

WordStatix - User manual

Version 2.0.0

Copyright © Massimo Nardello, Modena (Italy) 2016 - 2020.

WordStatix is a free and multiplatform software useful to create concordances, that are lists of the words used within a document along with their recurrence and context. The document may be structured in chapters, numbers or in any another way. The software allows to track specific words by prefix or suffix, to skip those which are meaningless (like articles or prepositions) or numbers, to create a simple statistic of the recurrence of all words or of some of them, possibly within the different sections of the document, and to create three kind of diagrams to visualize the statistical data in different ways.

WordStatix is free software and is released under the GPL license version 3 or following. It has been created with Free Pascal and Lazarus (<http://www.lazarus.freepascal.org>).

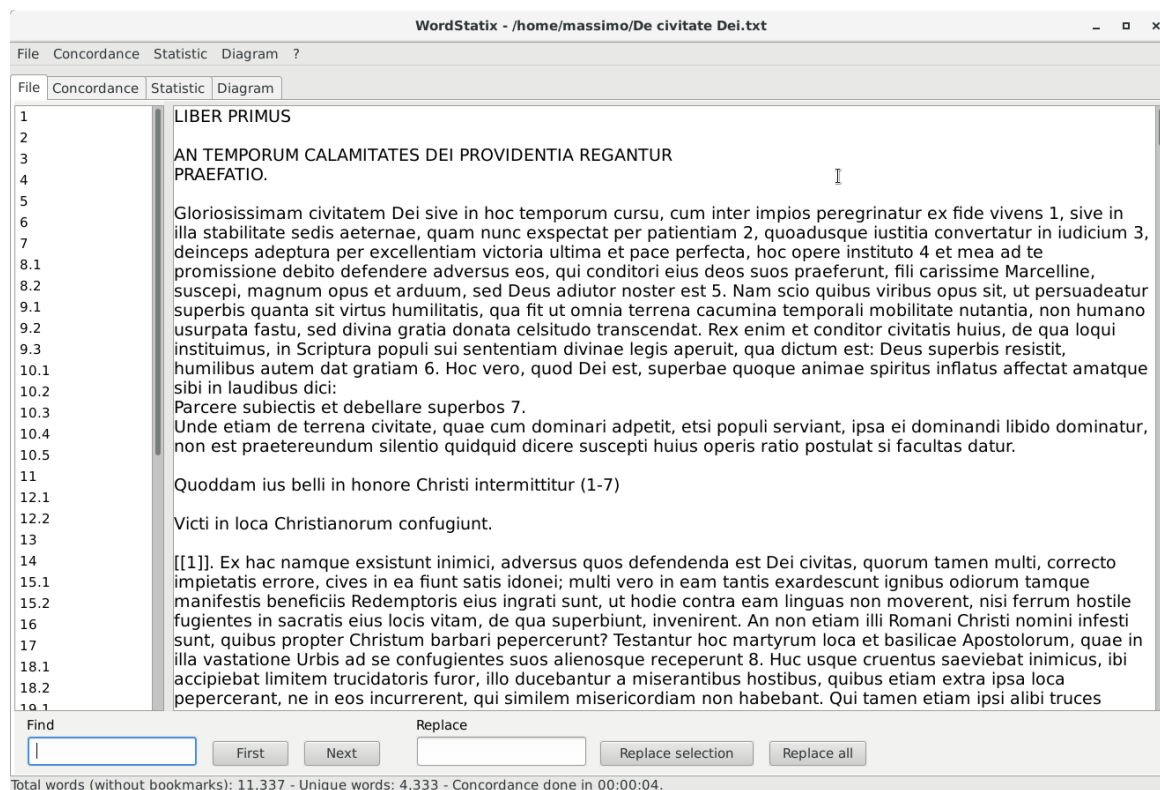
WordStatix runs natively in English, but is translated also in Italian.

The interface of WordStatix is divided in four sections, called “File”, “Concordance”, “Statistic” and “Diagram” which correspond to the main menu items with the same names. The sections can be selected with “Ctrl” + “1”, “Ctrl” + “2”, “Ctrl” + “3” and “Ctrl” + “4”. In the first section it’s possible to manage the document to be analyzed, in the second to create the concordance. in the third to create a statistic of the recurrence of some selected words, possibly in the various sections of the document, and in the fourth to create three kinds of diagrams to visualize the statistical data. Let’s take a look at each one of them.¹

¹ In the MacOS version of WordStatix, the menu shortcuts containing the “Ctrl” key must be activated using the “Cmd” button instead of “Ctrl”. For example, to save the current file the shortcut is “Cmd” + “S” instead of “Ctrl” + “S”.

File section

The File section, with some sample data, looks like this:



The bigger field on the right contains the document to be analyzed. This can be inserted in many ways:

- it can be open or imported with the menu item “File – Open” or with the shortcut “Ctrl” + “O”;
- it can be pasted after having copied from other sources, like a word processor, a browser or a pdf file, with “Ctrl” + “V” or with the popup menu “Paste” available with a right click of the mouse on this field;
- it can be typed directly by the user.

The file formats that WordStatix can actually open or import are the following:

- it can open directly text files, usually with “.txt” extension or with no extension at all, that can be created also with a word processor saving the document in text format;
- it can import Open Document (“.odt”) files, created with LibreOffice Writer or OpenOffice Writer; in this case the footnotes and the endnotes are shown within the text among single square brackets;
- it can import Word files (“.docx”, not “.doc”) files, created with Microsoft Word: in this case the footnotes and the endnotes are shown at the end of the text among single square brackets.

