

# Searchbot Help

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Searching with the Searchbot</b>	<b>3</b>
2.1	Perform a query . . . . .	3
2.2	Output file . . . . .	3
<b>3</b>	<b>Interface</b>	<b>4</b>
3.1	Textbox . . . . .	4
3.2	Query field . . . . .	5
3.3	Buttons . . . . .	5
3.3.1	Select database & output folders . . . . .	5
3.3.2	Send query . . . . .	5
3.4	Drop down menus . . . . .	5
3.4.1	Type of research . . . . .	5
3.4.2	Appearance mode . . . . .	6
<b>4</b>	<b>Annex</b>	<b>7</b>

## 1 Introduction

Welcome to the Searchbot help document. This file will explain how to install the Searchbot and will teach you all its functionalities. If you encounter an issue, or if you have suggestions to make the Searchbot even better, feel free to open an issue on the Github. If you liked this tool, please give the repository a star, so that others can appreciate it too!

## 2 Searching with the Searchbot

This tool allows you to look for a keyword in all PDFs files located in a set of directories called the database. It will give 2 outputs: the result of the search printed on the console, and a PDF file containing the result of the search.

### 2.1 Perform a query

Below is the step-by-step guide on how to perform a search using this tool. If you need more information about an element, please refer to the corresponding section further in this document. Steps 1 to 3 are interchangeable.

1. Enter a keyword in the query field.
2. Click the "Select database & output folder" button to choose respectively the folders in which the database is, and the output will be created in.
3. Choose the type of search you want to perform (Synonyms/Root search)
4. Click on the "Send query" button

If all the previous steps have been made correctly, the sentence "Looking for KEYWORD in the database, please wait..." will appear on the console. A window showing the progress of the search process will be created and will close by itself when the search is complete. A messagebox will then inform you that the output file has been created and saved, and will give you its name and location. Once the window is closed, the result of the search is displayed on the console, with the keyword colored to facilitate the reading.

### 2.2 Output file

As mentioned above, the result of the search is saved into a PDF file, at the location you gave before performing the search. This file is composed of all the pages of the documents from the database in which the keyword appears in. Moreover, all the occurrences of the keyword will be highlighted, no matter the text orientation or the context (table, image,...).

To keep track of the origin of each page, two marks have been used: first, documents have been separated by a page on which the name of the new files the pages have been extracted from. Second, a watermark has been inserted on the top left corner of each page, with the name of the file it is from, and the page number. You should then be able to know where each page comes from, making further researches in specific files easier.

