Textual Re-use on Ancient Greek Texts:
A case study on Plato's works

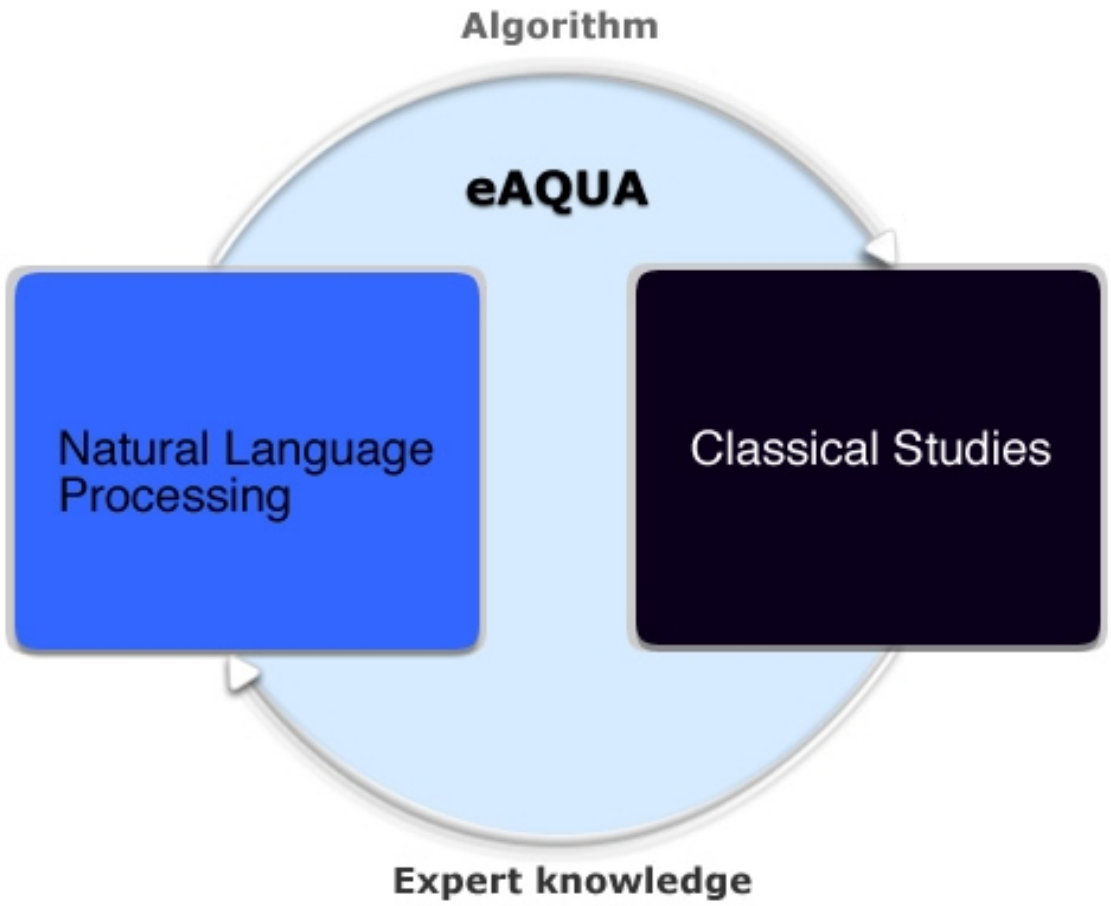
Workshop on Historical Texts
Boston (USA), 2010/01/13

Marco Büchler, Thomas Eckart
Natural Language Processing Group
Department of Computer Science
University of Leipzig

Annette Geßner
Ancient Greek
Institute of Classical Philology and Comparative Studies
University of Leipzig

20 years of Fall of the Berlin Wall (Leipzig, Berlin)

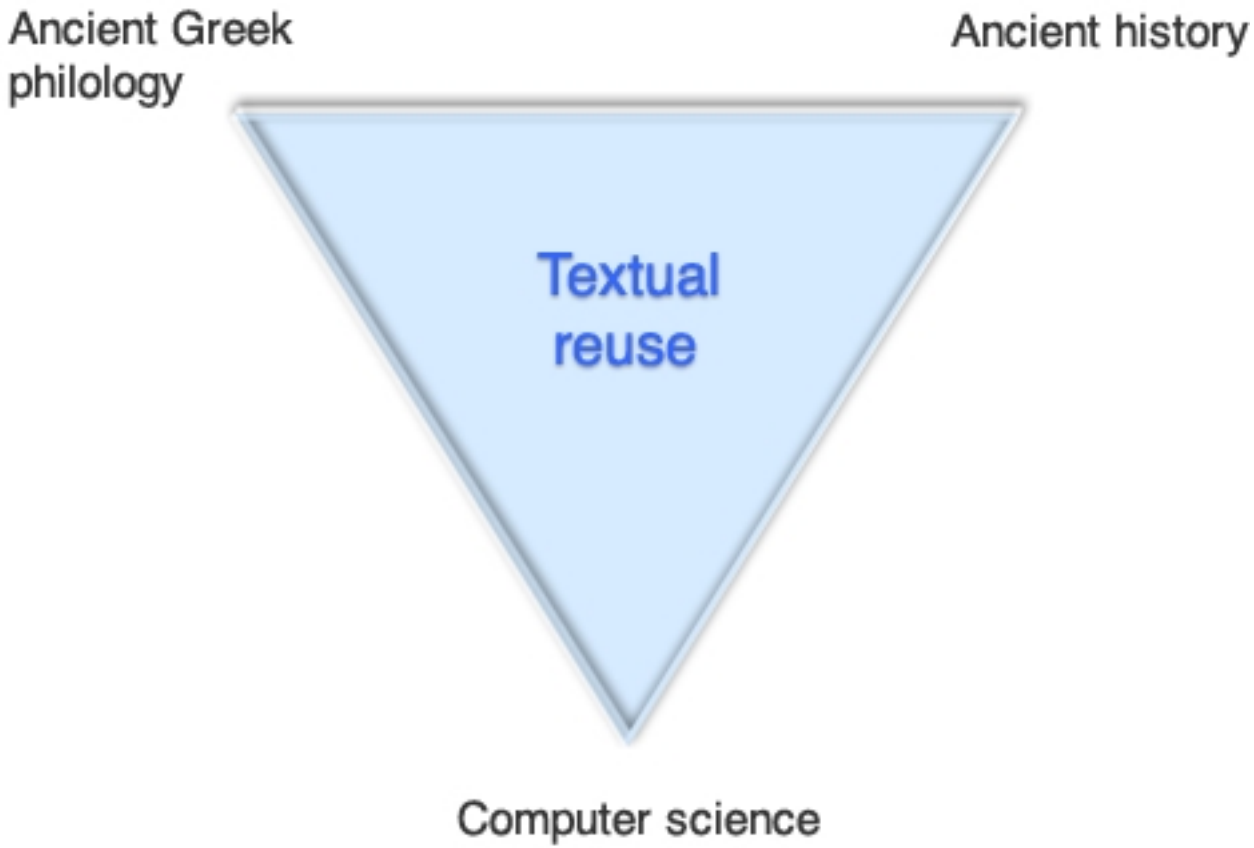
We are the people!



Marco Büchler, Thomas Eckart
Natural Language Processing Group
Department of Computer Science
University of Leipzig

Annette Geßner
Ancient Greek
Institute of Classical Philology and Comparative Studies
University of Leipzig

Different research interests & Agenda



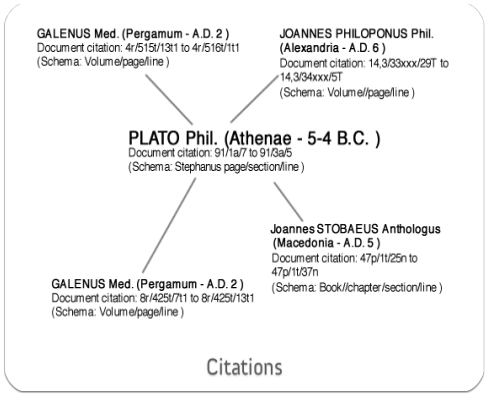
Epigram: Lots of texts produce much more text mining data which can easily be accessed by a powerful Visual Analytics component.

