows 7 or above (64-bit version preferred). Suggested hardware: 4GB RAM or above, 50GB+ free hard disk space, Solid-state drive recommended.

Download the program

Visit RCPCE Profession-specific Corpora webpage (<http://rcpce.engl.polyu.edu.hk/CGC/>).

Click the link ‘self-extracting executable’ of the item ‘ConcGramCore v.1.0.x (GUI)’ and download the file.



ConcGramCore

RCPCE rewrote the core automatic concgram identification feature of ConcGram and named the new software as ConcGramCore in March 2018.

ConcGramCore is an open-source [MIT license](#) software. Scholars and software developers are free to use and modify the program according to their research needs or to improve the codes to optimise for the future computer system.

ConcGramCore utilised the SQL query engine of SQLite3 and [Strawberry Perl](#) for the automatic concgram search process and is much more efficient to identify concgrams in a large corpus. The codes can be modified to run on Linux and Apple computers. The program was not designed to be the fastest but to allow simple maintenance and scalability. The core identification process is handled by SQLite which is a robust and well-maintained engine to get the job done. More importantly, all these technologies are free and still widely used (and supported).

ConcGramCore users can also have the options to select the desire segmentation methods. A simple segmentation method separate English words by punctuations, white spaces and paragraph marks. ConcGramCore also utilises the [Stanford Part-of-Speech Tagger](#) engine for more accurate segmentation and for compatibility to handle segmentation of other languages such as Arabic and Chinese (with modifications on the code). ConcGramCore processes corpora in batch. The output is saved automatically.

ConcGramCore runs on 32-bit and 64-bit Microsoft® Windows computer with 4GB or more memory installed. As the engine will use your harddisk space to swap the concgrams it found, the faster the harddisk, the better. Minimum 250GB free spaces and 4GB computer memory are recommended for a wide span search on a multi-million words corpus.

Instructions on how to use the program are available in the video tutorials below.

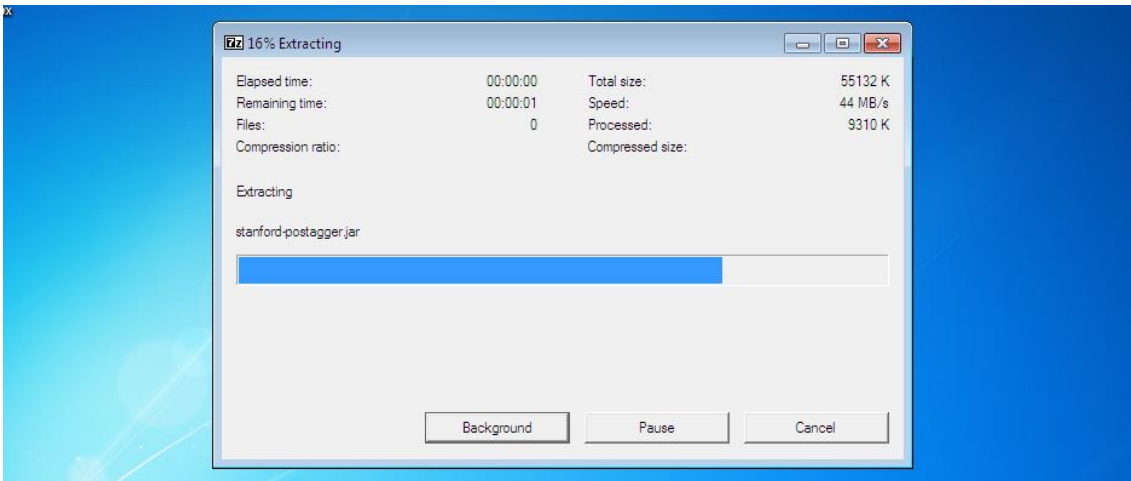
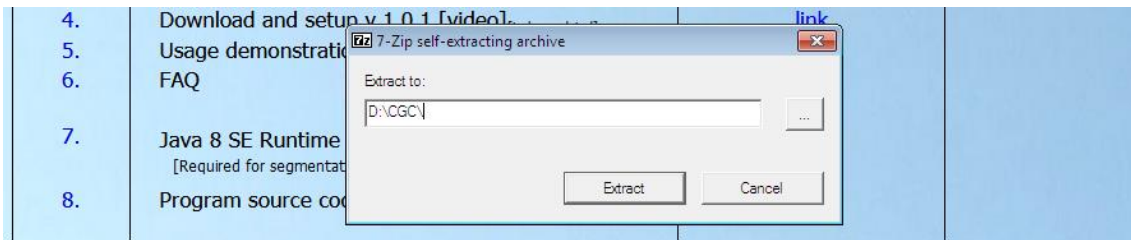
You will still need ConcGram or other corpus software (e.g. [AntConc](#)) to lookup the concgrams concordances that ConcGramCore found.

