

# Evaluation, CLIR Strategies and Metadata

Shibamouli Lahiri

December 6, 2010

# Current Search Interface



The screenshot shows a Mozilla Firefox browser window titled "CyDAR Multilingual Search - Mozilla Firefox". The address bar contains the URL "http://cs05.ist.psu.edu/8080/cydar/". The page features a central image of the Virgin Mary and the Christ Child. Below the image, the text "CyDAR Multilingual Search" is displayed in a large, bold font. A search prompt in English and Greek asks the user to enter a query. There are two radio buttons for "English" (selected) and "Greek". Below this is a text input field. A second text input field is labeled "Translation here (μετάφραση εδώ)". Below the input fields are five radio buttons for search engines: "CyDAR CLIR" (selected), "Perseus English", "Perseus Greek", "Gutenberg Greek", "CyDAR English", and "CyDAR Greek". At the bottom of the form are "Submit" and "Reset" buttons. The browser's status bar at the bottom shows "Done" and the system tray includes a clock displaying "7:21 PM".

# Search Results

CyDAR Multilingual Search - Mozilla Firefox  
File Edit View History Bookmarks Tools Help  
http://coo05.ist.psu.edu0080/cydar/search.jsp?query\_lang=English&search\_box=socrates&tr\_box=Σωκράτης&language\_button=CCLIR  
Most Visited Getting Started Latest Headlines  
Search the Web SEARCH #10 Set Your Status  
webshots Search the web My Homepage Upload

37 results for the query **socrates**. Showing 1-10.

[04 taifaco AKYG6 231-655-0002-0005](#)  
...the description is compared with reports of **Socrates**. Here is the passage in the memoirs of Xenophon (2.6.30-33), is part of the dialogue about friendship and friends: He said he Kritovoulos; and not, O **Socrates**, that I of course especially when standing...upon their souls and good bodies. And **Socrates** said; but, O Kritovoule, essential enestin én Now...

[04 taifaco AKYG6 231-655-0001-0009](#)  
...Did not legitimately apokteinynai him himself, O **Socrates**? egoge already gar, opera nyndi rou thou, and I heard Filolaou, OTE despite us diptato already he and certain other, as a st doei this poiesis. In between **Socrates**, Kevitos Simmia and ensuing debate are examined almost all afflicted these issues. He says that **Socrates** (62b): The former, then, en aporrizois called...

[04 taifaco AKYG6 231-655-0002-0011](#)  
...moderate, except if not for the wise **Socrates** Truly Estin writings, he graciously spared him beneath Xanthippi tής **Socrates** had for women after his death, as...the authority (the headless), did not show **Socrates'** style (heavy, for they are elderlymenoi reveal themselves and it was not his Sokratikis eftoniai) . For them, therefore, there was question of

11 results for the (Google-translated) query **Σωκράτης**. Showing 1-10.

[04 taifaco AKYG6 231-655-0103](#)  
...γῆρας, ὅπως δὲν εἶχε ἀποδεχθῆαι καὶ ὁ **Σωκράτης** παλαιότερα παρόμοια πρόταση τοῦ βασιλιᾶ Ἀρχελάου, ἀλλὰ γὰ ἀλλους λόγους· ἀναφέρει μία ἀπὸ τὴς πηγῆς, ὁ Ἀριστοτέλης (Ρητορικὴ 2.23, 1398a 24-26): ὄβρον γὰρ ἔση (sc. ὁ **Σωκράτης**) εἶναι τὸ μὴ δύνασθαι ἀμύνασθαι ὁμοίως καί...