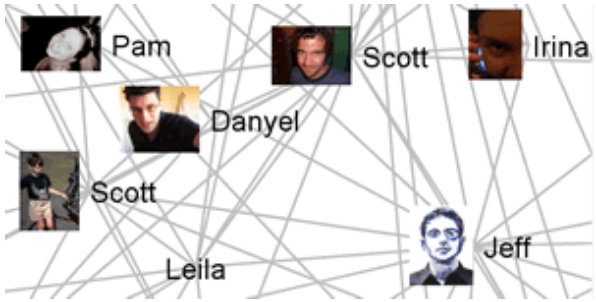
Graph

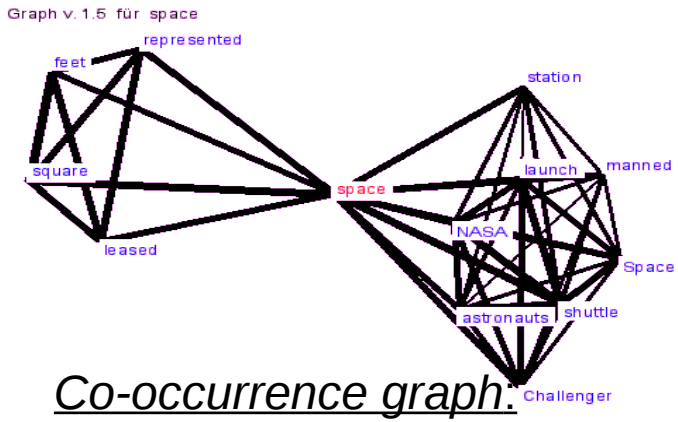
- *Formal*: Graph G=(V,E) V=collection of vertices, E=collection of edges

- *Simple*: pairwise relations between objects from a certain collection

Several examples of a graph:



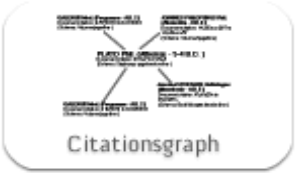




Textual reuse graph:
V: collection of text passages
E: collection of found reuses between **V**

Social Network graph:
V: collection of persons
E: collection of relations between persons in **V**

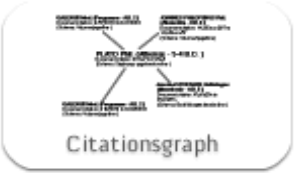
Co-occurrence graph:
V: collection of words
E: collection of found and statistically significant relations between **V**


Citationsgraph

- Naive method: comparing every sentence with all other sentences
- TLG: 5,500,000*5,500,000 = 3.025e13 comparisons
- Assumption: It can be compared 1000 sentences/sec.
- This process would run about **3.025e10 seconds** or more than **959 years.**

- Even if we would compare only sentences with all significant phrases we would need about one year.

- That's why:
- Usage of divide & conquer strategies
- Intelligent pre-clustering of data
- Using occurrences of Plato, work titles or roles of Plato's works
- Using significant terms of Plato's work



Graph

- *Formal*: Graph G=(V,E) V=collection of vertices, E=collection of edges
- *Simple*: pairwise relations between objects from a certain collection

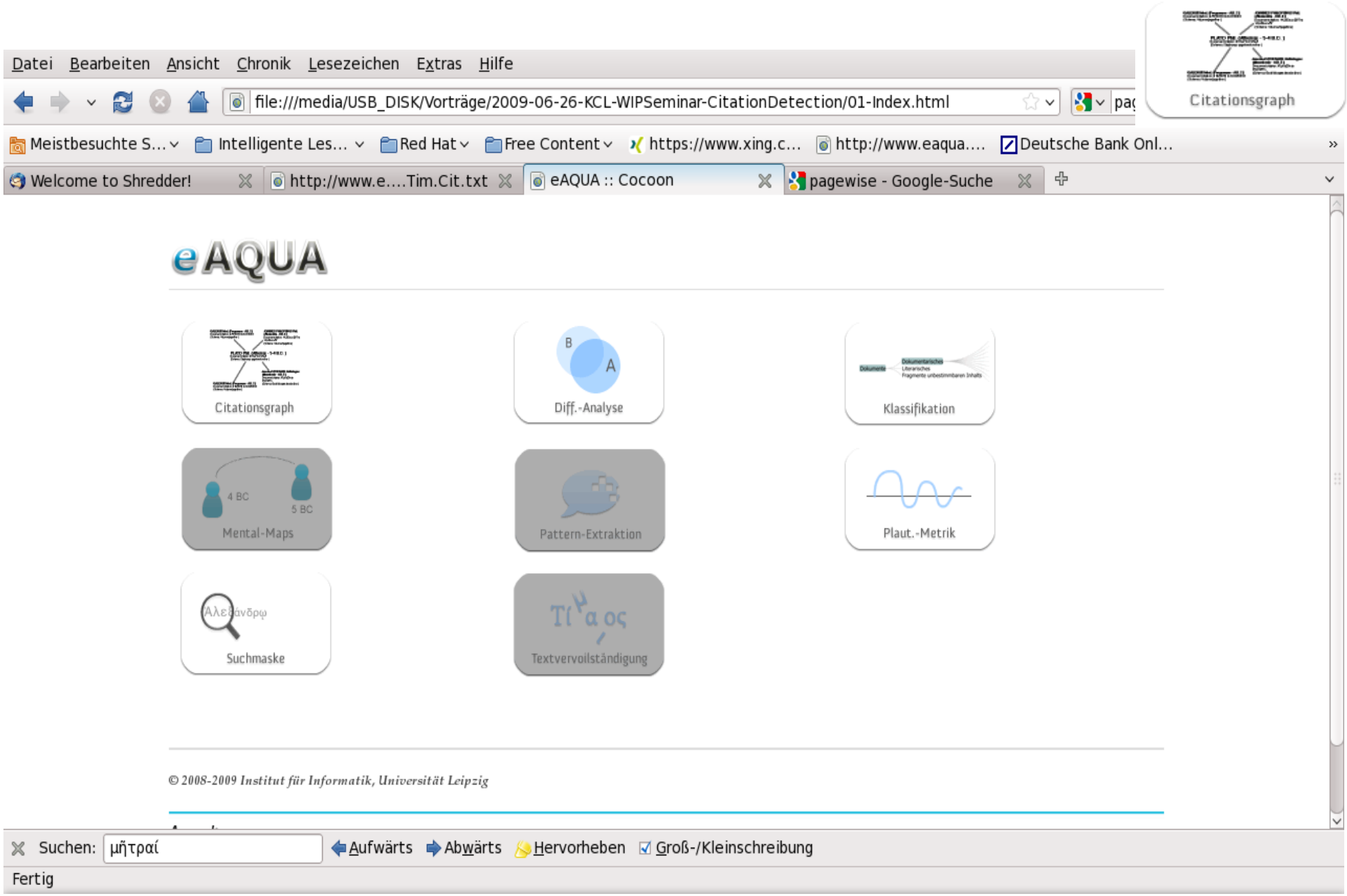
1 V = segment_corpus(C) with v_1, v_2, ..., $v_n \in V$, $\cup v_i = C$ and $v_i \neq v_j$
2 **for each** $v_i \in V$
3 F_i=train_features(v_i);
4 **for each** $v_i \in V$
5 **for each** $f_k \in F_i$
6 $e_i=(v_i,v_j) \in E$=select all v_j containing feature f_k
7 **for each** $e_i \in E$
8 s_i=scoring($e_i=(v_i,v_j) \in E$; F_i; F_j);
9 if($s_i < threshold$){$E=E\backslash\{e_i\}$}

Literal citations:

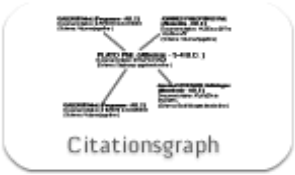
αἱ δ' ἐν ταῖς γυναιξὶν αὖ μῆτραί τε καὶ ὑστέραι λεγόμεναι διὰ τὰ αὐτὰ ταῦτα ζῷον ἐπιθυμητικὸν ...

