

ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ

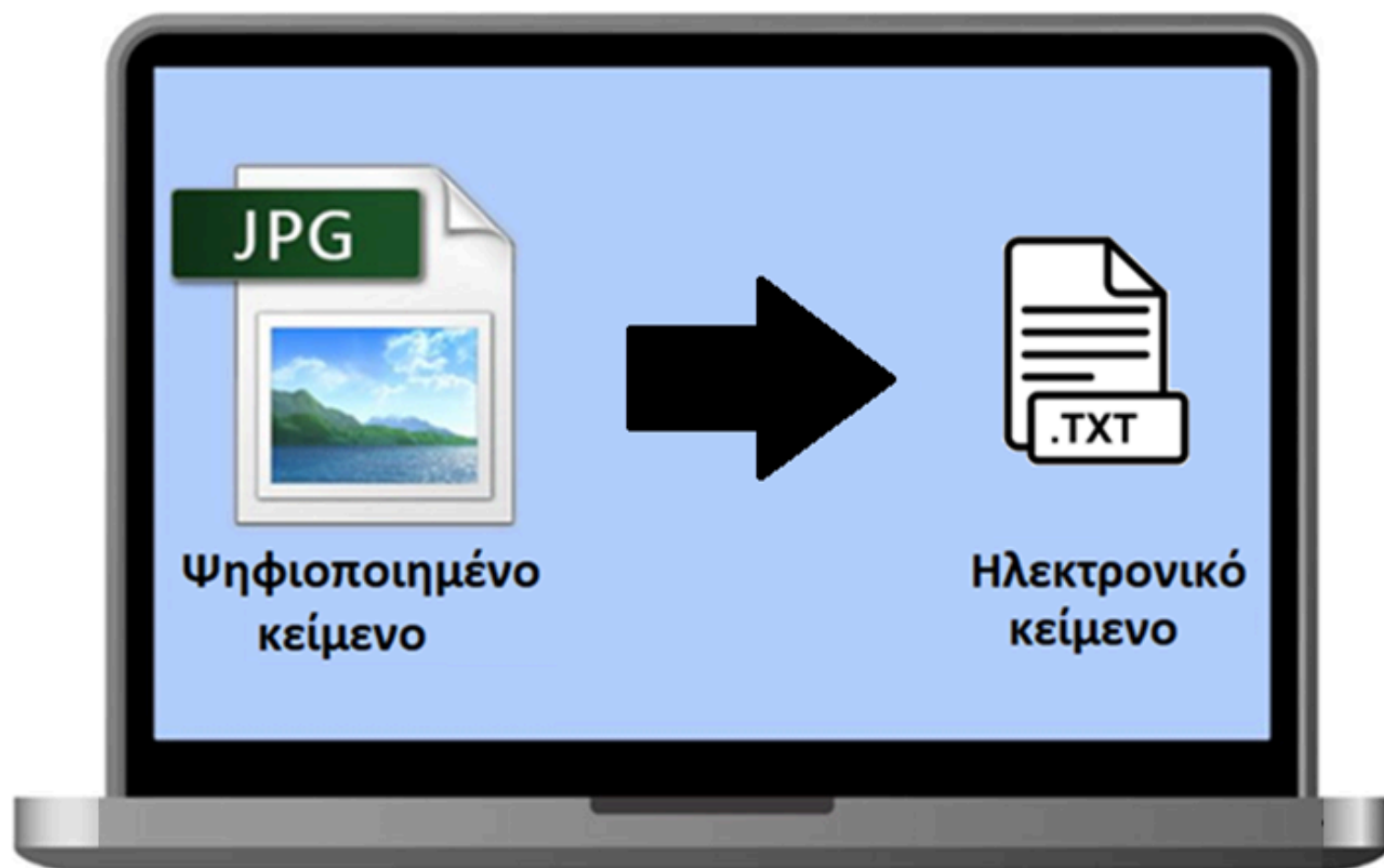
Κατηγορίες ψηφιακών εργαλείων επεξεργασίας χειρόγραφων κειμενικών τεκμηρίων:

- **Transcription** – Μεταγραφή / Παραγωγή ηλεκτρονικού κειμένου.
- **Analysis** – Κειμενική ανάλυση ηλεκτρονικού κειμένου
- **Encoding & Markup** - Κωδικοποίηση / σήμανση ηλεκτρονικού κειμένου.
- **Visualisation** - Οπτικοποίηση δεδομένων
- **Edition** – Έκδοση



Ψηφιακή μεταγραφή (*Transcription*)

Ψηφιακή μεταγραφή χειρογράφου είναι η μετατροπή του ψηφιοποιημένου κειμένου (digitized text) σε ηλεκτρονική μορφή (digital text) η οποία ξεπερνά τη δισδιάστατη εγγραφή πληροφοριών και καθιστά εφικτή την ψηφιακή ανάλυσή του.



ΙΛΙΑΔΟΣ ΑΛΦΑ
 Ἥμιν ἄεϊδε θεὰ Πηληϊάδεω Ἀχιλῆος
 οὐλομένην, ἣ μυρὶ Ἀχαιοῖς ἄλγε' ἔθηκε,
 πολλὰς δ' ἰφθίμους ψυχὰς Ἄϊδι προΐαψεν
 ἠρώων, αὐτοὺς δὲ ἐλώρια τεύχε' κύνεσσιν