[04 taifaco AKYG6 231-655-0119](#)  
...ἐπὶ τοὺς καλοὺς τὰ σώματα. καὶ ὁ **Σωκράτης** ἔση Ἀλλ', ὁ Κριτόβουλε, οὐκ ἔνεστιν ἐν τῇ ἐμῇ ἐπιστήμῃ τὸ τὰς χεῖρας προσφέροντα πατεῖν ὑπομένειν τοὺς καλοὺς. ... καὶ ὁ Κριτόβουλος ἔση Ὡς οὐ προσοίσιοντος τὰς χεῖρας, εἰ ἢ ἔχεις ἀγαθὸν εἰς φίλων κτήσιν, διδάσκει. οὐδὲ τὸ στόμα οὖν, ἔση ὁ **Σωκράτης**, πρὸς τὸ στόμα προσοίσεις; Θάρρει, ἔση ὅ...

[01 taifaco proto-0048-0043](#)  
...Ἡρακλεώτῃ. Τὰ μεγάλα πρότυπα του ἦταν ὁ **Σωκράτης** καὶ ὁ Διογένης, ἀλλὰ διεβίωσε μὲ τίς...

11 results for the (dictionary-translated) query **Σωκράτης**. Showing 1-10.

[04 taifaco AKYG6 231-655-0103](#)  
...γῆρας, ὅπως δὲν εἶχε ἀποδεχθῆαι καὶ ὁ **Σωκράτης** παλαιότερα παρόμοια πρόταση τοῦ βασιλιᾶ Ἀρχελάου, ἀλλὰ γὰ ἀλλους λόγους· ἀναφέρει μία ἀπὸ τὴς πηγῆς, ὁ Ἀριστοτέλης (Ρητορικὴ 2.23, 1398a 24-26): ὄβρον γὰρ ἔση (sc. ὁ **Σωκράτης**) εἶναι τὸ μὴ δύνασθαι ἀμύνασθαι ὁμοίως καί...

[04 taifaco AKYG6 231-655-0119](#)  
...ἐπὶ τοὺς καλοὺς τὰ σώματα. καὶ ὁ **Σωκράτης** ἔση Ἀλλ', ὁ Κριτόβουλε, οὐκ ἔνεστιν ἐν τῇ ἐμῇ ἐπιστήμῃ τὸ τὰς χεῖρας προσφέροντα πατεῖν ὑπομένειν τοὺς καλοὺς. ... καὶ ὁ Κριτόβουλος ἔση Ὡς οὐ προσοίσιοντος τὰς χεῖρας, εἰ ἢ ἔχεις ἀγαθὸν εἰς φίλων κτήσιν, διδάσκει. οὐδὲ τὸ στόμα οὖν, ἔση ὁ **Σωκράτης**, πρὸς τὸ στόμα προσοίσεις; Θάρρει, ἔση ὅ...

[01 taifaco proto-0048-0043](#)  
...Ἡρακλεώτῃ. Τὰ μεγάλα πρότυπα του ἦταν ὁ **Σωκράτης** καὶ ὁ Διογένης, ἀλλὰ διεβίωσε μὲ τίς...

Done  
gadgets 10:46 PM

- Original Cypriot documents (in Greek)
- English versions obtained using Google Translate

Each search result corresponds to a **page** of the original document.

- For example, **04\_taifac0\_AKYG6\_231-655-0103** corresponds to the 103rd page of the document  
**04\_taifac0\_AKYG6\_231-655**

# Evaluation of English Search Results

- Independent judge
- Set of English queries from Wikipedia articles on Ancient Greece and Ancient Cyprus
- Topical instead of named-entity queries
- The search results give us an idea of the book contents

Archaic period, Battle of Corinth, Athenian Golden Age, Peloponnesian War, Classical Antiquity, Syro-Hittite, Assyrian, Phoenician, battle of Leuctra, Hellenistic period, Delian League, Mycenaean civilization, Lelantine War, Chalcis, helots, Battles of Salamis, Battles of Plataea, Epaminondas, Battle of Chaeronea, League of Macedon, Diadochi, Eudoxus of Cnidus, Knossos, Peace of Nicias, chiton, Hellas, Flaccus, Aulus Gellius