Item	Description	Click to download
1.	What is concgram? [Quick check]	pdf
2.	ConcGramCore v.1.0.2 (GUI) [version history]	self-extracting executable
3.	Tutorial (ConcGramCore v.1.0.2)	pdf

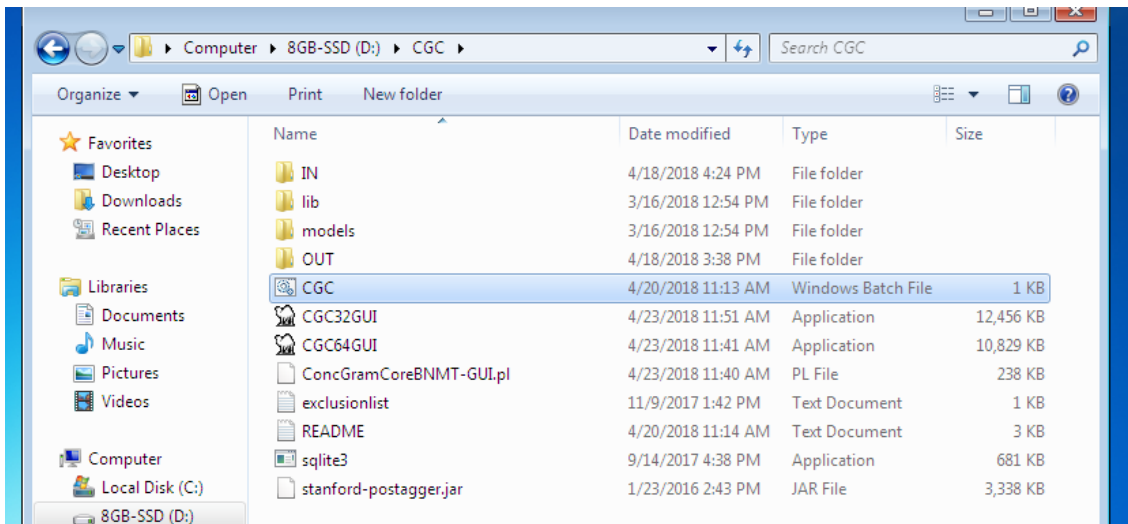
Setup the program

Run the downloaded Windows executable file ‘CGC-1.0.x.exe’.

In the ‘Extract to:’ dialog box, enter a path where the drive will have enough space to run the automatic search. ‘C:’ drive is not recommended because it is shared with Microsoft Windows. In this example, we will save the files in ‘D’ drive. Type ‘D:\CGC\’ in the dialog box and then press Extract button.



After extraction, the ConcGramCore program and its supporting files are saved into the path specified.



Open Windows Explorer and browse to the folder where the program was extracted, and double click the file 'CGC' (type is 'Windows Batch File') to run the program.

If you will use Stanford modules to do the segmentation, please install Java Runtime Engine. The download link of the Java 8 SE Runtime engine can be found on the 'ConcGramCore' webpage.

-END-