

# Text Search

SEARCH THE TLG CORPUS

**SIMPLE** | PROXIMITY | ADVANCED 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 three main options in Text Search:

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

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

**Advanced Proximity** (search for a wordform, lemma or grammatical category in proximity)

SEARCH THE TLG CORPUS

**SIMPLE** | PROXIMITY | ADVANCED PROXIMITY

Word Index    Lemma    Textual Search

Enter a word or lemma to search for

Case sensitive    Exact Match    Diacritics sensitive    Wildcard ?

Input

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**).

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

## TEXT SEARCH OPTIONS

Simple  
Proximity  
or  
Advanced  
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 **ἀνθρώπου**. 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

In exact match, the system will search for the exact string selected by the user.

Input may be in Greek, beta or transliteration

The screenshot shows a search interface for the TLG Corpus. At the top, it says "SEARCH THE TLG CORPUS" with a lightbulb icon. Below that are three search modes: "SIMPLE" (underlined), "PROXIMITY", and "ADVANCED PROXIMITY". There are three radio buttons for search types: "Word Index" (selected), "Lemma", and "Textual Search". A search input field contains the placeholder text "Enter a word or lemma to search for" and a red "GO" button. Below the input field are four checkboxes: "Case sensitive", "Exact Match", "Diacritics sensitive", and "Wildcard ?". To the right of the main search area, there is a section for "Search in:" with radio buttons for "Full Corpus" (selected) and "Author", and a text box labeled "Author Search". At the bottom right, there is an "Input" field with "Beta" and a keyboard icon. Red arrows point from external text boxes to the "Exact Match" checkbox, the "Diacritics sensitive" checkbox, the "Input" field, and the search input field.

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

The search is sensitive to Greek diacritics that are otherwise ignored.

- **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.
- **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.

## SEARCH FEATURES

Case Sensitive

Diacritics Sensitive

Non-Alphabetic characters

Adscript as Subscript

Substring match

Exact match

# Wildcard Searches

SEARCH THE TLG CORPUS

**SIMPLE** | PROXIMITY | ADVANCED PROXIMITY

Word Index  Lemma  Textual Search

Enter a word or lemma to search for

Case sensitive  Exact Match  Diacritics sensitive  Wildcard ?

Search in:  Full Corpus  
or:  Author

Author Search

Input: Beta

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 as of this Phase.

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

Specify number of lines to display

Display in Greek or beta code

Sort results by date or author name

Show or hide hyperlinks. No links is the ADA setting

Expand section to view text

Lines: 3 Display: Greek Sort: Author # Links: Active

« RESULTS »

Display results: as a list | per word | by author | grammar |    

Display all wordforms with their counts as a list

Choose one wordform at a time to display results

Choose an author from the sidebar to view the results for the author

Display results by their grammatical category (Lemma Search)

How to cite

Save search

Export citations in csv format

Print search results

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

Results maybe 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.

Users may save, print or export their search results by clicking on the appropriate icon.

# Proximity Search

SEARCH THE TLG CORPUS<sup>!</sup>  
SIMPLE | **PROXIMITY** | ADVANCED PROXIMITY

Search in:  Full Corpus  
or:  Author

Word Index  Lemma  Textual Search

and  and

Within  Words  first word  first word

Case sensitive  Exact Match  Diacritics sensitive  Wildcard ?

Searched: 7 out of 7 works

Lines:  Display:  Sort:  Links:  Results per page:

SELECTION

prattei

Deselect all

- πράττει (1,498)
- πράττει (1)
- πράττειεν (1)
- πράττειν (6)
- πράττειν (8,634)
- πράττεις (386)

swzesqai

Deselect all

- σώζεσθαι (1,290)
- σώζεσθαι (1,040)
- σώζεσθαι (1)

« RESULTS »

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

Prev | Next

- LIBANIUS Rhet. et Soph. Declamationes 1-51 {2200.005}** Declamation **49**    
(subdivision) **2** section **36** line **2**  
τοσαύτης ἐπιμελείας ὁρῶν ἔτι καὶ τὰ προσή-  
κοντα **πράττειν** δυνάμενος, ὁμῶς δὲ **σώζεσθαι** τὸν πα-  
τέρα καὶ μηδενὸς ἀηδοῦς πείραν λαμβάνειν ἐκάστοτε
- Anna COMNENA Hist. Alexias {2703.003}** Book **5** chapter **1** section **1** line    
**13**  
ὅτι ὁ Ῥομπέρτος εἰς τὸ ἐπὶν ἔαυ τὴν πολιορκίαν ταμειύεται, διεσκοπεῖτ  
ἐκαστος ὁ τι **πράττειν** χθὴ καὶ **σώζεσθαι** καὶ μὴ αἰθὶς ἐς τοσοῦτους ἐμ-  
πεπτοκέναι κινδύνους. (2) συλλεξάμενοι οὖν ἑαυτοὺς τὸ ἀπόρητον
- ISIDORUS Pelusiota Epist. et Scr. Ecl. Epistulae de interpretatione divinae**    
*scripturae* {2741.003} Book **2** epistle **270** line **49**  
Εἰ γὰρ οὕτω νοηθεῖ τὸ προφητικὸν χωρίον, καὶ τῷ  
Θεῷ τὸ πρέπον φυλαχθεῖ, τὰ αὐτοῦ **πράττειν** προ-  
ρημένῳ. **Σώζεσθαι** γὰρ ἐκ παντὸς τρόπου τοὺς εἰς (50)  
κακίαν καλινδουμένους βούλεται, καὶ πέμπει τοὺς

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

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.