Greek results but no English results

- poleis
- archon basileus
- Nea Nikomedia

The goal is to create a gold-standard set of documents for the above queries. We use ternary relevance judgment:

- 2 = Relevant
- 1 = Marginally Relevant
- 0 = Not Relevant



# Relevance Judgment (Contd.)

Battle of Chaeronea	League of Macedon	Peace of Nicias
2	2	1
2	2	1
0	2	0
0	2	0
0	2	0
0	0	1
0	0	0
0	0	0
0	0	0
0	0	0

<b>Metric</b>	<b>Value</b>
P@1	0.7931
P@2	0.8276
P@3	0.8276
P@4	0.8276
P@5	0.8276
P@10	0.7744
MAP@10	0.8077
MRR	0.8276
NDCG@10	0.8297

- Language Model
- LDA-ESA (Roth and Klakow, 2010)

The Language Model approach requires **sentence alignment** and **translation probability estimation**. LDA-ESA takes Wikipedia-based **link information** into account.

- Alignment of Europarl and Perseus
- Probability Estimation (Brown, et al., 1993)

- **01\_taifaceo\_proto:** Title of the Book, Publisher and Copyright, Preface, Introduction, Contents, Overall Description.
- **taifaceo-arxaio001-154:** Testimonia concerning Cypriot authors and their fragments as found in other ancient authors in the original language (Ancient Greek or Latin). Each passage comes with various other readings found in manuscripts and papyri or suggested by scholars.
- **03\_taifaceo-arxaio155-229:** Continuation of taifaceo-arxaio001-154.
- **05\_taifaceo\_transl\_005-229:** Modern Greek translation of different philosophers' works.
- **04\_taifaceo\_AKYG6\_231-655:** Modern Greek commentary on the works, a glossary, a list of abbreviations, bibliography, index, correlating table of fragments and an epilogue.

## ΚΕΙΜΕΝΟ – ΜΕΤΑΦΡΑΣΗ

### ΚΛΕΑΡΧΟΣ ΣΟΛΕΥΣ

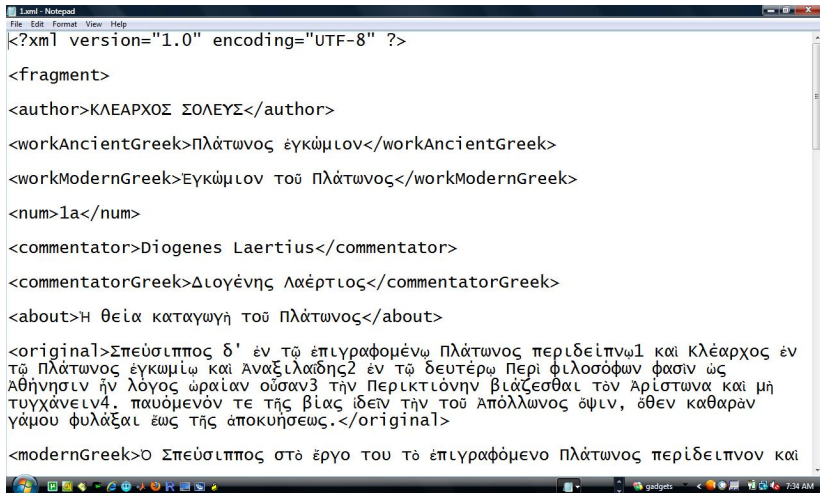
Πλάτωνος ἐγκώμιον

**F1a** (2aW) Diogenes Laertius 3.2

Σπεύσιππος δ' ἐν τῷ ἐπιγραφομένῳ Πλάτωνος περιδείπνω<sup>1</sup> καὶ Κλέαρχος ἐν τῷ Πλάτωνος ἐγκωμίῳ καὶ Ἀναξилаΐδης<sup>2</sup> ἐν τῷ δευτέρῳ Περί φιλοσόφων φασὶν ὡς Ἀθήνησιν ἦν λόγος ὥραϊαν οὔσαν<sup>3</sup> τὴν Περικτιόνην βιάζεσθαι τὸν Ἀρίστωνα καὶ μὴ τυγχάνειν<sup>4</sup>. πανόμενόν τε τῆς βίας ἰδεῖν τὴν τοῦ Απόλλωνος ὄψιν, ὅθεν καθαρὰν γάμου φυλάξει ἕως τῆς ἀποκνήσεως.