1. Step: Iterative training of possible n-grams candidates (Training)
   - *αἱ δ'*
   - *αἱ δ' ἐν*
   - *αἱ δ' ἐν ταῖς*
   - *αἱ δ' ἐν ταῖς γυναιξὶν*
   - *...*
2. Removing all n-grams having a smaller frequency of 2 and having less than 5 words (Selection)
3. Removing prefixes and suffixes
4. Creating inverted list for the above detected n-grams (Mapping n-gram to sentence)
5. Collect all sentences having same n-grams (Citation candidates)
6. Compute similarity by word overlap (Dice)

Literal citations: portal I



Literal citations: portal II



Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

file:///media/USB_DISK/Vorträge/2009-06-26-KCL-WIPSeminar-CitationDetection/02-RequestReferences(

Meistbesuchte S... Intelligente Les... Red Hat Free Content https://www.xing.c... http://www.eaqua.... Deutsche Bank Onl...

Welcome to Shredder! | http://www.e....Tim.Cit.txt | eAQUA :: Cocoon | pagewise - Google-Suche

eAQUA

PLATO - 0059 | Timaeus - 031

Results

α. τί τὸ ὂν ἀεί, γένεσιν δὲ οὐκ ἔχον, καὶ τί τὸ γιγνόμενον μὲν ἀεί, ὂν δὲ οὐδέποτε;

Source : PLATO (Atheniensis - 5-4 B.C.): Timaeus

Vide : Cf. et SOCRATICORUM EPISTULAE (0637) Lexicon: Cf. TIMAEUS Sophista Gramm. (2602) Scholia: Cf. SCHOLIA IN PLATONEM (5035)

Publication : Timaeus, ed. J. Burnet, Platonis opera, vol. 4. Oxford: Clarendon Press, 1902 (repr. 1968): St III.17a-92c. (Cod: 24,104: Dialog., Phil.)

Document citation : 27/3a/6 to 28/a/1 (Schema: Stephanus page/section/line)

References : There have been 52 references of this sentence found.

β. ὁ δὴ πᾶς οὐρανὸς _ἢ κόσμος ἢ καὶ ἄλλο ὅτι ποτὲ ὀνομαζόμενος μάλιστ' ἂν δέχοιτο, τοῦθ' ἡμῖν ὠνομάσθω_σκεπτέον δ' οὖν περὶ αὐτοῦ πρῶτον, ὅπερ ὑπόκειται περὶ παντὸς ἐν ἀρχῇ δεῖν σκοπεῖν, πότερον ἦν ἀεί, γενέσεως ἀρχὴν ἔχων οὐδεμίαν, ἢ γέγονεν, ἀπ' ἀρχῆς τινος ἀρξάμενος.

Source : PLATO (Atheniensis - 5-4 B.C.): Timaeus

Vide : Cf. et SOCRATICORUM EPISTULAE (0637) Lexicon: Cf. TIMAEUS Sophista Gramm. (2602) Scholia: Cf. SCHOLIA IN PLATONEM (5035)

Publication : Timaeus, ed. J. Burnet, Platonis opera, vol. 4. Oxford: Clarendon Press, 1902 (repr. 1968): St III.17a-92c. (Cod: 24,104: Dialog., Phil.)

Document citation : 28/1a/2 to 28/1a/7 (Schema: Stephanus page/section/line)

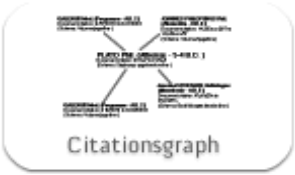
References : There have been 46 references of this sentence found.

γ. τὸ μὲν δὴ νοήσει μετὰ λόγου περι ληπτόν, ἀεὶ κατὰ ταὐτὰ ὄν, τὸ δ' αὖ δόξῃ μετ' αἰσθήσεως ἀλόγου δοξαστόν, γιγνόμενον καὶ ἀπολλύμενον, ὄντως δὲ οὐδέποτε ὄν.

Suchen: μῆτραί | Aufwärts | Abwärts | Hervorheben | Groß-/Kleinschreibung

Fertig

Literal citations: portal III



Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

file:///media/USB_DISK/Vorträge/2009-06-26-KCL-WIPSeminar-CitationDetection/03-ToySampleReference

pagewise

Meistbesuchte S... Intelligente Les... Red Hat Free Content https://www.xing.c... http://www.eaqua.... Deutsche Bank Onl...

Welcome to Shredder! | http://www.e....Tim.Cit.txt | eAQUA :: Cocoon | pagewise - Google-Suche

eAQUA

Author : *0059 - PLATO*
Publication : *031 - Timaeus*

Quotations Citations Map Graph

Found : *6*

Original

αἱ δ' ἐν ταῖς γυναιξὶν αὖ μῆτραί τε καὶ ὑστέραι λεγόμεναι διὰ τὰ αὐτὰ ταῦτα, ζῷον ἐπιθυμητικὸν ἐνὸν τῆς παιδοποιίας, ὅταν ἄκαρπον παρὰ τὴν ὥραν χρόνον πολὺν γίγνηται, χαλεπῶς ἀγανακτοῦν φέρει, καὶ πλανώμενον πάντῃ κατὰ τὸ σῶμα, τὰς τοῦ πνεύματος διεξόδους ἀποφράττον, ἀναπνεῖν οὐκ ἐῶν εἰς ἀπορίας τὰς ἐσχάτας ἐμβάλλει καὶ νόσους παντοδαπὰς ἄλλας παρέχει, μέχριπερ ἂν ἑκατέρων ἡ ἐπιθυμία καὶ ὁ ἔρως συναγαγόντες, οἷον ἀπὸ δένδρων καρπὸν καταδρέψαντες, ὡς εἰς ἄρουραν τὴν μήτραν ἀόρατα ὑπὸ σμικρότητος καὶ ἀδιάπλαστα ζῷα κατασπείραντες καὶ πάλιν διακρίναντες μεγάλα ἐντὸς ἐκθρέψωνται καὶ μετὰ τοῦτο εἰς φῶς ἀγαγόντες ζῴων ἀποτελέσωσι γένεσιν.

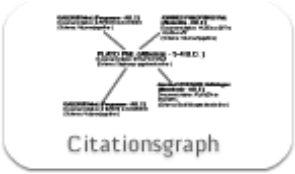
Source : PLATO (Atheniensis - 5-4 B.C.): Timaeus
Vide : Cf. et SOCRATICORUM EPISTULAE (0637) Lexicon: Cf. TIMAEUS Sophista Gramm. (2602) Scholia: Cf. SCHOLIA IN PLATONEM (5035)
Publication : Timaeus, ed. J. Burnet, Platonis opera, vol. 4. Oxford: Clarendon Press, 1902 (repr. 1968): St III.17a-92c. (Cod: 24,104: Dialog., Phil.)
Document citation : 91/1a/7 to 91/3a/5 (Schema: Stephanus page/section/line)

Quotations

α. «αἱ δ' ἐν ταῖς γυναιξὶν αὖ μῆτραί τε καὶ ὑστέραι λεγόμεναι διὰ τὰ αὐτὰ ταῦτα, ζῷον ἐπιθυμητικὸν ἐνὸν τῆς παιδοποιίας, ὅταν ἄκαρπον περὶ τὴν ὥραν χρόνον πολὺν γίγνηται, χαλεπῶς ἀγανακτοῦν φέρει καὶ πλανώμενον πάντῃ κατὰ τὸ σῶμα τὰς τοῦ πνεύματος διεξόδους ἀποφράττον ἀναπνεῖν οὐκ ἐῶν εἰς ἀπορίας τὰς ἐσχάτας

Suchen: μῆτραί | Aufwärts | Abwärts | Hervorheben | Groß-/Kleinschreibung

Fertig



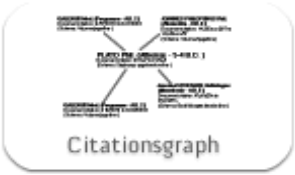
Plato: Timaeus 91b7 ff.

