

Ψηφιακή ανάλυση κειμένου (Digital Text Analysis)

Η ψηφιακή ανάλυση κειμένου περιλαμβάνει την χρήση λογισμικών για την ανάλυση των δεδομένων που περιέχονται στο κείμενο με στόχο την εξόρυξη χρήσιμων πληροφοριών με ταχύτητα και ακρίβεια που θα δώσουν τις απαντήσεις σε συγκεκριμένα ερευνητικά ερωτήματα.

Τα εν λόγω λογισμικά είναι, κατά κύριο λόγο, εφαρμογές λεξικομετρίας, που μετρούν τις εμφανίσεις και τις συνεμφανίσεις λεκτικών τύπων ή τις συνάψεις, και οδηγούν σε συμπεράσματα μέσω στατιστικών αναλύσεων, οι οποίες έχουν να κάνουν με τη συχνότητα, τη χρονολογική κατάταξη, τα συμφραζόμενα των εμφανίσεων.

Τα λογισμικά ανάλυσης ηλεκτρονικού κειμένου είναι κυρίως εφαρμογές λεξικομετρίας που βασίζονται στην στατιστική ανάλυση της συχνότητας και της χρονολογικής κατάταξης λέξεων και φράσεων ενός κειμένου και της σύγκρισής τους σε σχέση με άλλα κείμενα. Για παράδειγμα:

- **Κατάλογοι συχνότητας λέξεων**
Περιλαμβάνουν τις λέξεις που εμφανίζονται στα κείμενα του σώματος κειμένων με αλφαβητική σειρά ή με σειρά συχνότητας, και μας επιτρέπουν να διαπιστώσουμε στοιχεία για το θέμα και το ύφος των κειμένων, να εντοπίσουμε το ειδικευμένο λεξιλόγιο μιας θεματικής περιοχής ή ενός κειμενικού είδους.
- **Συμφραστικοί πίνακες**
Περιλαμβάνουν τα συμφραζόμενα μιας λέξης ή φράσης, δηλαδή τις συνάψεις της, και μας επιτρέπουν έτσι να διαπιστώσουμε τα ευρύτερα δομικά σύνολα και σχήματα στα οποία χρησιμοποιείται.

Εργαλεία και υπηρεσίες γλωσσικής τεχνολογίας της ΑΠΟΛΛΩΝΙΣ

Η ΑΠΟΛΛΩΝΙΣ, η εθνική ψηφιακή υποδομή ο κατεξοχόν κόμβος για την υποστήριξη και προώθηση των Ψηφιακών Ανθρωπιστικών Επιστημών και της Γλωσσικής Τεχνολογίας στην Ελλάδα, προσφέρει πρόσβαση σε διαδικτυακά εργαλεία και υπηρεσίες γλωσσικής τεχνολογίας, όπως:

- Μορφοσυντακτικούς και σημασιολογικούς αναλυτές.
- Εργαλεία εξαγωγής ορολογίας από κείμενα.
- Εργαλεία συντακτικής ή σημασιολογικής ανάλυσης κειμένων κ.λπ.
- Εργαλεία εξαγωγής ορολογίας από κείμενα.
- Εργαλεία ανεύρεσης των κυρίων ονομάτων (ονοματικών οντοτήτων) ενός πόρου και η κατηγοριοποίησή τους.

Ακολουθούν μερικά από τα εργαλεία που είναι διαθέσιμα από το [clarin:el](#) και το [DARIAH-GR/ΔΥΑΣ](#) ως διαδικτυακές υπηρεσίες.