Note that if the user imports a file with “.odt” or “.docx” extension, the name of the file that can be saved within WordStatix will have the “.txt” extension, since the software cannot save in other formats.

The menu item “File – New” is useful to clear everything and start a new document and the related concordance. The menu item “File – Save” or the shortcut “Ctrl” + “S” is useful to save the current document in text format, necessary only to keep it but not to create the concordance. This menu item is enabled only if the text has been modified after having opened or after its last save. The menu item “File – Save as...” is useful to save the file with a different name. The menu item “File – Exit” or the shortcut “Ctrl” + “Q” quits the software.

It’s also possible to clean a text pasted from another source from possible unwanted carriage returns, double spaces and lacking of a space after punctuation marks or before open parentheses . To use this feature, just separate each paragraph with an empty line and press “Ctrl” + “Shift” + “P”. All the lines of text not separated by an empty line will be gathered in one single paragraph, the double spaces will be replaced by a single space, all the punctuation marks will be followed by a space and all the open parentheses will be preceded by a space. Note that punctuation marks and carriage returns are not relevant for the concordance, so they can be added freely by the user.

It’s possible to type a special space between two words which will appear like a space, but will not be considered by the software as the beginning of a new word. This special space is useful to associate two or more words in a single sentence to be considered and processed as one word by the software. To insert this special space type “Ctrl” + “Space”. This feature works also in the fields “Find” and “Replace”, discussed below, so if necessary this space can be easily found and replaced with a normal space.

In Linux and Windows the font size of the text may be enlarged and reduced with “Ctrl” + “Shift” + “+” and “Ctrl” + “Shift” + “-”, or using the mouse wheel while holding the “Ctrl” key. In macOS it’s possible to use the popup menu available with “Ctrl” + “click” on the text.

The list on the left gathers all the bookmarks contained within the document. A bookmark is used to denote the beginning of a new section of the text, like a chapter, a number or the like. It has the following characteristics:

- it’s a single word, separated by the other words by spaces or punctuation marks, or placed at the beginning of a paragraph;
- it doesn’t contain commas, spaces or carriage returns inside it;
- it’s not an asterisk, since this is a special virtual bookmark used to identity the part of a text which precedes the first possible real bookmark;
- it’s contained between double square brackets without spaces inside (like “[chapter1]”);
- it’s unique in the text to be processed, so a bookmark can be used only once.

Anyway, the user doesn’t need to remember all these rules, since the software checks the bookmarks and possibly changes them, removing commas and changing spaces with “_” and asterisks with dots, to make them match these requirements. If a bookmark is used more than once, the user will be warned about it and the bookmark list will be cleared. It will be still possible to create a concordance, but the reference to the bookmarks might be incorrect. To fix this, find the repeated bookmark, possibly using the “Find” function described below, modify all its recurrences to make them unique and update the list of bookmarks with the menu item “File – Update bookmark list” or with the shortcut “Ctrl” + “U”. Likewise, if in the text there are double open square brackets (“[”) not followed by double closed square brackets (“]”) in the same paragraph and within 50 characters, the software will warn the user about that and will not create the bookmark list. It will be still possible to create a concordance, but the reference to the bookmarks might be incorrect. To fix the problem, just look in the text for isolated open double square brackets and remove them.

A bookmarks must be typed by the user within the document before the beginning of a new section, or possibly in any other place if it’s useful to create a structure of the document that is different from the original one. Obviously the bookmarks will not be considered by the software as words to

include in the concordance. It's better to use bookmarks not longer than 10 or 15 characters, just to make them look clearly in the diagrams. Anyway, the maximum length of a bookmark is of 50 characters.

To insert an existing word as bookmark, select it and use the menu item "File – Set Bookmark" or the shortcut "Ctrl" + "R". In this case the list of the bookmarks on the left will be updated automatically. On the contrary, when no word is selected, this menu item will simply insert the open and closed double square brackets ("[]"), and the user will have to type the bookmark inside them. In this case the bookmark list will not be updated automatically. This should be done manually with the menu item "File – Update bookmark list" or with the shortcut "Ctrl" + "U". Anyway, the bookmark list will be update automatically also just before creating a concordance.

