

Text Search

SEARCH THE TLG CORPUS[®]
SIMPLE | PROXIMITY

Word Index Lemma Textual Search

Enter a word or lemma to search for

Case sensitive Exact Match Diacritics sensitive Wildcard ?

Input

Search in: Full Corpus
or: Author

There are two main options in Text Search:

Simple (search for one string/word or a phrase)

and

Proximity (search for one or more words in proximity).

SEARCH THE TLG CORPUS[®]
SIMPLE | **PROXIMITY**

Word Index Lemma Textual Search

and and

Within Words first word first word

Case sensitive Exact Match Diacritics sensitive Wildcard ?

Input

Search in: Full Corpus
or: Author

There are three types of searches:

Word Index search: searches for a specific string (word or phrase) within a pre-indexed list of all words in the corpus (fast).

Lemma Search: searches for a lemma (i.e. dictionary form) and retrieves all its word-forms in the TLG corpus.

Textual Search: searches for a specific string (word or phrase) going through all texts in the corpus (slower).

User may search the full corpus or limit the search to a particular author(s) selected from the Canon.

TEXT
SEARCH
OPTIONS

Simple
or
Proximity

What kind of search type do I need?

Word Index Search

The **Word Index Search** searches through the TLG Word Index. A Word Index search is a more efficient way to search the corpus if you know the exact form you wish to search for. For example, enter the genitive form **ἄνθρώπου**. The TLG will retrieve all the instances of the word regardless of their spelling variants.

The search string is treated as a **prefix** of possible word-forms. (You can use space to terminate the word). Word Index search ignores accents and punctuation unless the user selects "Exact Search."

Lemma Search searches through the TLG Lemma database which contains lemmata extracted from a number of lexica. If a lemma does not exist in the database, the search engine will return no results. In those cases, Word Index or Textual search should be used. If you enter lemma **πόλις**, the search engine will retrieve all wordforms belonging to this lemma, including its variants.

The Textual Search goes through the actual text files, character by character, to find the string you have specified. The results will include all words that contain the specified string. Spaces can be used to indicate the beginning and end of words. Textual search will be slower than Word index and Lemma Search. Textual Search ignores accents and punctuation unless the user selects "Exact Search."

TYPES OF SEARCH

Word Index

Lemma

Textual Search

SEARCH THE TLG CORPUS
SIMPLE | PROXIMITY

Word Index Lemma Textual Search

Wildcard Non-alphab. chars Diacritics sensitive
 Case sensitive Adscript as subscript

Input

Input maybe
Greek, Beta or
Transliteration

If this option is specified, the search is sensitive to case. The system will search for upper or lower case depending on what the user enters.

When the adscript option is specified, the search treats unambiguous instances of iota adscript identically to iota subscripts.

This option allows search for non alphabetic characters, such as beta escapes that are not rendered in Unicode.

The search is sensitive to Greek diacritics, which are otherwise ignored.

SEARCH FEATURES

Case Sensitive

Diacritics Sensitive

Non-Alphabetic characters

Adscript as Subscript

Substring match

Exact match

- **Case Sensitive:** When this option is selected, the search is sensitive to case, which is otherwise ignored.
- **Diacritics Sensitive:** If this option is specified, the search is sensitive to Greek diacritics, which are otherwise ignored in the search string.
- **Non-Alphabetic Characters:** This option allows search for non alphabetic characters, such as beta escapes and characters not rendered in Unicode.
- **Adscript as Subscript:** The TLG corpus has not in general resolved adscript iota into subscript iota. Therefore, adscript iota appears in some editions. If the adscript option is specified, the search treats unambiguous instances of iota adscript identically to iota subscripts.
- **Non Alphabetic Characters:** This feature is used to search for symbols and beta escapes. The search engine will detect non-alphabetic characters in most cases and may return results even the box is not checked.
- **Substring Match:** When this option is selected in Lemma Search, the search engine will look for all compound forms of the lemma.
- **Exact match:** When this option is selected the search engine will search for the exact string entered in the search box. Exact match is possible in Word Index and Textual Search.

Wildcard Searches



The screenshot shows a search interface titled "SEARCH THE TLG CORPUS". Below the title are two options: "SIMPLE" and "PROXIMITY". Underneath, there are three radio buttons: "Word Index" (selected), "Lemma", and "Textual Search". A text input field contains the placeholder "Enter a word or lemma to search for" and a red "GO" button. Below the input field are four checkboxes: "Wildcard" (checked), "Case sensitive", "Exact Match", and "Diacritics sensitive". To the right, there is an "Input" dropdown menu set to "Beta" and a keyboard icon. A red arrow points to the "Wildcard" checkbox.

Wildcard search allows the substitution of one character for another. Wildcard is possible in Word Index and Textual Search.

The following wildcard combinations are possible:

- Search for ANA as part of a word: ANA
- Search for ANA as a prefix (word index)/at the start of a line (full text): ^ANA
- Search for ANA as a suffix (word index)/at the end of a line (full text): ANA\$
- Search for the word ANA (word index only): ^ANA\$
- Search for ANADU or ANEDU: AN[AE]DU
- Search for ANADU or ANEDU: ANADU|ANEDU
- Search for ANADU or ANEDU: (ANA|ANE)DU
- Search for A)NADU or A)NEDU: (A\)NA|A\)NE)DU
- Search for ANAD followed by anything but U: ANAD[^U]
- Search for a word consisting of ANAD, then any two letters, then U: ^ANAD..U
- Search for EDU or ANEDU: (AN)?EDU
- Search for a numeric digit: [0-9]
- Search for an Arabic numeral (a sequence of one or more digits): [0-9]+
- Search for a quotation mark Beta escape (" followed by zero or more digits): "[0-9]*

Note: In both Word index and Textual search, special beta character such as `*()/\|+` should be escaped by prefixing them with backslash (`\`).