### 3 Interface

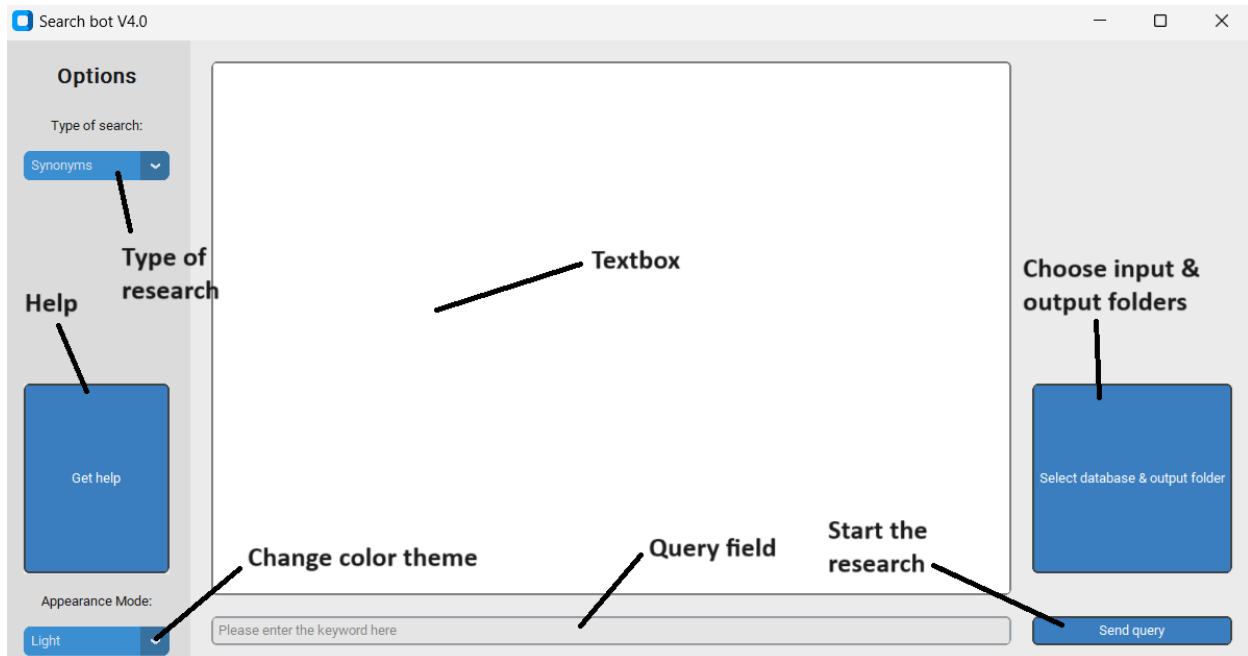


Figure 1: Anatomy of the searchbot

The interface displays several widgets, which will be explained one by one, so that you can have a full comprehension of the tool.

- The textbox
- Query field
- "Select database & output folders" button
- "Send query" button
- "Appearance mode" drop down menu
- "Type of search drop down menu

#### 3.1 Textbox

The textbox is the zone where the result of the search will be displayed. You cannot write or erase something inside, only the program can. This is to make sure that you do not modify the displayed result by mistake. You can still copy text from it and paste it elsewhere.

Please note that it does not display tables nor images. If you need these, please take a look inside the output file.

## 3.2 Query field

In this field, you can type the keyword you want the Searchbot to look for. It does not take casing into account, so do not worry about this. Special characters (like accents) are however needed, as the tool does not perform a normalization of the text or the keyword.

## 3.3 Buttons

### 3.3.1 Select database & output folders

This button will ask you to choose two folders, the first one for the database, the second one for the location in which the output file will be created.

The first folder should be the first folder in your database hierarchy. Indeed, the bot will open all subfolders and search in their PDFs files as well, so make sure that you do not have a folder higher in the hierarchy that you want to search in when you choose the database folder.

The second folder you will be asked to choose will contain the output file described in [2.2](#). If you do not choose a folder for the output file, the file will be saved by default in the database folder given previously.

### 3.3.2 Send query

This button should be clicked on when the first 3 steps, detailed in [2.1](#), have been completed. The Searchbot will then proceed to the research of your keyword in your database. You can activate the button using the "Enter" key

## 3.4 Drop down menus

### 3.4.1 Type of research

This option is at the core of the Searchbot, as it allows you to choose which kind of search you want to perform, based on the keyword you are looking for. You have two choices:

- Search using synonyms.
- Search using the root of the keyword.

The search using synonyms means that the Searchbot will search in the database your keyword and all its synonyms the tool found. If you are not sure about the word you want to find, this is the best option to choose. Please note that there is one major limitation: you will have no control over the synonyms found by the tool, which means that it may end up looking for words which are indeed synonyms, but are not related to the search you want to perform.

The root search is different, as it will take the root of your keyword, and look for words sharing the same root in your database. This is to be used when you know the word you are looking for, but you would also need all its variations.

### **3.4.2 Appearance mode**

This drop down menu is used to change the color theme of the tool. It has 3 possibilities: Light, System and Dark. The chosen color theme will remain the same until changed, and will be used for all windows of the Searchbot (for example, the progress bar window).

## 4 Annex

The list of external libraries used by the tool:

- Fitz: PDF handling
- CustomTkinter: graphical tools, messageboxes
- Webbrowser: opening this file in the browser
- nltk: NLP functions