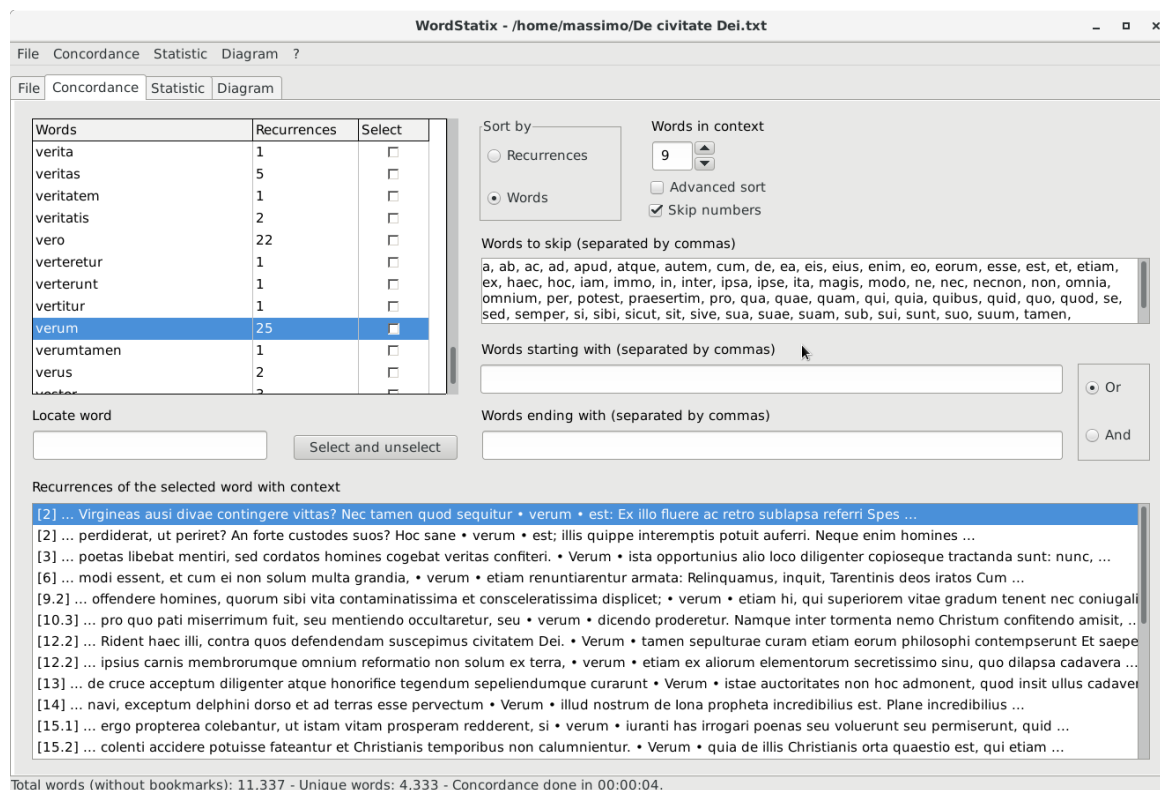
A click on a bookmark in the bookmark list selects the corresponding bookmark in the text. With this option it's possible to check very easily that all the bookmarks are in the right place before creating the concordance.

At the bottom of the section there are fields and buttons to find a text within the document in use and to replace the text found with another text. To use this feature, fill the field "Find" with the text to look for, then click on the button "First" to select its first recurrence and on the button "Next" to select its next recurrences. To replace the text that has been found with another text, fill the field "Replace" with the text to replace, click on "Find" to select the first occurrence and then click on "Replace selection" to replace the word just selected, or "Replace all" to replace all the recurrences of the searched text. In the first case, after having replaced the selection, the software will find the next word to be found and will selected it automatically. The search and the replace are case insensitive, so the software will consider lowercase and uppercase characters as the same. Note that the software will search and replace a text also if it's within a word and not only if it match a whole word. So use the "Replace all" function with care, to avoid to replace wrong part of the text. To insert this special space in these fields, that gathers two or more words so that the software may consider them as a single word, type "Ctrl" + "Space".

The menu items "File – Save concordance..." or the shortcut "Ctrl" + "Shift" + "S" are useful to save in a file the current text along with its concordance. With this feature the user may reload them together without the need to load the text alone and then recreate the concordance again, which may take some time. The text and the concordance are saved in a text file whose extension is ".wsx", just to distinguish it more easily from the other files present in the disk. Along with the text and the concordance, are saved also all the settings used to create the concordance, discussed later, namely the values of the fields and check boxes "Sort by", "Advanced sort", "Skip numbers", "Words to skip (separated by commas)", "Word starting with (separated by commas)" and "Word ending with (separated by commas)" along with their "And" or "Or" condition. Note that if the text or the settings useful to create a concordance, not just to sort its words, have been changed after processing it, it's necessary to recreate the concordance itself before saving it along with the text. In this way the user may be sure that in a ".wsx" file a concordance is perfectly aligned with the text and with the settings used to create it. If a word or a recurrence of the concordance have been deleted with the functionalities discussed later, they are still saved in the file of the concordance. The menu item "File – Open concordance..." or the shortcut "Ctrl" + "Shift" + "O" are useful to open a ".wsx" file in order to load a text along with its concordance. All the settings saved in the file will overwrite the current ones. So, if the concordance is immediately recreated after having loaded in a ".wsx" file, it will be perfectly identical to the one contained in the file.

Concordance section

The Concordance section, with some sample data, looks like this:



The top left grid shows the list of the words of the concordance and their recurrence in the document in use. To create the concordance and fill this grid use the menu item “Concordance – Create concordance” or the shortcut “Ctrl” + “N”. The software creates the list of the words of the concordance, sorts them according to the value of the field “Sort by” and then fills the concordance grid. This second and third procedures may be activated also later, while maintaining the same concordance (that is, the same internal list of words managed by the software), with the menu item “Concordance – Refresh concordance grid” or with the shortcut vb“Ctrl” + “Shift” + “U”. To interrupt the creation of the concordance already started or the sorting procedure, use the shortcut “Ctrl”+ “Shift” + “H”, as indicated also in the status bar at the bottom of the form when the concordance process is activated. While creating the concordance, which may take some time, no other section of the software may be selected, and many menu items are not enabled, but the user may use other application while leaving WordStatix complete its work. While the process goes on, the status bar reports the number of words analyzed, and at the end of the procedure also the time spent to complete it. The user will be informed by a message when the concordance will be ready. The words of the concordance are sorted by words of by number of recurrences, according to the value of the field “Sort by”, which is remembered by the software. If a concordance is sorted by number of recurrences, the words that have the same number of recurrence will be sorted also by their name. Changing the value of the field “Sort by” the sorting will not be updated correspondingly. To have the concordance properly resorted, it’s necessary to recreate the concordance or to update only the concordance grid using the menu item “Concordance – Refresh concordance grid” or shortcut “Ctrl”+ “Shift” + “U”. This feature will not recreate the concordance, but will just fill again the concordance grid with its words, a procedure which is faster that the first one. Note that in this way all the words and recurrences that have been possibly deleted with the menu items “Concordance – Add current word to skip list”, “Concordance – Remove current word”

