

Textpresso User Guide

[\[Contents\]](#)

[\[Simple Retrieval\]](#)

-
- ◇ [Description](#)
 - ◇ [What is *Textpresso* ?](#)
-

Description

This is a first draft User Guide for the Textpresso Information Retrieval system. It is designed to provide information on the concept behind Textpresso and details on how to use the system. If anything in this guide is incomplete, unclear or you have further questions, please email [Textpresso](#) to let us know!

For more detailed information on the concepts behind Textpresso, see the [About Textpresso](#) section. For examples of searches you can perform with Textpresso, see the [Examples](#) section.

[to the top](#)

[\[Contents\]](#)

[\[Simple Retrieval\]](#)

What is *Textpresso* ?

Simply put, Textpresso is a search tool to help retrieve and process information from a corpus of *C. elegans* abstracts and papers. Textpresso facilitates three levels of text searching:

- [Simple Retrieval](#)
- [Advanced Retrieval](#)
- [Fact Extraction](#)

Searches can be performed by entering keywords into the Textpresso search field, much like popular search engines such as [Google](#) and [PubMed](#) . Additionally, searches can also be performed by selecting classes from the **Textpresso Ontology** . But what exactly *is* an ontology?

The Documentation Page for the Textpresso Information Extraction System

An ontology is a set of vocabularies or a dictionary. The members of each ontology class are groups of words that have similar meanings, where the ontology class name represents the "sense" of those words. For example, the ontology class "Regulation" contains words such as "repress", "enhance", "suppress" and the ontology class "Gene" contains all known *C. elegans* three-letter gene names and the word "gene". The corpus of text is annotated with the ontology terms, where if term in the ontology is matched in the corpus of text, that word is tagged with the appropriate ontology term. Therefore, searching the text corpus by ontology classes allows the user a much broader and intuitive search that with keywords alone.

There are two main categories of ontology classes used in the Textpresso search engine; (i) words that describe biological entities (such as the "gene", "phenotype", "allele" and "cell" ontology classes) and (ii) words that describe the relationship between entities (such as the "regulation", "purpose", "localization" and "association" ontology classes). A third category of ontology classes exists that is design for use in semantic analysis and is not search-able in the Textpresso system. Below is a table of the current ontology classes search-able in Textpresso.

Biological Entities	Relationships Between Entities
Allele	Action
Cell or Cell Group	Association
Cellular Component *	Biological Process *
Clone	Characterization
Drugs	Comparison
Entity Feature	Consort
Gene	Descriptor
Life Stage *	Effect
Molecular Function	Involvement
Mutant	Localization
Nucleic Acid	Method
Organism	Pathway
Phenotype	Purpose
Sex	Regulation
Strain	Spatial Relation
Transgene	Time Relation

* The Textpresso system incorporates the controlled vocabulary developed by the Gene Ontology Consortium to describe the biology of a gene product in any organism. There are three ontologies that describe the **molecular function** of a gene product, the **biological process** in which the gene product participates, and the **cellular component** where the gene product can be found.

The Documentation Page for the Textpresso Information Extraction System

The extensive use of ontologies to search text differentiates Textpresso from other Information Extraction systems. **In order to search effectively by ontologies however, the user must understand thoroughly the sense meaning of each of the ontology classes.** See the [Ontology](#) section for a complete list of ontology classes, their definitions and examples.

The following sections give detailed instructions on how to perform searches using the Textpresso system, from the most basic text retrieval to sophisticated information extraction. The Textpresso search engine is specifically designed for ease and clarity of use. The system can also be customized by the user to suit his/her individual preferences.

Finally, Textpresso is a dynamic and evolving project at [WormBase](#) . The Textpresso ontology continues to grow with curator-level input and as more biological information becomes available. We also depend heavily on user feedback to develop and refine the Textpresso search tools and features. If you would like to comment or contribute to Textpresso, please email [Textpresso](#) !

[to the top](#)

[\[Contents\]](#)

[\[Simple Retrieval\]](#)