Textpresso User Guide

[Simple Retrieval, 593K]

Simple Retrieval

The **Simple Retrieval** search engine is the most basic of the searches *Textpresso* offers. It is designed to combine a keyword search with up to four ontology category searches. The output of the search is displayed on a " **Summary Page** ", which summarizes details of the publications which contain matches. The user selects which matches they would like to view and the results are shown on the " **Results Page** ". Additionally, the user can perform an "author search" and a publication "year search" using the Simple Retrieval tool.

Simple Retrieval search features

1. Search Buttons

Clicking on the "Search!" button initiates a search with the inputed parameters. There are two additional buttons. The "Load last query!" button is a useful feature that remembers the last user query and loads it again. The "Undo last changes!" button allows the user to conveniently undo the last change they did to their search.

2. Text Corpus Selection and Author/Year Search

The user has the option to chose to search any combination of the "Titles", "Abstracts" and "Papers" by selecting the boxes beside these options. The user also can perform a search for a particular author by selecting the "Author" option and typing the author name in the search field. Likewise, the user can search for publications from a particular year by selecting the "Year" option and typing a four digit year in the search field.

3. Search Field

This is where keywords to be searched are entered, for example, "let–60", "dumpy" or "1999". Multiple entries are separated with a comma, for example, "lin–3, BW136". The wildcard insertion (**∗**) is used by default after each word. For example, if you type "egl" in the search box, the return would include matches with egl, egl–1, egl–2, egl–30 etc. The user can select the exact match box to turn off the wild card insertion. Note: the keyword search is *not* case sensitive.

4. Ontology Category Menu

Above the search field box are drop down menu's where the user can specify up to four different *Textpresso* Ontology categories with which to search. Note: each ontology category will only be searched for once, even if the user selects the same ontology category more than once.

5. Sentence Search vs Publication Search

The user has the option to search the input parameters within individual sentences or entire publications. Note: this only applies when "Abstracts" and/or "Papers" are selected to search against.

Search Summary Page Features

1. Matches Information

This is where the number of matches of the search parameters is displayed. If the system is searching either sentences or publications, the total number of sentences containing the search parameters is displayed. Also shown is the total number of publications that contain one or more hits.

2. Summary Display Controls

The summary page is returned with ten summaries displayed per page (this is the default, the number of results summaries displayed per page can be customized in the " **Customization** " page). The summary display controls allow the user to display any page by selecting the page from the drop down menu and pressing the "Display" button. Alternatively, the user can navigate the summary page using the "Previous" and "Next" buttons.

3. Email Settings

The user can opt to have the summary page sent to them via email. To do this the user must enter their email address in the text box and press the "E–mail" button. By selecting the include matches option, the email will also contain the resulting matches from the search. **Beware, this can result in very large emails!**

4. "View all matches" Button

Clicking this button brings the user to a results page containing all the results for a given search.

5. "Abstract Expansion" Buttons

By default (and for the sake of clarity) only the first two sentences of an abstract are displayed in the "Abstract" column. Pressing the "Expand abstract" button will display the full abstract in the column. This full abstracts can be collapsed again by pressing the "Collapse abstract" button.

6. "View matches" Button

Clicking this button brings the user to a results page containing the results for a that particular publication.

7. "PDF" Button

Clicking on the PDF button displays the pdf version of that publication *(only available to Caltech users)* .

8. "Related articles" Button

This button outlinks to the PubMed web–site page of citations that are related to that particular publication.

9. "Results in PDF" Button

Clicking on the Results in PDF button brings the user to a web page where they can opt to download all the resulting hits for their query in PDF format. *(This may take a few minutes, depending on the number of resulting hits)*

to the top

[Description]   [Contents]   [Advanced Retrieval]

Result Page Features

1. Query Display

At the top of the results page the search query is displayed. Note that the different search parameters are displayed with different color as a visual aid.

2. Results Display

The result display is returned with ten matches displayed per page (this is the default, the number of matches displayed per page can be customized in the " **Customization** " page). The summary display controls allow the user to display any page by selecting the page from the drop down menu and pressing the "Display" button. Alternatively, the user can navigate the summary page using the "Previous" and "Next" buttons.

3. Publication Identifier

The "File ID" identifies the publication from which the match comes according to WormBase abstract nomenclature. The type of publication is displayed in parenthesis after the File ID, i.e. Abstract.

4. Sentence Identifier

The "Sentence ID" specifies the sentence number of the match in the publication.

5. Search Matches

The matching sentences are displayed in boldface font. For the sake of context, the sentences that surround the match in the publication may also be displayed (the default number is ten, the number of surrounding sentences displayed per match can be customized in the " **Customization** " page).