and “Concordance – Delete selected recurrences”, which are described below, will be restored. Anyway, before the concordance grid is refreshed, if some words or recurrences have been deleted, the user is requested to confirm the update and the consequent loss of the changes made. Furthermore, the statistic, and also the diagrams which depends on it, can be sorted independently from the concordance, as will be discussed below.

The option “Advanced sort” makes the software sort properly special characters, like accented ones. Without this options, the sort is much faster, but these special characters are sorted after the more common ones. If this doesn’t matter, leave this option unchecked. The option “Skip numbers” makes the software skip the words that contain only numbers and punctuation marks.

In the field “Words to skip (separated by commas)” the user may type some words, separated by commas, that should not enter in the concordance since they are meaningless, like articles and propositions. This filter is not case sensitive, so the software will not distinguish among lowercase and uppercase characters. It’s possible to type the words to skip directly in this list, or to add to it a word already present in the concordance grid with the menu item “Concordance – Add current word to skip list” or with the shortcut “Ctrl” + “Shift” + “K”. In this way, that word will be also removed from the concordance. To remove just the word from the grid without inserting it in the skip list, use the menu item “Concordance – Remove current word” or the shortcut “Ctrl” + “K”. To restore the deleted words and recurrences, use the menu item “Concordance – Refresh concordance grid” or shortcut “Ctrl”+ “Shift” + “U”. Note that to change the sort order of the words it’s necessary to update the concordance grid and loose all the deletions that have been made.

The skip list will be remembered by the software. Furthermore, it’s possible to save it in a text file with the menu item “Concordance – Save skip list...”, and open an existing skip list file with the menu item “Concordance – Open skip list...”. In this way the user may keep different skip lists corresponding to different languages or various kinds of documents to analyze.

In the field “Words starting with (separated by commas)” it’s possible to type one or more prefix of words separated by commas. When creating the concordance, only the words beginning with one of those prefix will be included. This filter is not case sensitive, so the software will not distinguish among lowercase and uppercase characters.

In the field “Words ending with (separated by commas)” it’s possible to type one or more suffix of words separated by commas. When creating the concordance, only the words ending with one of those suffix will be shown. The filter is not case sensitive, so the software will not distinguish among lowercase and uppercase characters.

At the right of the fields “Words starting with (separated by commas)” and “Words ending with (separated by commas)” there is an option box useful to specify the condition of the filter performed with these two fields, provided that both of them contain some text. If the condition is “Or”, all the words beginning with one of the prefixes inserted in the field “Words starting with (separated by commas)” or ending with one of the suffixes inserted in the field “Words ending with (separated by commas)” will be included in the concordance. If the condition is “And”, only the words that begin with one of the items inserted in the field “Words starting with (separated by commas)” and at the same time end with one of the items inserted in the field “Words ending with (separated by commas)” will be included in the concordance.

When the user adds new words to the fields “Words to skip (separated by commas)”, “Words starting with (separated by commas)” and “Words ending with (separated by commas)”, the software sorts them automatically, so that they can be seen more easily.

Note that if the options “Sort by” or “Advanced sort” are changed, there’s no need to recreate the concordance from scratch to have it corresponding to these options, but it’s enough to sort again the words of the existing concordance with the menu item “Concordance – Refresh concordance grid” or with the shortcut “Ctrl”+ “Shift” + “U”. On the contrary, if the option “Skip numbers” or the fields “Words to skip (separated by commas)”, “Words starting with (separated by commas)” and

“Words ending with (separated by commas)” are changed, it’s necessary to recreate the concordance to have it corresponding to these options.

When a word is selected in the concordance grid, in the bottom list “Recurrence of the selected word with context” is shown each single recurrence of that word in the document analyzed. Each recurrence will be reported in a single row of the list along with its context, that are the words before and after it in the document analyzed. The number of words before and after the selected one are determined by the field “Words in context”, whose value can be changed by the user and that will be remembered by the software. At the beginning of each recurrence, among single square brackets, is shown the last bookmark before that recurrence, to indicate the section of the document in which it’s located. If no bookmark is present, it will be shown an asterisk, that is the mentioned virtual bookmark.

A single recurrence of a word in the bottom grid “Recurrence of the selected word with context” can be removed, if for some reason it’s not to be included in the concordance nor computed in the statistic and in the diagrams. To delete a recurrence, use the menu item “Concordance – Delete selected recurrences” or the shortcut “Ctrl” + “Shift” + “D”. To restore the removed recurrences and words, just refresh the concordance grid with the menu item “Concordance – Refresh concordance grid” or with the shortcut “Ctrl”+ “Shift” + “U”.