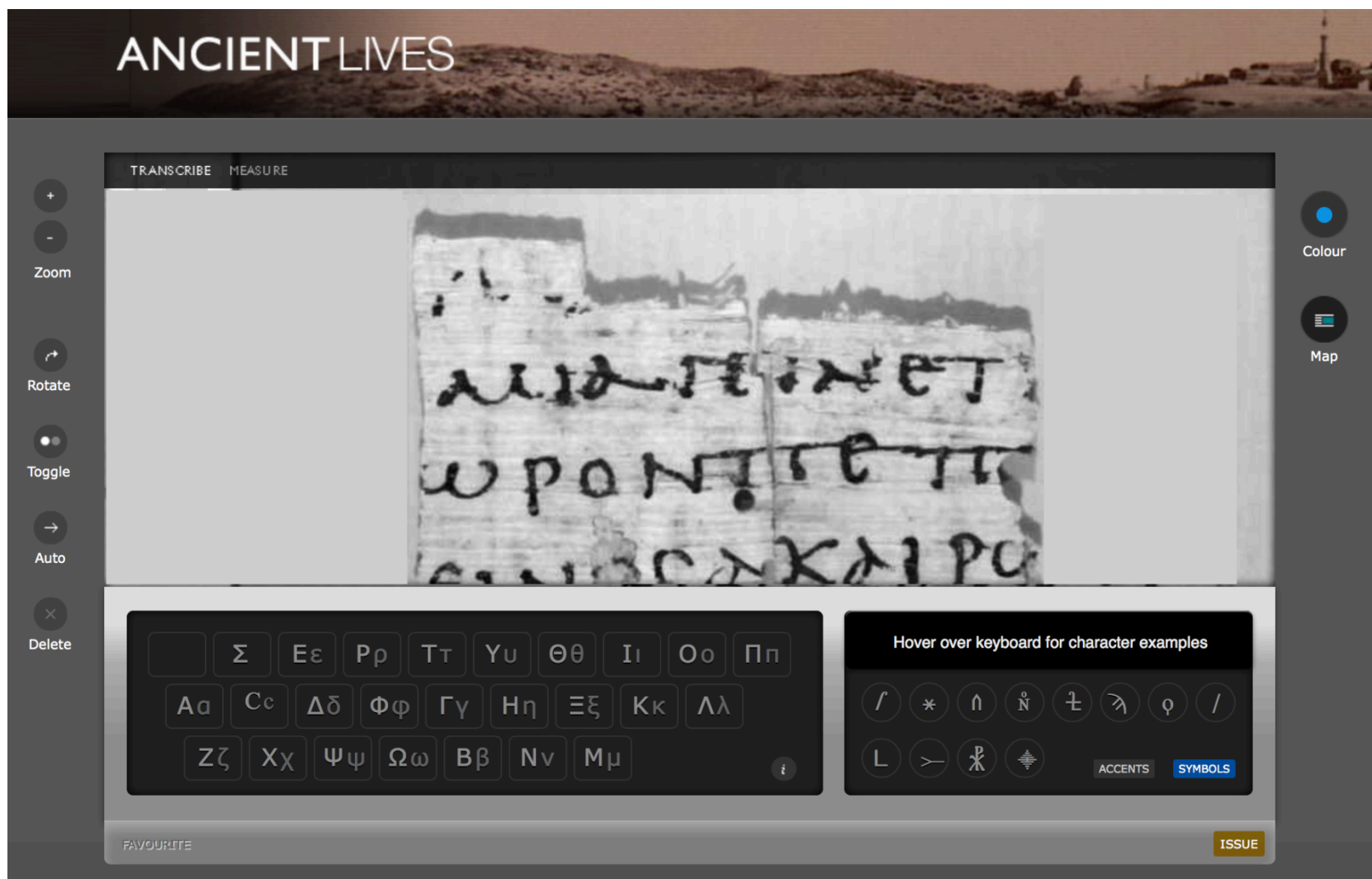
ΙΛΙΑΔΟΣ Α

Μῆνιν ᾄειδε θεὰ Πηληϊάδεω Ἀχιλῆος
 οὐλομένην, ἣ μυρὶ Ἀχαιοῖς ἄλγε' ἔθηκε,
 πολλὰς δ' ἰφθίμους ψυχὰς Ἄϊδι προΐαψεν
 ἠρώων, αὐτοὺς δὲ ἐλώρια τεύχε' κύνεσσιν

Χειρόγραφος κώδικας του Προσιμίου της Ιλιάδας, μεταγραφή κειμένου και μετάφραση.
 Ο κώδικας αυτός (10ος αι. μ.Χ) βρίσκεται στη Μαρκεσιανή Βιβλιοθήκη της Βενετίας.

Μεταγραφή με δακτυλογράφηση

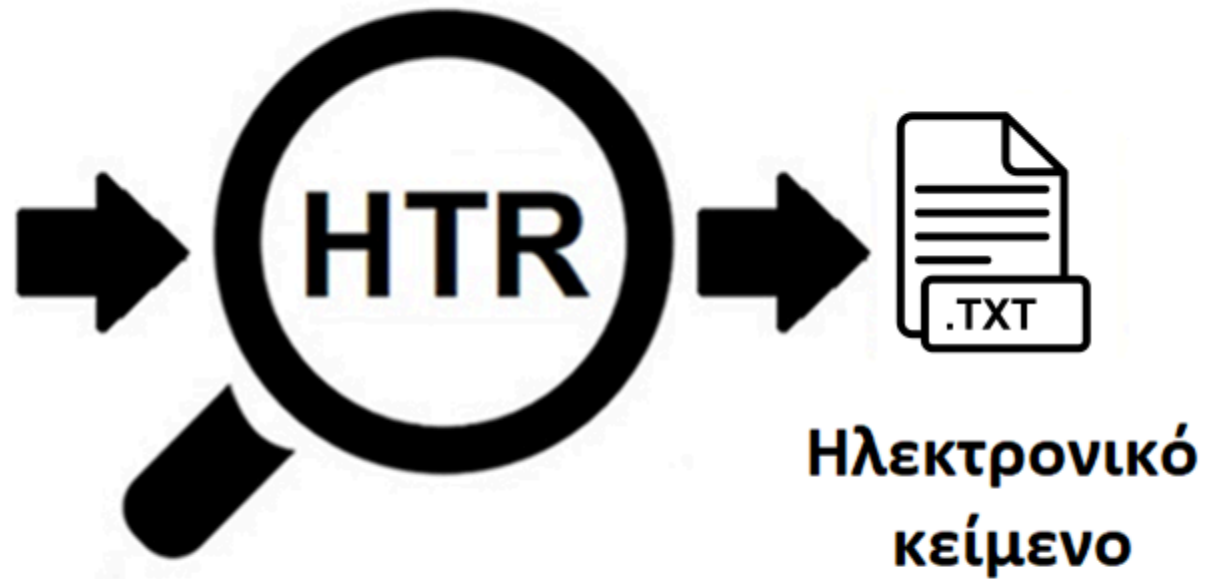
Παράδειγμα: <https://www.ancientlives.org/>