- **GrNE-Tagger** - Εργαλείο που αναγνωρίζει αυτόματα (βάσει κανόνων) και ταξινομεί Ονοματικές Οντότητες, δηλαδή κύρια ονόματα, στις ακόλουθες πέντε κατηγορίες: PERSON, LOCATION, ORGANIZATION, FACILITY, GPE (Geo-political entity).
- **Μετρητής συχνότητας εμφάνισης λέξεων** - Εργαλείο που επεξεργάζεται κείμενα και αναγνωρίζει τις ελληνικές λέξεις, τις διαχωρίζει από οποιαδήποτε άλλα στοιχεία του κειμένου (αριθμούς, ημερομηνίες, ξένες λέξεις κ.λπ.), και μετράει τη συχνότητα εμφάνισής τους.
- **Εργαλείο Παραγωγής Περιλήψεων** - Το εργαλείο παράγει περιλήψεις, συνδυάζοντας δεδομένα από πολλά κείμενα. Είναι ανεξάρτητο γλώσσας και μπορεί να εφαρμοστεί σε σύνολο δεδομένων.
- **Συντακτικός αναλυτής IEL** - Επεξεργάζεται ελληνικά κείμενα και το αποτέλεσμα της επεξεργασίας είναι συντακτικά αναλυμένο και επισημειωμένο κείμενο.
- **Σύστημα μεταγραφής μεταξύ Ελληνικών και Greeklish** - Σύστημα μεταγραφής της γραφής Greeklish σε Ελληνικά και το αντίστροφο.
- **HGazetteer** - Το εργαλείο HGazetteer εντοπίζει σε κείμενα προκαθορισμένες λέξεις ή φράσεις, οι οποίες περιέχονται σε λίστες (gazetteers) γνωστών λέξεων/φράσεων.

Voyant Tools

Το on-line λογισμικό κειμενικής ανάλυσης Voyant παρέχει 29 εργαλεία στους χρήστες για εντοπισμό της συχνότητας εμφάνισης και την σύγκριση και προσδιορισμό της χρήσης, συσχέτισης και σημασίας σημαντικών όρων / λέξεων σε κείμενα.

Το Voyant είναι ένα διαδικτυακό περιβάλλον για την ψηφιακή ανάλυση κειμένων. Προσφέρει στους χρήστες τις παρακάτω δυνατότητες:

- Εντοπισμό της συχνότητας εμφάνισης σημαντικών όρων / λέξεων σε κείμενα.
- Σύγκριση και προσδιορισμό της χρήσης, συσχέτισης και σημασίας σημαντικών όρων / λέξεων σε κείμενα.

Η συνολική λίστα των εργαλείων που είναι διαθέσιμα στην τρέχουσα έκδοση του λογισμικού και αποτελείται από τα εξής 29 εργαλεία:

ΣΥΧΝΟΤΗΤΑ

Προβολή σε πίνακες και απεικόνιση σε γραφήματα:

- Της συχνότητας εμφάνισης όλων των όρων σε ένα σώμα ή μέσα σε ένα ενιαίο έγγραφο.
- Της συχνότητας εμφάνισης των κορυφαίων λέξεων.
- Της αλλαγής της συχνότητας εμφάνισης των λέξεων.
- Της κατανομής και του τρόπου συγκέντρωσης των όρων.
- Της πορείας της εμφάνισης του κάθε όρου μέσα σε αυτό.

ΣΥΜΦΡΑΖΟΜΕΝΑ

Προβολή σε πίνακες:

- Κάθε εμφάνισης μιας λέξης-κλειδιού μαζί με ένα μικρό μέρος των συμφραζομένων.
- Επαναλαμβανόμενων ακολουθιών λέξεων.
- Των όρων που εμφανίζονται πιο συχνά σε άμεση γειτνίαση με λέξεις-κλειδιά σε ολόκληρο το σώμα.

Παράδειγμα αξιοποίησης του λογισμικού Voyant

Κειμενική ανάλυση της Οδύσσειας για την εξέταση του ρόλου και της σημασίας των δύο βασικών θεών. Αναζήτηση του αριθμού και της συχνότητας εμφάνισης των λημμάτων για τα λήμματα «Δίας» και «Αθηνά».