The words in the concordance grid may also be individually selected with a click on the check box in the column “Selected” or with the space bar when the grid is focused (click with the mouse on it to focus it). Then it’s possible to show only the words that are selected in this way with the menu item “Concordance – Show only selected words” or with the shortcut “Ctrl” + “L”. When this filter is enabled, using the same menu item or shortcut remove the filter and show all the words.

This selection is necessary to create the statistic and the diagrams, described below, but also to move the recurrences of some words to those of another word. This feature is useful to associate together the recurrences of different words that should be considered a single item in the statistic and in the diagrams. The menu item “Concordance – Associate selected words” or the shortcut “Ctrl” + “J” are useful to remove all the selected words and to move their recurrences to the current word. Just select the words whose recurrences must be moved checking the “Select” check box in the concordance grid, move on the word that should receive them, so that its recurrences are shown in the bottom list “Recurrence of the selected word with context”, and then activate this functionality. The recurrences of the current word will contain also those of the the selected words, after those already existing, while the selected words will be deleted. To restore the recurrences and the deleted words, just refresh the concordance grid with the menu item “Concordance – Refresh concordance grid” or with the shortcut “Ctrl”+ “Shift” + “U”.

To locate a word in the concordance grid type its first letters in the field “Locate word” and press “Return”. In this way the first occurrence of the word beginning with the typed letters will be selected. To select the next recurrences, press “Ctrl” + “Return”. The search is case insensitive. To select and deselect alternatively all the words in the concordance grid, use the button “Select and unselect”.

It’s possible to save the list of the words and of the recurrences created by a concordance in a report. This could be saved as a text file or as an HTML file with the menu item “Concordance – Save report...” or with the shortcut “Ctrl” + “Shift” + “S”. To save the report in HTML format, add the extension “.html” (not just “.htm”) to the file to be saved. If the extension is different, the report will be saved in text format. The HTML file may be open in a browser or with a word processor. The single words will be formatted as “Heading 2” paragraphs, to be tracked and managed more easily. Before saving the file of the report the software asks for a possible title to be inserted at the beginning of it.

Note that the only the words that are actually visible in the concordance grid will be saved in the file. So, if the menu item “Concordance – Show only selected words” is checked, only the selected words will be saved. Also the possible deleted words and recurrences will be excluded from the file.

The concordance is not case sensitive, so lowercase and uppercase characters will be considered as the same and printed always in lower case. On the contrary, the list of the recurrences is case sensitive, so each word is printed as it is in the original text. If only lowercase or uppercase recurrences of a word must be considered, just delete the recurrences that contain the unwanted version of the word.

Statistic section

The Statistic section, with some sample data, looks like this:

	Total	*	1	2	3	4	5	6	7	8.1	8.2	9.1	9.2	9.3	10.1	10.2	10.3
dei	26	3	1	0	0	1	0	0	0	7	0	1	0	0	0	1	0
deo	12	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0
deorum	13	0	1	1	0	4	1	3	0	0	0	0	0	0	0	0	0
deorumque	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
deos	22	1	0	0	5	1	0	1	0	0	0	0	0	0	0	0	0
deum	12	0	0	0	0	0	0	0	0	0	1	0	0	1	2	0	0
deus	17	2	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0
Total	103	6	3	1	5	6	1	4	0	7	3	3	0	2	2	2	0

Total words (without bookmarks): 11,337 - Unique words: 105 - Concordance done in 00:00:02.