αἱ δ' ἐν ταῖς γυναιξὶν αὖ μῆτραί τε καὶ ὑστέραι λεγόμεναι διὰ τὰ αὐτὰ ταῦτα ζ
ἐπιθυμητικὸν ἑνὸν τῆς παιδοποιίας ὅταν ἄκαρπον παρὰ τὴν ὥραν χρόνον πολὺν γίγνηται
χαλεπῶς ἀγανακτοῦν φέρει καὶ πλανώμενον πάντῃ κατὰ τὸ σῶμα τὰς τοῦ πνεύματος
διεξόδους ἀποφράττον ἀναπνεῖν οὐκ ἐῶν εἰς ἀπορίας τὰς ἐσχάτας ἐμβάλλει καὶ νόσους
παντοδαπὰς ἄλλας παρέχει μέχριπερ ἂν ἑκατέρων ἡ ἐπιθυμία καὶ ὁ ἔρως συναγαγόντες οἷον
ἀπὸ δένδρων καρπὸν καταδρέψαντες ὡς εἰς ἄρουραν τὴν μήτραν ἀόρατα ὑπὸ σμικρότητος
καὶ ἀδιάπλαστα ζῷα κατασπείραντες καὶ πάλιν διακρίναντες μεγάλα ἐντὸς ἐκθρέψωνται καὶ
μετὰ τοῦτο εἰς φῶς ἀγαγόντες ζῴων ἀποτελέσωσι γένεσιν

αἱ δ' ἐν ταῖς γυναιξὶν αὖ μῆτραί τε καὶ ὑστέραι λεγόμεναι διὰ τὰ αὐτὰ ταῦτα ζῷον
ἐπιθυμητικὸν ἑνὸν τῆς παιδοποιίας ὅταν ἄκαρπον περὶ τὴν ὥραν χρόνον πολὺν γίγνηται
χαλεπῶς ἀγανακτοῦν φέρει καὶ πλανώμενον πάντῃ κατὰ τὸ σῶμα τὰς τοῦ πνεύματος
διεξόδους ἀποφράττον ἀναπνεῖν οὐκ ἐῶν εἰς ἀπορίας τὰς ἐσχάτας ἐμβάλλει καὶ νόσους
παντοδαπὰς ἄλλας παρέχει μέχριπερ ἂν ἑκατέρων ἡ ἐπιθυμία καὶ ὁ ἔρως ξυναγαγόντες οἷον
ἀπὸ δένδρων καρπὸν καταδρέψαντες ὡς εἰς ἄρουραν τὴν μήτραν ἀόρατα ὑπὸ σμικρότητος
καὶ ἀδιάπλαστα ζῷα κατασπείραντες καὶ πάλιν διακρίναντες μεγάλα ἐντὸς ἐκθρέψωνται καὶ
μετὰ τοῦτο εἰς φῶς ἀγαγόντες ζῴων ἀποτελέσωσι γένεσιν

περὶ δὲ τῆς μήτρας ὅτι τε ζῷόν ἐστι καὶ αὕτη καὶ τὰ ἀπὸ τοῦ πατρὸς ἐξερχόμενα μόρια
ταῦτα πάλιν λέγει Πλάτων αἱ δ' ἐν ταῖς γυναιξὶν αὖ μῆτραί τε καὶ ὑστέραι λεγόμεναι διὰ τὰ
αὐτὰ ταῦτα ζῷον ἐπιθυμητικὸν ἑνὸν τῆς παιδοποιίας ὅταν ἄκαρπον παρὰ τὴν ὥραν χρόνον
πολὺν γίνηται χαλεπῶς ἀγανακτοῦν φέρει καὶ πλανώμενον πάντῃ κατὰ τὸ σῶμα τὰς τοῦ
πνεύματος διεξόδους ἀποφράττον καὶ ἀναπνεῖν οὐκ ἐῶν εἰς ἀπορίας τὰς ἐσχάτας ἐμβάλλει
καὶ νόσους παντοδαπὰς ἄλλας παρέχει μέχριπερ ἂν ἑκατέρων ἡ ἐπιθυμία καὶ ὁ ἔρως
ξυναγαγόντες οἷον ἀπὸ δένδρων καρπὸν καταδρέψαντες ὡς εἰς ἄρουραν τὴν μήτραν ἀόρατα
ὑπὸ σμικρότητος καὶ ἀδιάπλαστα ζῷα κατασπείραντες καὶ πάλιν διακρίναντες μεγάλα ἐντὸς
ἐκθρέψωνται καὶ μετὰ τοῦτο εἰς φῶς ἀγαγόντες ζῴων ἀποτελέσωσι γένεσιν

Literal citations: Visu



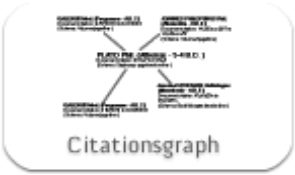
Plato: Timaeus 91b7 ff.

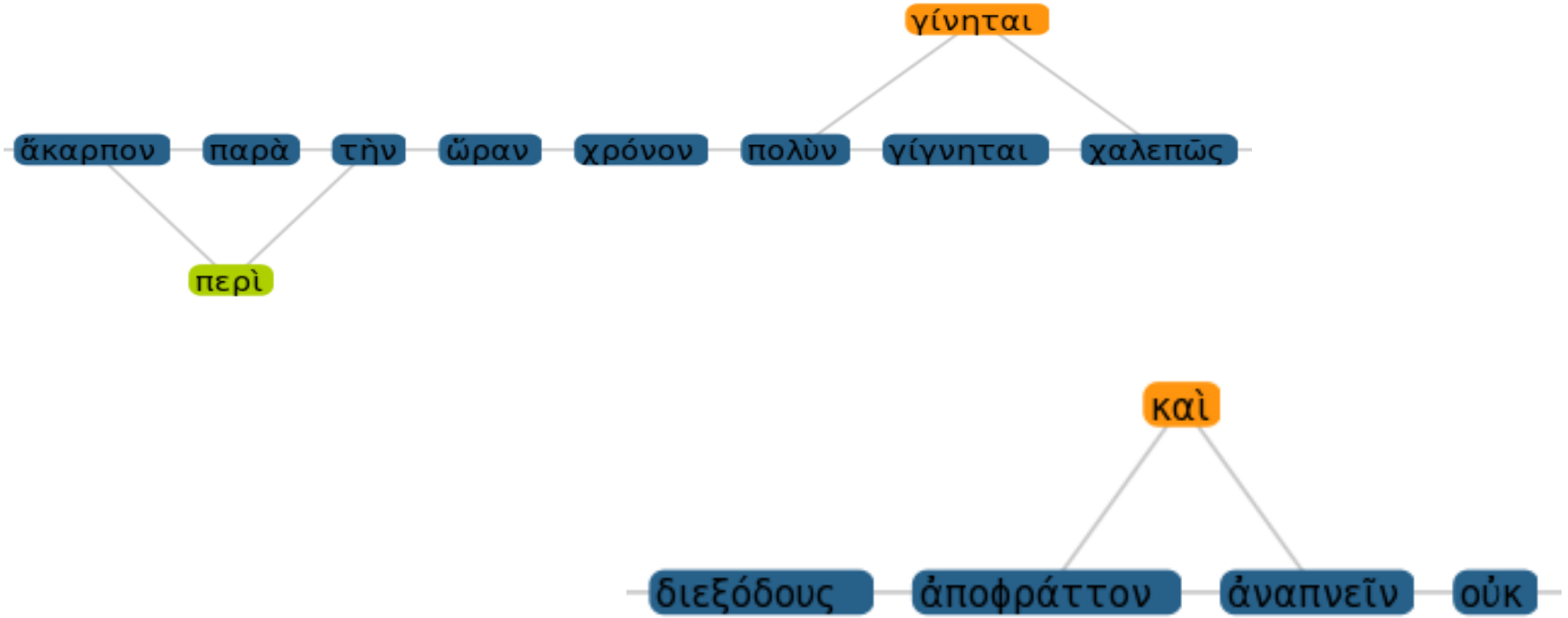
αἱ δ' ἐν ταῖς γυναιξὶν αὖ μῆτραί τε καὶ ὑστέραι λεγόμεναι διὰ τὰ αὐτὰ ταῦτα ζ
ἐπιθυμητικὸν ἐνὸν τῆς παιδοποιίας ὅταν ἄκαρπον παρὰ τὴν ὥραν χρόνον πολὺν γίγνηται
χαλεπῶς ἀγανακτοῦν φέρει καὶ πλανώμενον πάντῃ κατὰ τὸ σῶμα τὰς τοῦ πνεύματος
διεξόδους ἀποφράττον ἀναπνεῖν οὐκ ἐῶν εἰς ἀπορίας τὰς ἐσχάτας ἐμβάλλει καὶ νόσους
παντοδαπὰς ἄλλας παρέχει μέχριπερ ἂν ἑκατέρων ἡ ἐπιθυμία καὶ ὁ ἔρως συναγαγόντες οἷον
ἀπὸ δένδρων καρπὸν καταδρέψαντες ὡς εἰς ἄρουραν τὴν μήτραν ἀόρατα ὑπὸ σμικρότητος
καὶ ἀδιάπλαστα ζῷα κατασπείραντες καὶ πάλιν διακρίναντες μεγάλα ἐντὸς ἐκθρέψωνται καὶ
μετὰ τοῦτο εἰς φῶς ἀγαγόντες ζῴων ἀποτελέσωσι γένεσιν