Αντλούμε / αντιγράφουμε το κείμενο ολόκληρης της Οδύσσειας από την ιστοσελίδα του Περσέα και το εισάγουμε / upload του κειμένου στο δικτυακό περιβάλλον του Voyant ώστε να πραγματοποιήσουμε την ανάλυση του κειμένου ολόκληρης της Οδύσσειας και να εντοπίσουμε τα λήμματα που μας ενδιαφέρουν, που στη συγκεκριμένη περίπτωση είναι ο «Δίας» και η «Αθηνά» και τη συχνότητα εμφάνισής τους, ώστε π.χ. να μελετήσουμε την αντίληψη του ομηρικού ανθρώπου γι' αυτούς τους θεούς.

Αφού φορτωθεί το κείμενο στο λογισμικό, μας εμφανίζεται η ανάλυση της Οδύσσειας. Στο “συννεφάκι” εμφανίζονται οι πιο συχνές λέξεις και τα πιο συνηθισμένα λήμματα με τον αντίστοιχο αριθμό εμφάνισής τους.

Χρησιμοποιώντας το **εργαλείο Terms του Voyant** εντοπίζουμε όλες τις λέξεις της Οδύσσειας, μαζί με τον αριθμό εμφάνισης της κάθε μίας, σε λίστα. Μέσα από το συγκεκριμένο εργαλείο, διαπιστώνουμε ότι το λήμμα «Δίας» συνολικά σε όλες τις μορφές του αναφέρεται 111 φορές, ενώ το «Αθηνά» 161 φορές.

Χρησιμοποιώντας το **εργαλείο Trends του Voyant** παρακολουθούμε την πορεία των δύο λημμάτων μέσα στο κείμενο της Οδύσσειας, μέσω του διαγράμματος που μας παρέχει το εργαλείο. Στο διάγραμμα αυτό εκτός από την απεικόνιση της πορείας των δύο λημμάτων μέσα στο κείμενο, αναγράφονται και τα τμήματα του έργου (Document segments), αλλά και οι σχετικές συχνότητες των λημμάτων. Παρατηρούμε λοιπόν, ότι στο πρώτο και στο τρίτο τμήμα του κειμένου, υπάρχει μία κορύφωση για το λήμμα «ἄθηνη», ενώ το λήμμα «ζεύς» στο πρώτο μέρος του έργου, δεν έχει τόσο μεγάλη συχνότητα. Παρ' όλα αυτά, η παρουσία του είναι συνεχής μέχρι το τέλος. Επίσης, στο δεύτερο κομμάτι του έργου βλέπουμε ένα κοινό σημείο των λημμάτων, ενώ από το πέμπτο μέχρι το έβδομο τμήμα παρουσιάζουν μία κοινή διαδρομή. Γενικά, όμως θα μπορούσαμε να πούμε, ότι υπάρχει μία σχετική παραλληλία στην πορεία τους, που γίνεται πιο εμφανής στα τελευταία τμήματα του έργου, με το λήμμα «ἄθηνη» να έχει τις περισσότερες φορές μεγαλύτερη συχνότητα.

Εργαλεία και υπηρεσίες γλωσσικής τεχνολογίας του TLG

Ψηφιακή βιβλιοθήκη αρχαίων και μεσαιωνικών ελληνικών κειμένων από τον Όμηρο έως και το τέλος της Βυζαντινής Αυτοκρατορίας που επιτρέπει την αναζήτηση στο σώμα των κειμένων και την σύνδεση της κάθε λέξης με μορφολογικά εργαλεία και λεξικά. Στο TLG μπορεί κανείς να βρει αρχικές εκδόσεις κειμένων, σημασίες λέξεων, αντίγραφα κειμένων, σημειώσεις σε κείμενα ή κειμενικούς δείκτες, αλλά και πηγές ανοιχτού περιεχομένου που παρέχουν ελεύθερη πρόσβαση σε πολλές αρχαίες πηγές κειμένων.