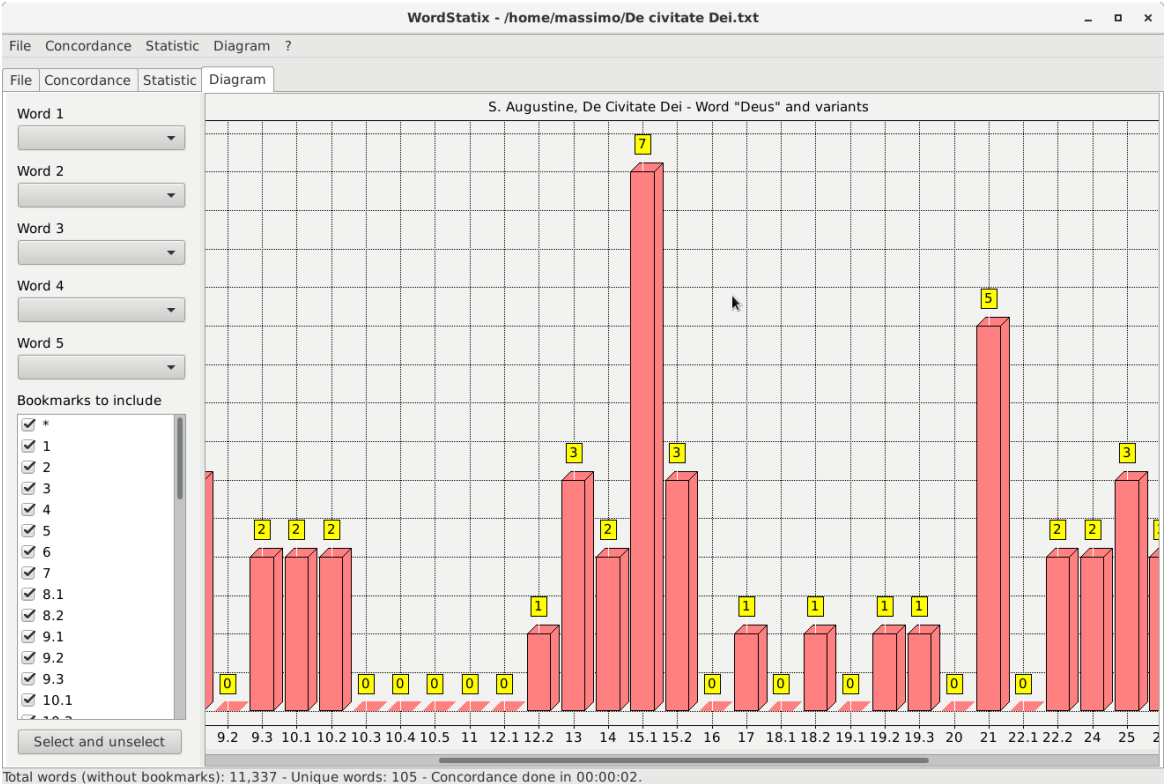
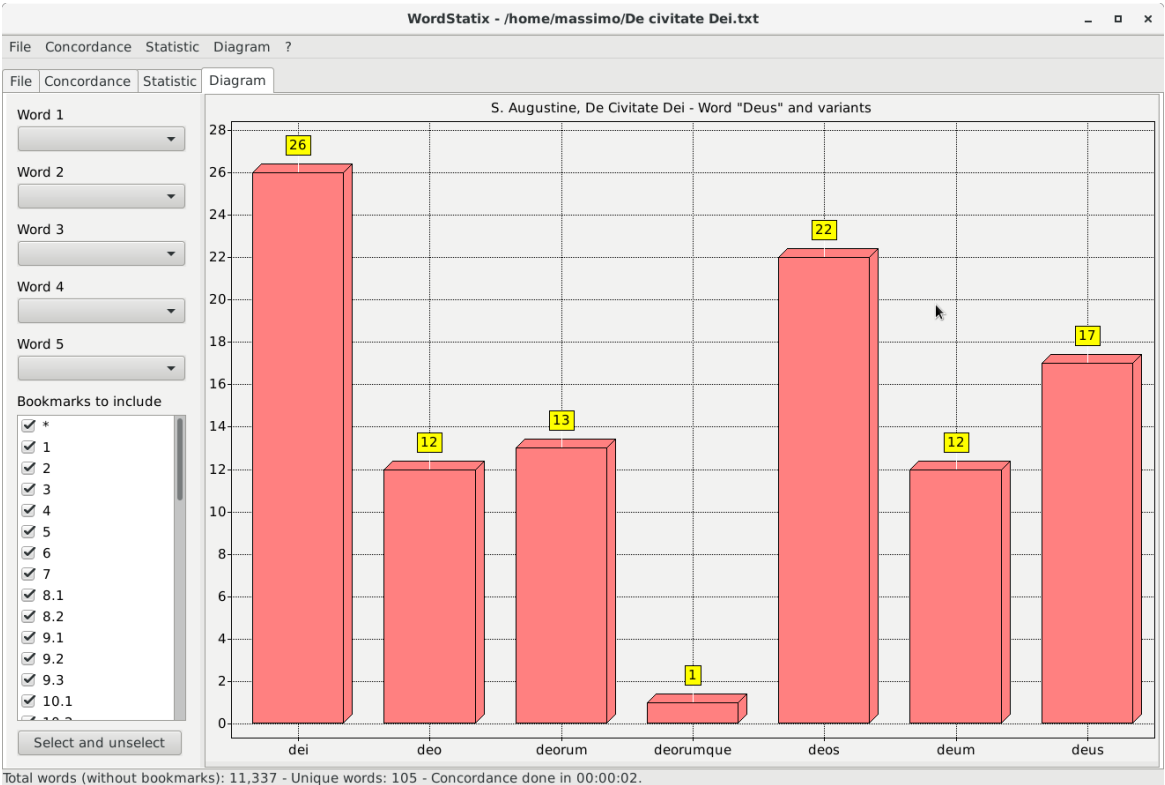
In this section it's possible to show in a grid the recurrences of the words selected in the concordance grid. The software will show both their total number in the second column of the grid (under the label "Total") and their recurrence within each section of the document, identified by the various bookmarks, in the other columns of the grid. The words are shown in the first left column, while the bookmarks are shown in the top row. The words that precede the first possible bookmark are gathered under the "*" virtual bookmark.