αἱ δ' ἐν ταῖς γυναιξὶν αὖ μῆτραί τε καὶ ὑστέραι λεγόμεναι διὰ τὰ αὐτὰ ταῦτα ζῷον
ἐπιθυμητικὸν ἐνὸν τῆς παιδοποιίας ὅταν ἄκαρπον περὶ τὴν ὥραν χρόνον πολὺν γίγνηται
χαλεπῶς ἀγανακτοῦν φέρει καὶ πλανώμενον πάντῃ κατὰ τὸ σῶμα τὰς τοῦ πνεύματος
διεξόδους ἀποφράττον ἀναπνεῖν οὐκ ἐῶν εἰς ἀπορίας τὰς ἐσχάτας ἐμβάλλει καὶ νόσους
παντοδαπὰς ἄλλας παρέχει μέχριπερ ἂν ἑκατέρων ἡ ἐπιθυμία καὶ ὁ ἔρως ξυναγαγόντες οἷον
ἀπὸ δένδρων καρπὸν καταδρέψαντες ὡς εἰς ἄρουραν τὴν μήτραν ἀόρατα ὑπὸ σμικρότητος
καὶ ἀδιάπλαστα ζῷα κατασπείραντες καὶ πάλιν διακρίναντες μεγάλα ἐντὸς ἐκθρέψωνται καὶ
μετὰ τοῦτο εἰς φῶς ἀγαγόντες ζῴων ἀποτελέσωσι γένεσιν

περὶ δὲ τῆς μήτρας ὅτι τε ζῷόν ἐστι καὶ αὕτη καὶ τὰ ἀπὸ τοῦ πατρὸς ἐξερχόμενα μόρια
ταῦτα πάλιν λέγει Πλάτων αἱ δ' ἐν ταῖς γυναιξὶν αὖ μῆτραί τε καὶ ὑστέραι λεγόμεναι διὰ τὰ
αὐτὰ ταῦτα ζῷον ἐπιθυμητικὸν ἐνὸν τῆς παιδοποιίας ὅταν ἄκαρπον παρὰ τὴν ὥραν χρόνον
πολὺν γίνηται χαλεπῶς ἀγανακτοῦν φέρει καὶ πλανώμενον πάντῃ κατὰ τὸ σῶμα τὰς τοῦ
πνεύματος διεξόδους ἀποφράττον καὶ ἀναπνεῖν οὐκ ἐῶν εἰς ἀπορίας τὰς ἐσχάτας ἐμβάλλει
καὶ νόσους παντοδαπὰς ἄλλας παρέχει μέχριπερ ἂν ἑκατέρων ἡ ἐπιθυμία καὶ ὁ ἔρως
ξυναγαγόντες οἷον ἀπὸ δένδρων καρπὸν καταδρέψαντες ὡς εἰς ἄρουραν τὴν μήτραν ἀόρατα
ὑπὸ σμικρότητος καὶ ἀδιάπλαστα ζῷα κατασπείραντες καὶ πάλιν διακρίναντες μεγάλα ἐντὸς
ἐκθρέψωνται καὶ μετὰ τοῦτο εἰς φῶς ἀγαγόντες ζῴων ἀποτελέσωσι γένεσιν

Literal citations: Visualisation



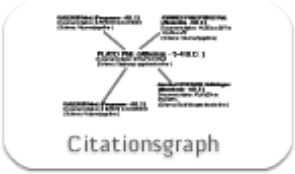


Live demonstration

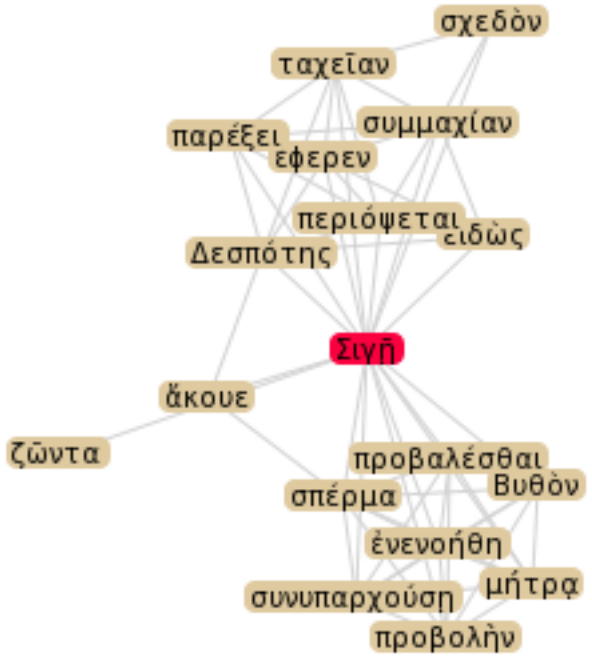
Similarity thresholds for some authors

First result: As time passes, text reuse by later authors becomes much less literal.

Motivation I (stable graph)