Η ψηφιακή βιβλιοθήκη κειμένων αρχαίας ελληνικής λογοτεχνίας / ηλεκτρονικός Θησαυρός (TLG) προσφέρει στους χρήστες τις παρακάτω δυνατότητες:

- Περιήγηση στα κείμενα που είναι διαθέσιμα σε όλη τη βάση δεδομένων του TLG.
- Αναζήτηση λημμάτων σε όλες τις παραλλαγές τους και για συγκεκριμένο συγγραφέα, ιστορική περίοδο, είδος, διάλεκτο ή σε συγκεκριμένα αποσπάσματα κειμένων σε κείμενα και σε σημαντικά λεξικά της αρχαίας ελληνικής λογοτεχνίας, ώστε να κατανοήσουμε καλύτερα τη χρήση τους.
- Εντοπισμό για κάθε σημαντικό όρο / λέξη στα κείμενα των διαθέσιμων στατιστικών (συχνότητα εμφάνισης), μορφολογικών και γραμματικών αναλύσεων.
- Αξιοποίηση των υπερσυνδέσμων στα κείμενα για αυτόματη μεταφορά στο περιβάλλον χρήσης του Περσέα για επιπλέον μορφολογική ανάλυση ή εξαγωγή της αγγλικής μετάφρασης ενός κειμένου ή μιας λέξης.

Η ψηφιακή βιβλιοθήκη του TLG παρέχει εργαλεία:

- Για τη διεξαγωγή απλών ή σύνθετων αναζητήσεων στο πλήρες σώμα του TLG (Full Corpus Search), στη συντομευμένη έκδοσή του (AbridgedTLG), στον Κανόνα των ελλήνων συγγραφέων και έργων (Canon Search), αλλά και στα λεξικά (Lexica).
- Για τη λεξιλογική, μορφολογική έρευνα και γενικά την κειμενική ανάλυση και ερμηνεία.

Παρακάτω μερικά από τα βασικά ψηφιακά εργαλεία του TLG είναι:

- Το εργαλείο Canon Search που επιτρέπει την αναζήτηση με βάση τον συγγραφέα (π.χ. το όνομα του συγγραφέα, τη χρονολογία του, το είδος του, αλλά και τη γεωγραφική προέλευσή του).
- Το εργαλείο Text Search που επιτρέπει αναζήτηση είτε στο πλήρες σώμα του TLG, είτε στη συντομευμένη έκδοσή του, για

μία μορφή ευρετηριασμένης λέξης (word form), για ένα λήμμα ή για μία μη ευρετηριασμένη συμβολοσειρά. Η αναζήτηση μίας ευρετηριασμένης λέξης επιστρέφει όλες τις μορφές της λέξης, ανεξάρτητα από τη γραμματική και την ορθογραφική της παραλλαγή. Η αναζήτηση ενός λήμματος δίνει στους χρήστες όλες τις λεξικές καταχωρήσεις μιας λέξης. Στο πλαίσιο αυτής της αναζήτησης, υπάρχει και η κειμενική αναζήτηση (textual search), που ερευνά στα αρχεία κειμένου του TLG και όχι στο ευρετήριο λέξεων.

- Browse: Επιτρέπει τη μελέτη ενός κειμένου ή την σύγκριση δύο κειμένων, επιλέγοντας την παράλληλη αναζήτηση (parallel browsing).
- Lexica: παρουσιάζει λεξικογραφικές πηγές, μερικές από τις οποίες συνδέονται σε εξωτερικές ιστοσελίδες, όπως αυτή της ψηφιακής βιβλιοθήκης του Περσέα. Η λειτουργία αυτή προσφέρει τη δυνατότητα στους χρήστες, να δουν τις καταχωρήσεις των λέξεων σε μία ποικιλία λεξικών.
- N – Grams: Ένα εργαλείο που επιτρέπει την αντιστοίχιση ή επισήμανση όμοιων φράσεων, ανάμεσα σε έργα του ίδιου ή διαφορετικού συγγραφέα.
- Statistics: Ένα εργαλείο που επιτρέπει την στατιστική ανάλυση για κάθε συγγραφέα, έργο ή λήμμα.
- Vocabulary tools: Υποστηρίζει τη δημιουργία λεξιλογίου με τον εντοπισμό του πιο συχνά χρησιμοποιημένου λήμματος σε έναν συγγραφέα ή σε ένα συγκεκριμένο έργο.