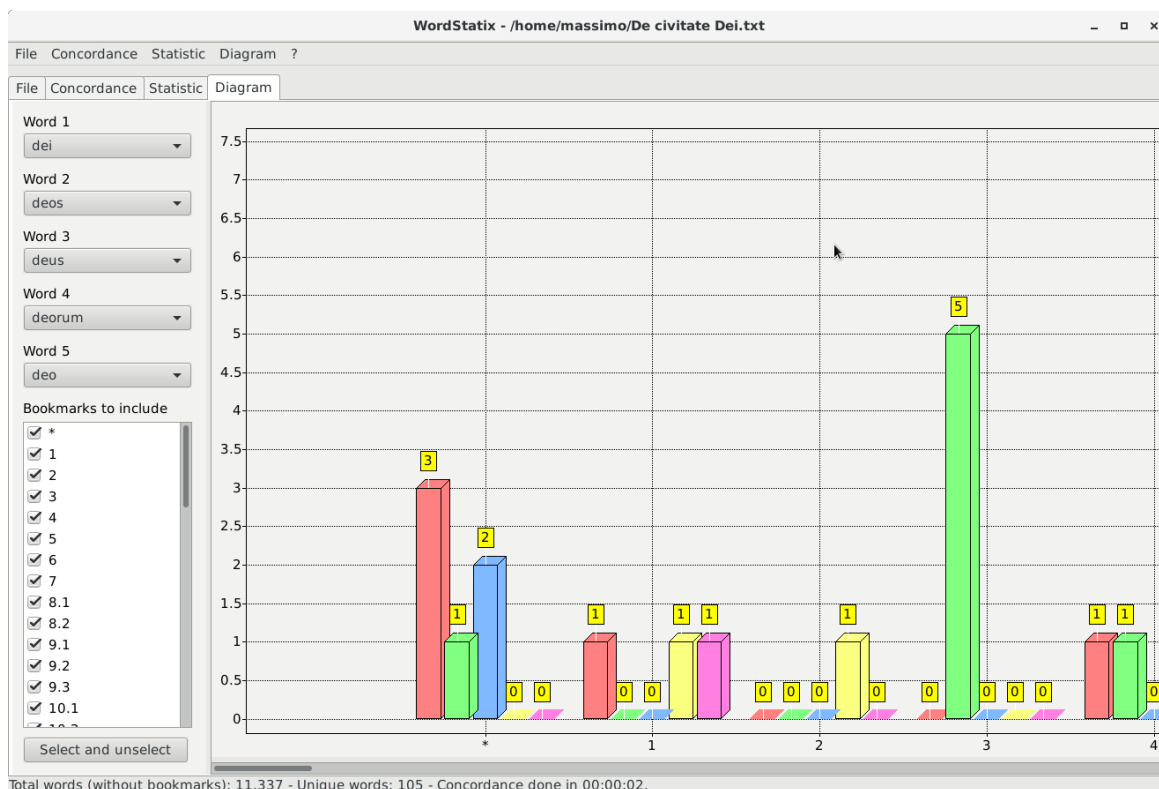
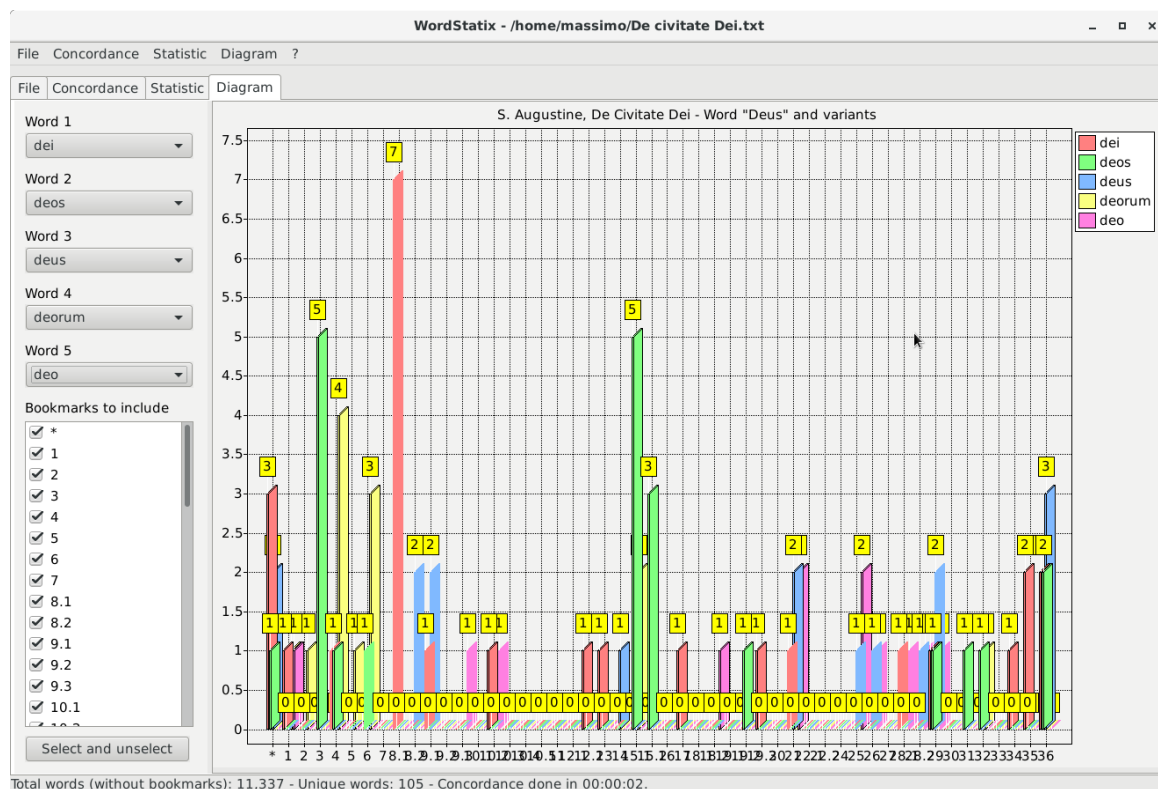
To create or update the statistic, use the menu item "Statistic – Create statistic" or the shortcut "Ctrl" + "T". When a statistic is created, it's automatically sorted by recurrences or by the name of the words just like the concordances. To sort the statistic on a different field, use the menu items "Statistic – Sort by name" and "Statistic – Sort by recurrences". The sorting of the concordance will not be modified. Note that when the words are sorted by recurrence, those which have the same recurrence will be sorted also by name. If the option "Advanced sort" in the Concordance section is checked, the sort of the statistic on the name of the words will order properly also the special characters. Otherwise they will be sorted after the more common ones.