<sup>1</sup> περιδείπνω **B<sup>2</sup>P<sup>3</sup>**: περι δείπνω **F**: περι δείπνου **P<sup>1</sup>QWH**    <sup>2</sup> Ἀναξилаΐδης Cobet: ἀναξιλίδης **FP<sup>3</sup> (P<sup>9</sup>)**: ἀναξιλήδης **B**: ἀναξιλήδης **FP<sup>2</sup>**: ἀναξιάδης **P<sup>1</sup>(P<sup>9</sup>)W**: Ἀναξιλείδης Lang    <sup>3</sup> γάμου ante οὔσαν add. Marcovich    <sup>4</sup> ἐπι super τυγχάνειν add. **F<sup>2</sup>**

# XML Formatted



```

Lxml - Notepad
File Edit Format View Help
<?xml version="1.0" encoding="UTF-8" ?>

<fragment>

<author>ΚΛΕΑΡΧΟΣ ΣΟΛΕΥΣ</author>

<workAncientGreek>Πλάτωνος ἐγκώμιον</workAncientGreek>
<workModernGreek>Εγκώμιον τοῦ Πλάτωνος</workModernGreek>

<num>1a</num>

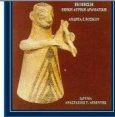
<commentator>Diogenes Laertius</commentator>
<commentatorGreek>Διογένης Λαέρτιος</commentatorGreek>

<about>Ἡ θεία καταγωγή τοῦ Πλάτωνος</about>

<original>Σπεύσιππος δ' ἐν τῷ ἐπιγραφομένῳ Πλάτωνος περιδειπνω1 καὶ κλέαρχος ἐν
τῷ Πλάτωνος ἐγκωμίῳ καὶ Αναξилаιδης2 ἐν τῷ δευτέρῳ Περι φιλοσόφων φασιν ὡς
Ἀθήνησιν ἦν λόγος ὡραίαν οὔσαν3 τὴν Περικτιόνην βιάζεσθαι τὸν Ἀρίστωνα καὶ μὴ
τυγχάνειν4. παυόμενόν τε τῆς βίας ἰδεῖν τὴν τοῦ Ἀπόλλωνος ὄψιν, ὅθεν καθαρὰν
γάμου φυλάξει εἰς τῆς ἀποκησεως.</original>

<modernGreek>Ὁ Σπεύσιππος στὸ ἔργο του τὸ ἐπιγραφόμενο Πλάτωνος περιδειπνον καὶ

```



## CyDAR Metadata Search

Check to select

<input type="checkbox"/>	<b>Author:</b>	<input type="text" value="Select one"/>	<input type="text" value="Or type here..."/>	<input type="text" value="AND"/>
<input type="checkbox"/>	<b>Commentator:</b>	<input type="text" value="Select one"/>	<input type="text" value="Or type here..."/>	<input type="text" value="AND"/>
<input type="checkbox"/>	<b>Fragment:</b>	<input type="text" value="Select one"/>	<input type="text" value="Or type here..."/>	<input type="text" value="AND"/>



- Metadata Search
- Sentence Alignment and Probability Estimation for the Language Model
- Comparison of Language Model with other approaches
- Implementation and evaluation of LDA-ESA

# Acknowledgments

- Prof Zoe Stamatopoulou, Classics and Ancient Mediterranean Studies.
- Partha Mukherjee, IST.