Παράδειγμα αξιοποίησης του TLG

Κειμενική ανάλυση της Οδύσσειας για την αναζήτηση της συχνότητας εμφάνισης του λήμματος «θεός» και εντοπισμός των αποσπασμάτων του κειμένου που το περιέχουν στην πιο συχνή του μορφή.

Χρησιμοποιώντας το **εργαλείο στατιστικής ανάλυσης Statistics του TLG** κάνουμε αναζήτηση του λήμματος «θεός» στον Όμηρο και την Οδύσσεια προκειμένου να εντοπίσουμε τη συχνότητα εμφάνισης. Παρουσιάζεται ένα 'συννεφάκι' με τις λέξεις που συναντάμε συχνά στο έπος αυτό, αλλά και κάποια διαγράμματα με τα λήμματα που εμφανίζονται, κατατάσσοντάς τα σε διάφορες κατηγορίες. Αν εστιάσουμε στο τρίτο διάγραμμα παρατηρούμε παρατηρούμε ότι το λήμμα «θεός» τι ανήκει στα πιο συνηθισμένα λήμματα (Most frequent lemmata), καθώς εμφανίζεται στην Οδύσσεια συνολικά 369 φορές.

Χρησιμοποιώντας το **εργαλείο Text Search του TLG** κάνουμε αναζήτηση των λεξικών καταχωρήσεων / εμφανίσεων του λήμματος «θεός» όλες τις μορφές της λέξης, ανεξάρτητα από τη γραμματική και την ορθογραφική της παραλλαγή. Επιλέγοντας ενδεικτικά από τη λίστα με τους πιθανούς τύπους (SELECTION) του λήμματος «θεός», αυτούς με τον μεγαλύτερο αριθμό συχνότητας εμφάνισης στην Οδύσσεια, ώστε να μας παρουσιαστούν μόνο εκείνα τα αποσπάσματα που περιέχουν τη χρήση τους.

Εργαλεία και υπηρεσίες γλωσσικής τεχνολογίας του Perseus Digital Library

Ψηφιακή βιβλιοθήκη αρχαιοελληνικών και λατινικών κειμένων που επιτρέπει την αναζήτηση στο σώμα των κειμένων, την πρόσβαση σε πρωτότυπες και μεταφρασμένες εκδόσεις, λεξικά και την χρήση εργαλείων στατιστικής και μορφολογικής ανάλυσης.

Η ψηφιακή βιβλιοθήκη του Περσέα επιτρέπει στους χρήστες τις παρακάτω δυνατότητες:

- Αναζήτηση πρωτογενών πηγών του αρχαιοελληνικού κόσμου και του ρωμαϊκού πολιτισμού που περιλαμβάνουν έργα πολλών αρχαίων ελλήνων συγγραφέων και των κυριότερων εκπροσώπων της λατινικής λογοτεχνίας.
- Αναζήτηση βιβλιογραφιών, ερμηνευτικών υπομνημάτων και εκδόσεων που παράγονται από και για κλασικιστές και μεταδεδομένων για συγγραφείς της λατινικής και ελληνικής λογοτεχνίας.
- Λεξιλογικές αναλύσεις των κειμένων (συχνότητα εμφάνισης λημμάτων) και συσχέτισής τους με αρχαιολογικές πηγές με χρήση ψηφιακών εργαλείων.