This series of values can be exported in a CSV file, which can be easily open by a spreadsheet like Calc or Excel. To export the current statistic in a file of this kind, use the menu item "Statistic – Save statistics...".

Diagram section

The Diagram section, with some sample data, looks like this:





In the Diagram section it's possible to visualize in a diagram the numeric data of the statistic, with three different kinds of diagrams.

1. A diagram of the total recurrences of the words of the statistic not gathered in bookmarks (first picture above). The values are taken from the rows of the various words following the second

column with the label “Total”. To create this kind of diagram, obviously after having created a statistic, use the menu item “Diagram – Total words without bookmarks” or the shortcut “Ctrl” + “I”. At the bottom of the diagram are shown the names of the different words, without any mention of the bookmarks, and the height of the various bars indicates the total of their recurrences in the statistic. Only the words contained in the bookmarks selected in the list “Bookmarks to include” will be taken into consideration to create the diagram, even if the bookmarks will not be printed in the diagram. So, in order to exclude the words contained in one or more bookmarks deselect their names in this list with the mouse or the space bar. The button “Select and unselect” allows to select and unselect alternately all the bookmarks in this list. For this kind of diagram, the fields above the list of bookmarks are useless.

2. A diagram of the total recurrences of the words of the statistic gathered within their bookmarks (second picture above). The values are taken from the last row of the statistic grid, at the right of the label “Total”. To create this kind of diagram, use the menu item “Diagram – Total words with bookmarks” or the shortcut “Ctrl” + “Shift” + “I”. At the bottom of the diagram are shown the names of the various bookmarks (so chapters, numbers, etc.), and the height of the various bars indicates the total recurrences of all the words used in the statistic for each bookmark. The words that precedes the first possible bookmark will be gathered under the “*” virtual bookmark. Only the words contained in the bookmarks selected in the list “Bookmarks to include” will be taken into consideration to create the diagrams. So, in order to exclude the words contained in one or more bookmarks deselect their names in this list with the mouse or the space bar. The button “Select and unselect” allows to select and unselect alternately all the bookmarks in this list. For this kind of diagram, the fields above the list of bookmarks are useless.

3. A diagram of the total recurrences of up to five words among those used in the statistic gathered within their bookmarks (third picture above). To create the third kind of diagram, select at maximum five words in the fields “Word 1”, “Word 2”, “Word 3”, “Word 4” and “Word 5” at the left of the section. Then use the menu item “Diagram – Single words with bookmarks” or the shortcut “Ctrl” + “Shift” + “Alt” + “I”. At the bottom of the diagram are shown the names of the various bookmarks (so chapters, numbers, etc.), while the height of the various bars indicates the recurrences of each one of the selected words for each bookmark. The words that precedes the first possible bookmark will be gathered under the “*” virtual bookmark. The different colors of the bars are related to the selected words, as indicated in the labels at the right top of the diagram. Note that to create this kind of diagram the field “Word 1” must contain a word, and the following fields must be compiled contiguously, so that no field above a compiled field is empty. To remove any word from one of these fields, except from the first one, select it and press the “Del” or the “Backspace” key. All the fields will be cleared when a new statistic is created. Only the words contained in the bookmarks selected in the list “Bookmarks to include” will be taken into consideration to create the diagrams. So, in order to exclude the words contained in one or more bookmarks, deselect their names in this list with the mouse or the space bar. The button “Select and unselect” allows to select and unselect alternately all the bookmarks of this list.

Before creating a diagram, the user is requested to insert a title, that will be printed at the top of it. If no title is provided, no title will be printed. Each diagram may be zoomed in width, not in height, with the menu item “Diagram – Zoom in” or with the shortcut “Ctrl” + “+” to see more clearly the various items and values (forth picture above). It can be zoomed out with the menu item “Diagram – Zoom out” or with the shortcut “Ctrl” + “-”. The normal width of the diagram, which corresponds to the width of the form of the software, can be restored with the menu item “Diagram – Normal width” or with the shortcut “Ctrl” + “0”. Note that when the form of the software is resized, the

diagram will be resized as well to its width. When the diagram is zoomed, it can be scrolled toward left or right dragging the bar at the bottom with the mouse or with the left and right arrow keys. It's possible also to zoom a part of the diagram holding the mouse button and dragging it right and down, drawing a box. To restore the normal zoom, just click on the diagram.

The menu item "Diagram – Show values" allows to show or to hide the values of the bars of the diagram, which are shown in yellow boxes. The menu item "Diagram – Show grid" allows to show or to hide the grid of the diagram and the left labels with the scale of values of the bars. These two options are mutually exclusive, because if there were neither left labels nor values marks, it would be impossible to understand the values indicated by the bars.

The menu item "Diagram – Save diagram" allows to save the actual diagram as a picture in ".jpeg" or ".png" format. Note that the diagram will be saved with the width corresponding to the actual zoom level, and with or without the grid and the yellow boxes, according to the state of the menu item "Diagram – Show values" and "Diagram – Show grid". On the contrary, the height of the diagram will be always of 1000 pixels.