WILDCARD SEARCH

Available in
Word index and
Textual Search

Phrasal Searches

A **phrase** (not just a single word) can be entered into the Simple search bar. Entering the text "MHNIN AEIDE QEA PHLHIADEW" in the search bar as a Word Index search returns all instances of the four-word phrase in the corpus. Proximity phrasal search has not been implemented in Phase 1 (use Proximity instead).

There are other ways to search for phrases in the TLG, though phrasal search is more powerful than each of them:

1. You can search for phrases by using Proximity search: "Search for MHNIN and AEIDE and PHLHIADEW within 2 words". However, this limits you to just three words in the phrase. Moreover, proximity search no longer differentiates between a word coming before or after a second word; so you cannot specify that the words in the phrase occur in that specific order.
2. You can use Textual Search. Textual search bypasses the word index, and searches for the phrase through the entire corpus, letter by letter; so it is much slower than searches involving the word index. Textual search deals with spaces between words, so you can have an arbitrarily long phrase. However you will have to insert spaces around the phrase, to make sure that the initial and final words are not treated as word fragments: a search for "O GE" will match "*DANAI=O GENE/QLH", while " O GE " limits the search to the two words.
3. You can use the new Browse by N-grams feature to find a sequence of two or three content words highlighted in a text, and look up all recurrences of that phrase by clicking on it. This search is limited to two or three content words, it ignores the order of words and ignores stop words, such as articles and particles.

PHRASAL SEARCHES

How to search for phrases

Display Preferences

The screenshot shows a search interface with the following elements and annotations:

- Lines of context:** 3 (dropdown)
- Display:** Greek (dropdown)
- Sort by:** Author (dropdown)
- RESULTS** (header)
- Display results:** As a list | per word | by author | grammar (checkboxes)
- Save your search** (button)
- Print your results** (button)

Annotations with arrows pointing to these elements:

- Greek or Beta code** points to the 'Display' dropdown.
- Sort results by date or in alphabetical order** points to the 'Sort by' dropdown.
- Save your search** points to the 'Save your search' button.
- Print your results** points to the 'Print your results' button.
- Display all wordforms and wordcounts as a list.** points to the 'As a list' checkbox.
- Display wordforms. Choose one wordform at a time from sidebar to view results.** points to the 'per word' checkbox.
- Display results by author. Choose author from sidebar to view results per author.** points to the 'by author' checkbox.
- Display results listed by their grammatical category (i.e. gender, number, case or tense, voice, mood)** points to the 'grammar' checkbox.

Results may be displayed in Greek or beta code and sorted alphabetically or by the date of the author.

Results may be displayed as a list of word-forms (user may select all or some of the forms to display the results), as a clickable list of word-forms (user may select one at a time) by author and by grammatical category.

Proximity Search

SEARCH THE TLG CORPUS

SIMPLE | PROXIMITY

Word Index Lemma Textual Search

and and

Within Words first word first word

Case sensitive Exact Match Diacritics sensitive Wildcard ?

Input

Search in: Full Corpus
or: Author

Searched 10,022 works

Lines of context: Display: Sort by: Results per page:

SELECTION

πρᾶττω

Select all

- πρᾶττω (266)
- πρᾶττω (1)
- πρᾶττωμεν (202)
- πρᾶττων (1,830)
- πρᾶττωνται (5)
- πρᾶττωντες (1)
- πρᾶττωσ' (1)
- πρᾶττωσι (60)
- πρᾶττωσιν (80)

σωζεσθαι

Select all

- σωζεσθαι (1,167)
- σωζεσθαι (981)
- σωζεσθαι (1)

RESULTS

Display results: As a list | **by author**

Prev | Next

- PLOTINUS Phil. *Enneades* {2000.001} Ennead 3 chapter 2 section 8 line 43**
τα τὰ κατὰ τὸν βίον γνώμη τῆ ἑαυτῶν πρᾶττειν, κἂν μὴ αὐτῆ πρᾶττωσιν, ἣ θεοῖς φίλα, σωζεσθαι δὲ μόνον παρὰ θεῶν οὐδὲ ταῦτα ποιήσαντας, δι' ὧν κελεύουσιν αὐτοὺς οἱ (45)
- THEODORETUS Scr. Eecl. et Theol. *Graecarum affectionum curatio* {4089.001} Book 6 section 69 line 10**
πάντα τὰ κατὰ τὸν βίον γνώμη τῆ ἑαυτῶν πρᾶττειν, κἂν μὴ αὐτῆ πρᾶττωσιν ἣ θεοῖς φίλα, σωζεσθαι δὲ μόνον παρὰ θεῶν οὐδὲ ταῦτα ποιήσαντας, δι' ὧν κελεύουσι τούτους οἱ θεοὶ σωζεσθαι. Καὶ τοίνυν

TEXT SEARCH

Proximity

Word Index

Lemma

Textual

Proximity search allows the user to enter up to three words/lemmata in proximity. Up to 50 words of context may be specified (15 is the default).

The same options used in Simple Search apply to Proximity search.

The lists of word-forms are shown in the left sidebar and the user may choose some or all. Results may be displayed as a list or by author (Word Index), as a list, by author or by grammar (Lemma) and as a list in Word Index.

Phrasal search is not available in Proximity as of Phase 1.