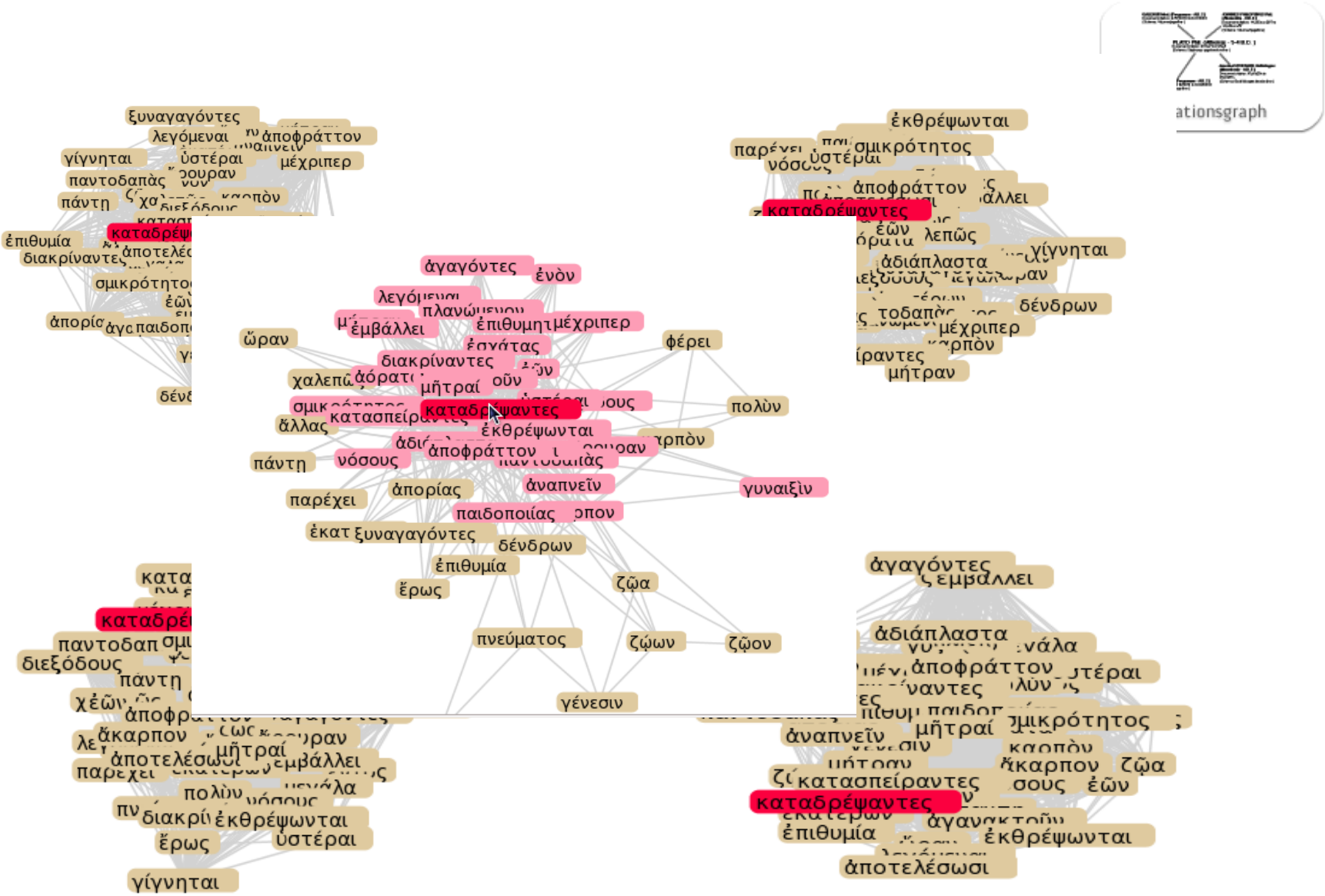
Time slice x



Time slice x+1

- Graph is built based on co-occurrence analysis.

Motivation II (unstable graph)



Wort *καταδρέψαντες (567423)*
Anzahl *3*
Häufigkeitsklasse *20*
Normalisiert gleiche Formen: *καταδρέψαντες (3);*

Kookkurrenzähnliche Formen: *ἐκθρέψωνται (0.9765); ἀποφράττον (0.9708); ἀποτελέσωσι (0.9595); ἀδιάπλαστα (0.9586); κατασπείραντες (0.9581); μῆτραί (0.954); ἀγανακτοῦν (0.8432); διακρίναντες (0.7892); ξυναγαγόντες (0.7889); μέχριπερ (0.573); σμικρότητος (0.5081); παιδοποιίας (0.4973); ὑστέραι (0.4865); ἐνὸν (0.4757); ἐπιθυμητικὸν (0.4757); ἐσχάτας (0.4432); ἐῶν (0.4108); ἀγαγόντες (0.4); ἐμβάλλει (0.4); ἄκαρπον (0.3892); παντοδαπὰς (0.3892); ἀναπνεῖν (0.3892); διεξόδους (0.3784); πλανώμενον (0.3459); δὲ (0.3027); γίγνηται (0.2919); λεγόμεναι (0.2919); ἐπὶ (0.2919); μὲν (0.2811); καὶ (0.2811); αὐτῶν (0.2811); ὥσπερ (0.2811); ἢ (0.2703); κατὰ (0.2703); ἀπορίας (0.2703); τῶν (0.2703); πάλιν (0.2703); ἀρχῆς (0.2595); λοιπὸν (0.2595); πολλῶν (0.2595); ἔξω (0.2595); καθάπερ (0.2595); σώματος (0.2595); ὁμοίως (0.2595); μάλιστα (0.2595); ἤδη (0.2595); ἔτι (0.2595); αὐτοῖς (0.2595); τούτων (0.2595); ἔξωθεν (0.2595); λικμῶσίν (0.1455); κατεσθίουσαι (0.1157); συντετρῆσθαί (0.1031); ἀμύξαντες (0.102);*

Wort *Πλάτων (675)*
Anzahl *8754*
Häufigkeitsklasse *9*
Normalisiert gleiche Formen: *Πλάτων (8754); ΠΛΑΤΩΝ (51); πλάτων (9); πλατῶν (7);*

Kookkurrenzähnliche Formen: *Ἀριστοτέλης (0.45); Πλάτωνος (0.45); Τιμαίῳ (0.41); Σωκράτης (0.39); Φαίδρῳ (0.39); Πλάτωνι (0.38); Πλά (0.37); στοτέλης (0.3316); Πορφύριος (0.33); Φαίδωνι (0.33); Πολιτείᾳ (0.32); Πρόκλος (0.32); Σοφιστῇ (0.32); Ἱπποκράτης (0.31); αὐτὴν (0.31); Παρμενίδης (0.31); Παρμενίδῃ (0.31); Ἰάμβλιχος (0.31); φιλόσοφος (0.3); Τίμαιος (0.3); ἐνταῦθα (0.3); Ἐμπεδοκλῆς (0.3); Φιλήβῳ (0.3); δοκεῖ (0.29); ἕκαστον (0.29); μέν (0.29); οὕτω (0.28); οὐσίας (0.28); τουτέστι (0.28); αἰτίαν (0.28); ἀρχὴν (0.28); ἁπλῶς (0.28); ὥσπερ (0.28); εἶναι (0.28); περὶ (0.28); τὴν (0.28); ἐξηγούμενος (0.28); Πλωτῖνος (0.28); διότι (0.28); εἰπὼν (0.28); ἀνάγκη (0.27);*

Toy sample corpus

1. *Copy from one, it is plagiarism; copy from two, it is research.*
2. *Plagiarism is not the same as copyright infringement.*
3. *Plagiarism is to to copy from one but to copy from two is research.*
4. *The concept of copyright originates with the Statute of Anne (1710) in Great Britain.*
5. *In a legal context, an infringement Frers to the violation of a law or a right.*

1. Step:

Co-occurrence analysis

(copy,from)
(copy,one)
(copy,it's)
(copy,plagiarism)
(copy,research)
...
(plagiarism,from)
(plagiarism,one)
(plagiarism,it's)
(plagiarism,copy)
(plagiarism,research)
...
(plagiarism,copyright)
(plagiarism,infringement)