Παρακάτω μερικά από τα βασικά ψηφιακά εργαλεία του Περσέα είναι:

- Το εργαλείο μελέτης λέξεων (Word Study Tool) που βοηθά τους χρήστες να αναλύσουν τις μορφές της ελληνικής και λατινικής γλώσσας.
- Το εργαλείο λεξιλογίου (Vocabulary tool) που ορίζει κάθε μία από τις λέξεις στο κείμενο και απαριθμεί τη συχνότητα λέξεων. Το εργαλείο αυτό είναι σημαντικό για τη δημιουργία και παροχή λεξιλογικών λιστών.
- Το εργαλείο αναζήτησης (Lookup Tool) που επιτρέπει τη γενική αναζήτηση με τη χρήση λέξεων – κλειδιών και φράσεων και την ειδική αναζήτηση που αφορά τη συχνότητα λέξεων.
- Το εργαλείο λεξικού (Dictionary Tool) που αφορά την αναζήτηση λημμάτων σε ελληνικά και λατινικά λεξικά.
- Το εργαλείο της μορφολογικής ανάλυσης ελληνικών και λατινικών λέξεων (Greek & Latin Morphological Analysis).
- Το εργαλείο εύρεσης παραδειγμάτων ενός ή περισσότερων λέξεων που εμφανίζονται είτε σε ελληνικά, είτε σε λατινικά κείμενα (Greek Words in Context και Latin Words in Context).
- Το εργαλείο αναζήτηση λημμάτων (Search for Lemmatized Words in Greek and Latin).

Παράδειγμα αξιοποίησης του Περσέα

Κειμενική ανάλυση της Οδύσσειας για το λήμμα «θεός», αναζήτηση ερμηνείας (λεξικό) και της συχνότητας εμφάνισης του λήμματος και εντοπισμός των αποσπασμάτων του κειμένου που το περιέχουν.

Αναζήτηση της συχνότητας εμφάνισης του λήμματος «θεός» στον Όμηρο και στις ραψωδίες της Οδύσσειας στο λογισμικό του Περσέα, ώστε να μελετήσουμε τα αποτελέσματα που αυτός θα μας δώσει και να εξετάσουμε το κατά πόσο συνάδουν με αυτά του TLG.

Χρησιμοποιώντας το **εργαλείο μελέτης λέξεων Word Study Tool του Περσέα** εντοπίζουμε τη γραμματική αναγνώριση του λήμματος και τη σημασία του. Επιπλέον μας δίνεται η επιλογή να δούμε την καταχώρησή του στα λεξικά που διαθέτει ο Περσέας.

Χρησιμοποιώντας το **εργαλείο Word frequency statistics του Περσέα** ανακτούμε στατιστικές πληροφορίες για τη συχνότητα της λέξης «θεός» στον Όμηρο και στις ραψωδίες της Οδύσσειας. Μέσα λοιπόν, από τα αποσπάσματα που μας δίνονται με τις επισημασμένες χρήσεις του λήμματος «θεός», παρατηρούμε ότι και ο Περσέας επιβεβαιώνει την παρουσία του λήμματος με τους περισσότερους πιθανούς του τύπους, “συμφωνώντας” με το TLG στο ότι ο αριθμός της συχνότητάς εμφάνισής του είναι το 368. Παράλληλα, είναι δυνατόν να αναζητήσουμε (search) τη χρήση του λήμματος αυτού σε όλα τα σημεία από τις ραψωδίες της Οδύσσειας και τα αντίστοιχα τμήματα των στίχων τους με δυνατότητα να αναπτύξουμε το κάθε τμήμα από την εκάστοτε ραψωδία και να δούμε ολόκληρο το αρχαίο κείμενο, που εμπριέχει το σημείο ή τα σημεία με τη χρήση του λήμματος, που και σε αυτήν την περίπτωση υπογραμμίζεται για τον εύκολο εντοπισμό της. Επίσης, μπορούμε να συνδεθούμε με τις δύο αγγλικές μεταφράσεις για το έργο της Οδύσσειας, αλλά και με ολόκληρο το πρωτότυπο κείμενό της.

Τα προηγούμενα παραδείγματα αφορούσαν σε κειμενική ανάλυση της Οδύσσειας για την μελέτη του ομηρικού θείκου στοιχείου αξιοποιώντας τις δυνατότητες που προσφέρει η ψηφιακή βιβλιοθήκη του TLG (Thesarus Lingue Graecae), αλλά και το διαδικτυακό περιβάλλον του Περσέα.και το λογισμικό του του Voyant.