6. Links to Wormbase

Some words and terms in the matching sentences will link to their corresponding report pages in Wormbase, a database repository for the biology and genome of *C. elegans*.

to the top

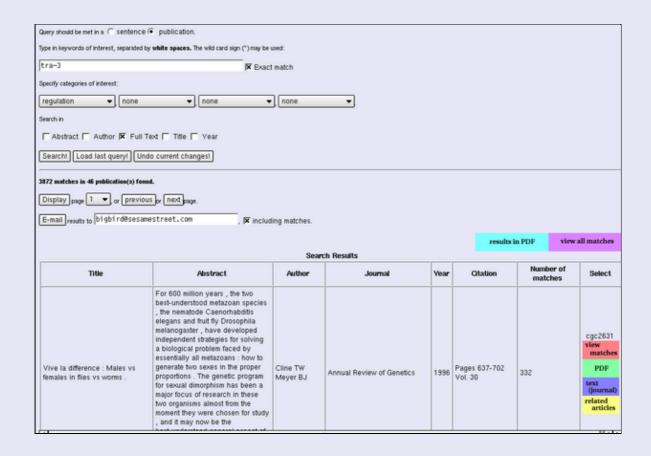[Description]   [Contents]   [Advanced Retrieval]

Performing a Text Search

Below is an example of a text search using the Simple Retrieval tool in Textpresso. Please see the " Examples " section for many more examples of Textpresso searches.

The Documentation Page for the Textpresso Information Extraction System

The user first selects whether they would like to search for the terms in entire publications or the individual sentences of the publications. In this instance, the "publications" option was selected. The user then enters the parameters for the text search. In this case, the user has chosen one ontology class, "regulation", and entered the keyword "tra−3". The user must then decide whether they wish to search for these parameters in publication titles, abstracts or papers (where more than one of these options can be selected). In this case the user has opted to search papers. The user can modify the search parameters with the "Load last query" and/or the "Undo last change" buttons. When the user is satisfied with the search parameters, the "Search" button is pressed to initiate the search.

If the search is successful, it returns a summary page from which the user can select the matches to be displayed. The summary pages are navigated and the user can choose to display matches for any publication with the "View matches" button. All matches are shown in another window by pressing the "View all matches" link. Alternatively, the selected summaries and their corresponding matches can be sent to the user in plain text format via email.

The sentences that match the search parameters and were selected by the user for display are then shown on the results page. In this case, any sentence in a publication that contains either a member of the "allele" ontology class or the keyword "tra−3" is displayed.

to the top

[Description]   [Contents]   [Advanced Retrieval]

Performing an Author Search

Below is an example of an author search using the Simple Retrieval tool in Textpresso. Please see the " Examples " section for many more examples of Textpresso searches.

The sentence or publication and the ontology class parameters are not used in this search. The user enters an author name in the text box, entering the lastname and then, optionally, the firstname initial (for example "Sternberg" or "Sternberg, p"). "Author" is selected to search against (make sure to unselect the other options). The "Search" button is then pressed to initiate the search.

An index of all the author names in Textpresso is searched and if the search is successful, a page is returned which gives details of all abstracts and papers in Textpresso where the author name is matched.

to the top

[Description]   [Contents]   [Advanced Retrieval]

Performing a Year Search

Below is an example of a text search using the Simple Retrieval tool in Textpresso. Please see the " Examples " section for many more examples of Textpresso searches.

The Documentation Page for the Textpresso Information Extraction System

Query should be met in a ○ sentence ● publication.

Type in keywords of interest, separated by **white spaces.** The wild card sign (*) may be used:

```
1999                                          ☒ Exact match
```

Specify categories of interest:

| none ▼ | none ▼ | none ▼ | none ▼ |

Search in

☐ Abstract ☐ Author ☐ Full Text ☐ Title ☒ Year

[Search!] [Load last query!] [Undo current changes!]

1513 matches in 1513 publication(s) found.

[Display] page [1 ▼], or [previous] or [next] page.

[E-mail] results to `bigbird@sesamestreet.com`, ☒ including matches.

results in PDF   view all matches

Search Results

Title	Abstract	Author	Journal	Year	Citation	Number of matches	Select
Caenorhabditis elegans .	Caenorhabditis elegans is a free-living soil nematode , about 1 mm in length , that is found around the world . It is currently a common laboratory model for many aspects of cellular , developmental , and molecular biology . [Expand abstract]	Pilgrim DB	"Encyclopedia of Reproduction". Ed.: E Knobil and J.D. Neill. Pub: Academic Press.	1999	Pages 449-457 Vol. 1	1	cgc3279 view matches PDF

The sentence or publication and the ontology class parameters are not used in this search. The user enters a year in the text box, in the format *yyyy* (for example "2000" or "1986"). "Year" is selected to search against (make sure to unselect the other options). The "Search" button is then pressed to initiate the search. An index of all publication dates of all the abstracts and papers in Textpresso is searched and if the search is successful, a page is returned which gives details of all abstracts and papers published that year.

to the top

[Description]   [Contents]   [Advanced Retrieval]