2. Step:

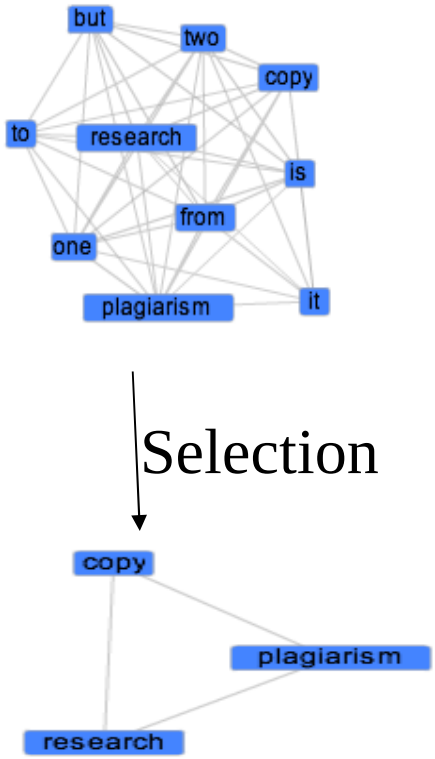
Graph based similarity

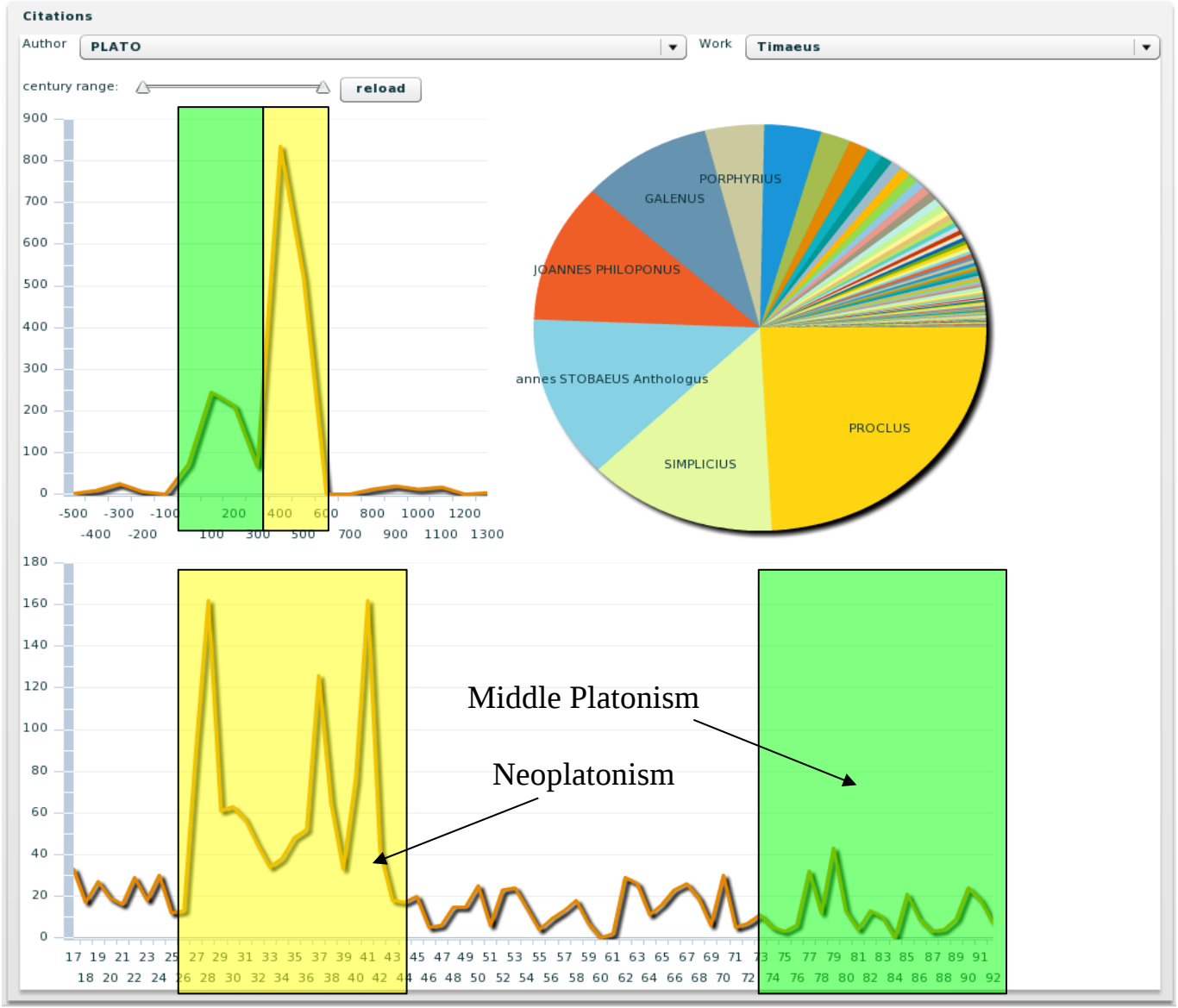
(copy,from, 1.0)
(copy,one, 1.0)
(copy,it's, 1.0)
(copy,plagiarism, 0.8)
(copy,research, 1.0)
...
(plagiarism,from, 0.8)
(plagiarism,one, 0.8)
(plagiarism,it's, 0.8)
(plagiarism,copy, 0.8)
(plagiarism,research, 0.8)
...
(copy,copyright, 0.1)
(copyright, infringement 0.1)

3. Step: Intersection

(copy,from, 1.0)
(copy,one, 1.0)
(copy,it's, 1.0)
(copy,plagiarism, 0.8)
(copy,research, 1.0)
...
(plagiarism,from, 0.8)
(plagiarism,one, 0.8)
(plagiarism,it's, 0.8)
(plagiarism,copy, 0.8)
(plagiarism,research), 0.8
...

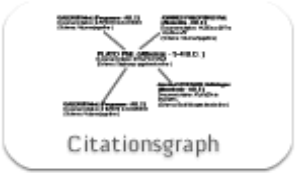
4. Step:

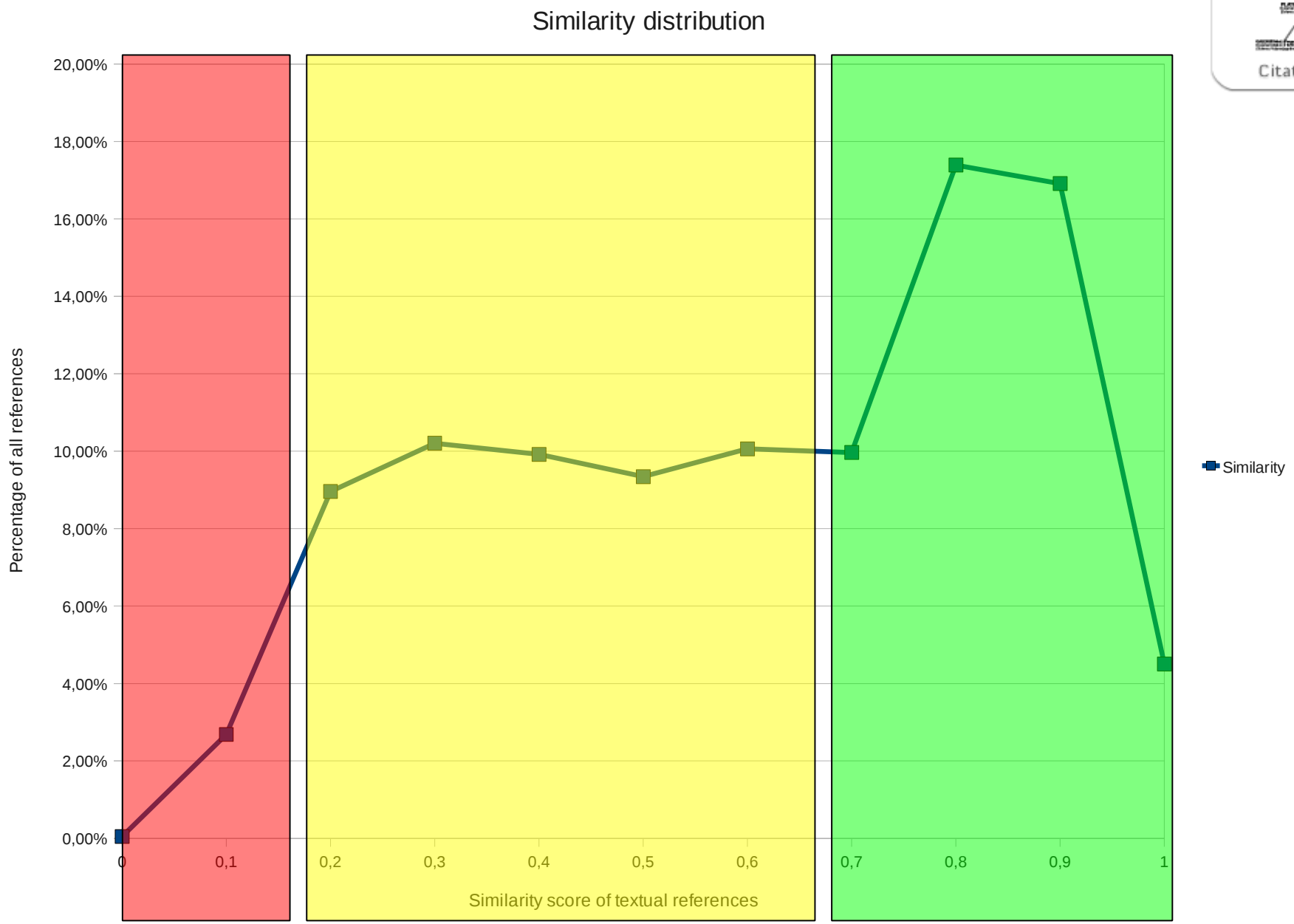




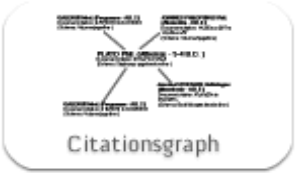
Live demonstration

Literal citations: Similarity thresholds





Some results – problem focused



- What is a good similarity threshold (literal citations)?
  - Dissimilarity vs. Fragment
  - Plato: Low threshold provides good results as well
  - Atthidographers: Poor quality – precision less than 20% --> Rethinking

- Multi word expressions like *King Alexander the Great* (literal citations)

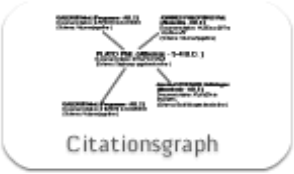
- Phrases
  - „τοῦ Κυρίου ἡμῶν Ἰησοῦ Χριστοῦ“ (Engl.: *Our Lord Jesus Christ*)
  - Again: „We are the people!“

- Editorial references to publications

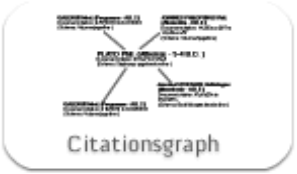
- Works in different editions

We are the people!

Start and end of a reuse

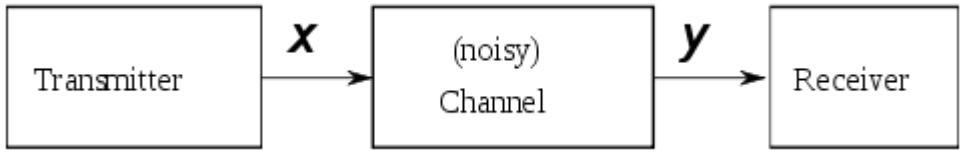


αἱ	*μῆτραί τε καὶ ὑστέραι λεγόμεναι ...*
αἱ δὲ ἐν ταῖς γυναιξὶν	*μῆτραί τε καὶ ὑστέραι λεγόμεναι ...*
αἱ δ' ἐν ταῖς γυναιξὶν αὖ	*μῆτραί τε καὶ ὑστέραι λεγόμεναι ...*
αἱ δ' ἐν ταῖς γυναιξὶ	μῆτραί τε καὶ ὑστέραι λεγόμεναι ...



Textual reuse graph

- *Formal*: Graph G=(V,E) V=collection of vertices, E=collection of edges
- *Simple*: pairwise relations between objects from a certain collection



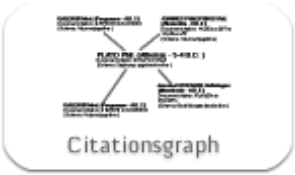
Given: Multiple „Receiver“ in terms of texts in different editions

Modelling a noisy channel to ...

- ... identify the systematic influence of editors
- ... Computer Scientists: formalise noise (what is can be systematically observed)
- ... Classicists: to improve understanding of variants and transmissions of texts to ...

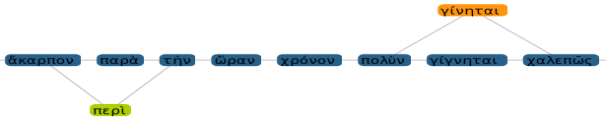
... e. g. extract fragmentary authors (presentation of Monica Berti)

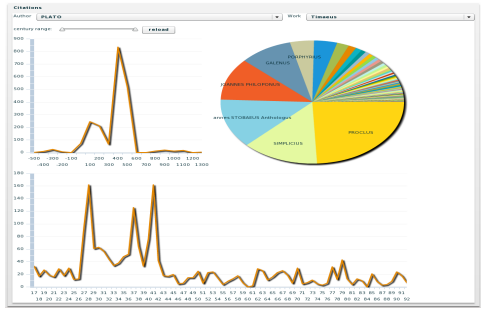
From a bird's eye view



	TLG	PHI5	PHI7	Duke	Greek & Roman corpus	...
TLG						
PHI5						
PHI7						
Duke						
Greek & Roman corpus						
...						

Reuse ratio = Number of inter corpus reuse/Number of intra corpus reuse



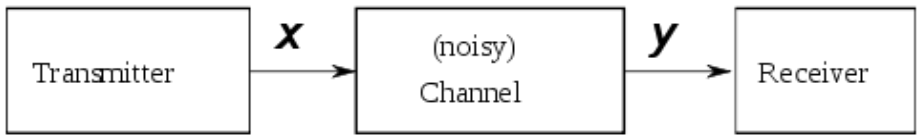


Ancient Greek philology

Ancient history

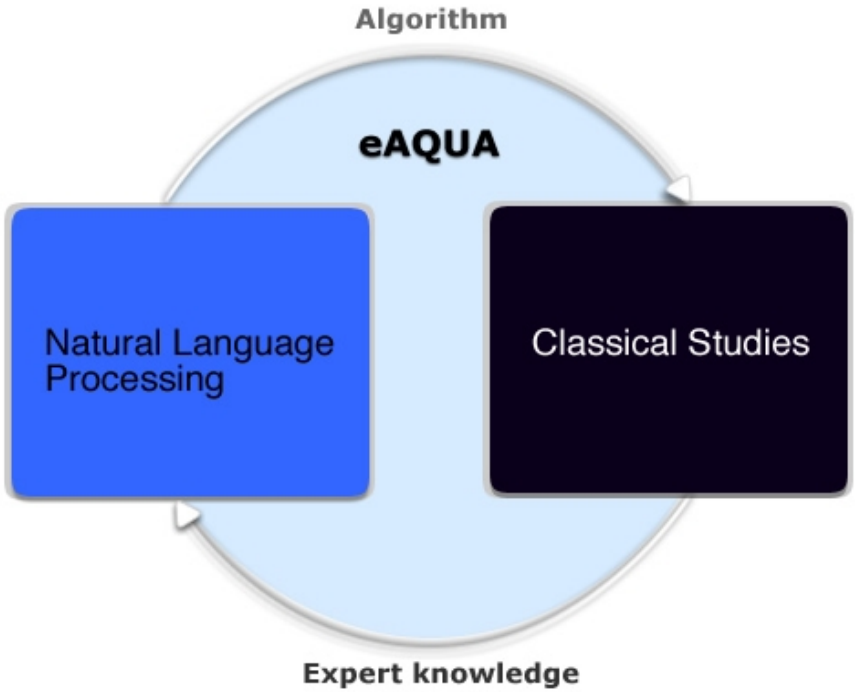
Textual reuse

Computer science



Epigram:

Lots of texts produce much more text mining data which can easily be accessed by a powerful Visual Analytics component.



http://www.eaqua.net

Marco Büchler, Thomas Eckart
Natural Language Processing Group
Department of Computer Science
University of Leipzig
mbuechler@eaqua.net

Annette Geßner
Ancient Greek
Institute of Classical Philology and
Comparative Studies
University of Leipzig
agessner@eaqua.net