- Περσέα: Χρησιμοποιείται συνήθως για την εύρεση πρωτογενών πηγών,
- Voyant: Χρησιμοποιείται συνήθως για τη σύγκριση σημαντικών όρων και για τον εντοπισμό της σημασίας μεμονωμένων όρων ή όρων που συσχετίζονται με άλλους (συμφραζόμενα).

- TLG: Χρησιμοποιείται συνήθως για τη μελέτη μετρικών, συντακτικών και υφολογικών στοιχείων.
- Για την εμφάνιση συχνότητας λέξεων, για τον εντοπισμό της οποίας αξιοποιείται συνήθως το Voyant και το TLG,
- Για τη λεξιλογική και μορφολογική έρευνα αξιοποιείται συνήθως ο Περσέας και το TLG

Ωστόσο, πρέπει να επισημανθεί ότι αυτά τα ερευνητικά μεθοδολογικά στοιχεία, δεν αναιρούν, άλλες δυνατότητες που διαθέτει το TLG, ο Περσέας και το Voyant, ούτε περιορίζουν το πεδίο χρήσης τους. Δεν σημαίνει επίσης, ότι για τον σκοπό που αξιοποιείται το ένα, δεν μπορεί να χρησιμοποιηθεί το άλλο, δεδομένου ότι περιλαμβάνουν αρκετές όμοιες λειτουργίες. Ο κάθε ερευνητής προσάρμοσε το εκάστοτε ψηφιακό εργαλείο στις ανάγκες της έρευνάς του ή ανέδειξε εκείνα τα στοιχεία του, που ο ίδιος θεώρησε πιο σημαντικά.

Ανάλυση σωμάτων κειμένων

Ανάλυση σωμάτων κειμένων (σύνολα κειμένων προς ανάλυση, συγκροτημένα και μορφοποιημένα με βάση συγκεκριμένα κριτήρια και σε σχέση τις ερευνητικές ανάγκες)

Η ψηφιακή κειμενική ανάλυση είναι ιδιαίτερα χρήσιμη για την διαχείριση και ανάλυση του μεγάλου όγκου δεδομένων που περιέχονται στα σώματα λογοτεχνικών έργων και την διενέργεια συγκρίσεων, που διαφορετικά θα ήταν ανέφικτες καθώς το μεγάλο εύρος τους καθιστά αδύνατη την χειρωνακτική εξαγωγή αποτελεσμάτων.

Στα ηλεκτρονικά σώματα κειμένων τα λογισμικά κειμενικής ανάλυσης είναι κυρίως εφαρμογές:

1. Θεματικής μοντελοποίησης που βασίζονται σε στατιστικά μοντέλα για τον εντοπισμό των θεμάτων που διατρέχουν μια συλλογή κειμένων.
2. Λεξικομετρίας που βασίζονται στην στατιστική ανάλυση της συχνότητας και της χρονολογικής κατάταξης λέξεων και φράσεων ενός κειμένου και της σύγκρισής τους σε σχέση με άλλα κείμενα. Για παράδειγμα:
 - Κατάλογοι συχνότητας λέξεων
Περιλαμβάνουν τις λέξεις που εμφανίζονται στα κείμενα του σώματος κειμένων με αλφαβητική σειρά ή με σειρά συχνότητας, και μας επιτρέπουν να διαπιστώσουμε στοιχεία για το θέμα και το ύφος των κειμένων, να εντοπίσουμε το ειδικευμένο λεξιλόγιο μιας θεματικής περιοχής ή ενός κειμενικού είδους.
 - Συμφραστικοί πίνακες
Περιλαμβάνουν τα συμφραζόμενα μιας λέξης ή φράσης, δηλαδή τις συνάψεις της, και μας επιτρέπουν έτσι να διαπιστώσουμε τα ευρύτερα δομικά σύνολα και σχήματα στα οποία χρησιμοποιείται.