TEXT  
SEARCH

Proximity

Word Index

Lemma

Textual



# Advanced Proximity Search (Beta)

SEARCH THE TLG CORPUS

SIMPLE | PROXIMITY | **ADVANCED PROXIMITY (Beta Release)**

Search in:  Full Corpus  
or:  Author

WI  L  G       WI  L  G       WI  L  G

swzesqai    and    πράσσω    and    Noun Fem. Dat. Sg.    **GO**

Within 10 Words      near first word      near first word

Case sensitive     Diacritics sensitive     Exact match (WI only)     Wildcard (WI only) ?    Input Beta

Lines: 7    Display: Greek    Sort: Author #    Links: Active    Results per page: 50

« RESULTS »

Display results: As a list |

Prev | Next

1. **LIBANIUS Rhet. et Soph. Orationes 1-64 {2200.004}** Oration 23 section

**19** line 7  
πλούτων ἐτέρωσε καταθέμενοι γυμνοὶ πρὸς τὸν  
ἔξω δρόμον ἦσαν ἔτοιμοι, καίτοι χρῆν αὐτοὺς καὶ (5)  
τούς ἐν πενία κωλύειν καὶ κατέχειν, τὰ μὲν πειθοί,  
τὰ δὲ καὶ ἀπειλαίς. οἱ δ' ἐπήνουν τε τοὺς ἐκείνο  
ποιούντας καὶ ἀνθρώπους ἐλεγον εἶναι **σώζεσθαι** βου-  
λομένους, ὡσπερ ὃν εὐ **ἐπραξαν** ἐν τῇ πόλει δίχης  
παρὰ τῆς πόλεως αὐτοῖς ὀφειλομένης. οἱ τοῦθ' ἐν (10)  
μόνον εὐ πεποιθήκασιν, ὅτι τὸν ἄλλον χρόνον ἀρνού-  
μενοι μὴ πλουτεῖν νῦν αὐτοὺς ἐξήλεγξαν ὡς πλείστα  
κεκτημένοι πόνῳ πολλῷ καὶ πραγματεῖα μεταθέντες

2. **MENANDER Rhet. Περὶ ἐπιδεικτικῶν {2586.002}** Spengel page 374 line

**32**  
ὅτι νικήσας ὁ βασιλεὺς οὐ τοῖς ὁμοίοις ἡμίνατο τοὺς  
ἀρξάντας ἀδίκων ἔργων, ἀλλ' ἐμέρισε κατὰ τὸ δί- (30)  
καιον τὰς πράξεις τιμωρία καὶ φιλανθρωπία, καὶ ὅσα  
ἠγείτο πρὸς σωφρονισμόν ἀρκεῖν ἐργασάμενος, ἐν-  
(375) ταῦθα στήσας **φιλανθρωπία** τὰς **πράξεις** ἀνήκε συγχω-  
ρήσας τὸ λείψανον τοῦ γένους **σώζεσθαι**, ἅμα μὲν ἵνα

**SELECTION**

σώζεσθαι

Deselect All

- σώζεσθαι (1,290)
- σώζεσθαι (1,040)
- σῶζεσθαι (1)

πράσσω

- πρᾶττοντος (200)
- πρᾶττω (285)
- πρᾶττη (274)
- ἐπραξας (256)
- πράσσω (236)
- πρᾶττοι (234)
- πράξαντος (215)
- πρᾶττε (213)
- πράξη (211)
- πρᾶττοντι (207)
- πρᾶττωμεν (207)
- πράξω (203)
- πράξαντα (199)
- ἐπράξαμεν (192)
- πραξάντων (190)
- πράξαντας (187)
- πρᾶσαντες (176)

## TEXT SEARCH

## Search in Advanced Proximity

Advanced proximity search allows you to search for any combination of wordforms (WI), lemmata (L) or grammatical categories (G) in proximity. Up to 3 search items may be selected and up to 50 words of context may be specified. Unlike regular Proximity, Advanced Proximity allows you to search for any combination of grammatical categories, wordforms or lemmata in proximity.

Advanced proximity is released in Beta version and should be used with the understanding that, due to the ambiguity of many Greek forms, the system may produce spurious results.

# Advanced Proximity Search (Beta)

SEARCH THE TLG CORPUS<sup>+</sup>  
SIMPLE | PROXIMITY | **ADVANCED PROXIMITY** (Beta Release)

Search in:  Full Corpus  
or:  Author

WI  L  G and WI  L  G and WI  L  G  
swzesgai and πράσσω Fem. Dat. Sg. GO

Within 10 Words near first word near first word

Case sensitive  Diacritics sensitive  Exact match (WI only)  Wildcard (WI only) ? Input: Beta

Lines: 7 Display: Greek Sort: Author # Links: Active Results per page: 50

SELECTION  
σώζεσθαι  
Deselect All  
1.  σώζεσθαι (1,290)  
2.  σώζεσθαι (1,040)  
3.  σώζεσθαι (1)

πράσσω  
Fem. Dat. Sg. (limits apply \*)  
Deselect All  
1.  τῆ (506,850)  
2.  αὐτῆ (49,727)  
3.  φύσει (31,079)  
4.  ἡ (29,455)  
5.  ταύτη (27,744)  
6.  δυνάμει (21,653)  
7.  τινί (17,087)  
8.  ἡμέρα (16,335)  
9.  πόλει (16,265)  
10.  τινι (16,179)  
11.  ψυχῆ (13,058)  
12.  τῶι (11,507)  
13.  νῆ (11,240)

« RESULTS »  
Display results: As a list |