Οπτική σάρωση και αναγνώριση χειρόγραφων χαρακτήρων



**Ψηφιοποιημένο
κείμενο**



Αναζήτηση και εντοπισμός λέξεων (KeyWord Spotting)

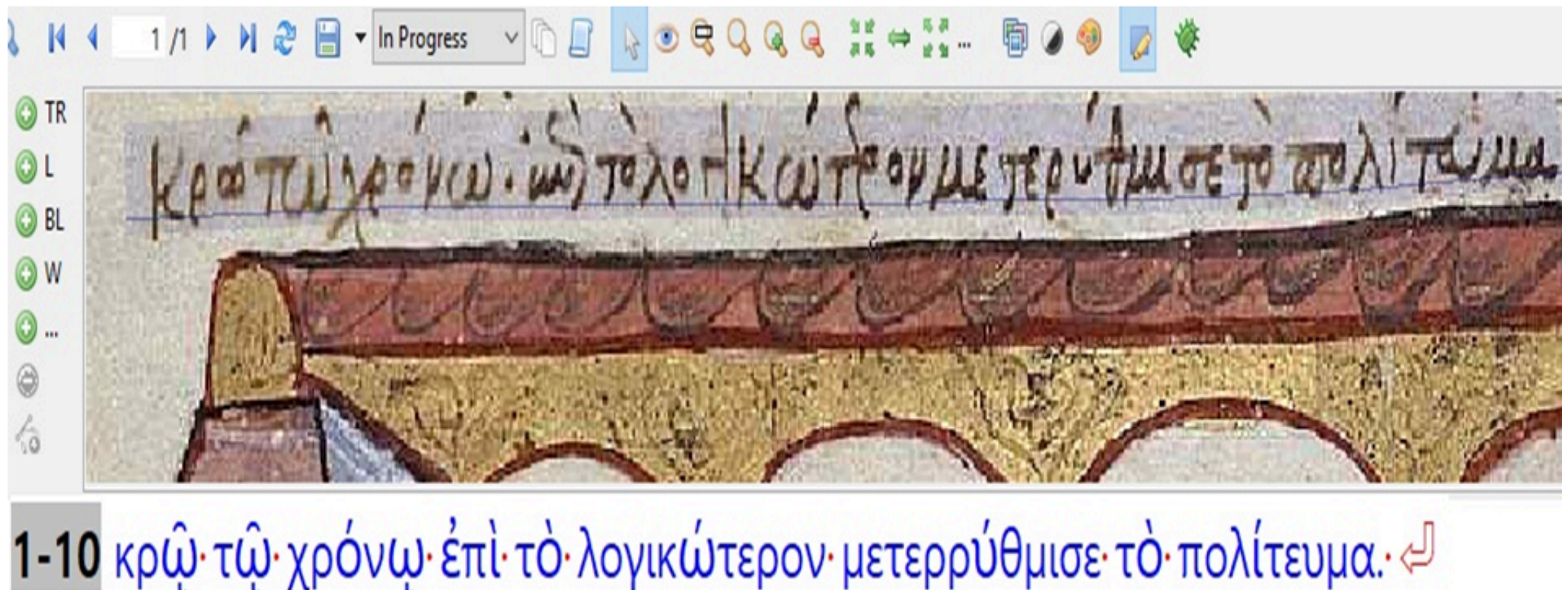
Οι μέθοδοι αναζήτησης λέξεων σε χειρόγραφα χρησιμοποιούν εικόνες λέξεων-κλειδιών (query keyword images) ως πρότυπα, τις οποίες προσπαθούν να εντοπίσουν απευθείας επάνω στις εικόνες των ιστορικών εγγράφων. Βασίζονται στην κατάτμηση των κειμένων σε γραμμές, λέξεις ή χαρακτήρες.



Αναγνώριση χειρόγραφων χαρακτήρων με μηχανική μάθηση (HTR)

Παράδειγμα: <http://transkribus.eu/>

Βασίζεται στην τεχνητή νοημοσύνη (νευρωνικά δίκτυα). Οι χρήστες πρώτα καταχωρούν την μεταγραφή ενός μέρους του ψηφιοποιημένου κειμένου, κατόπιν το λογισμικό μαθαίνει να αναγνωρίζει τους χαρακτήρες και μετά ολοκληρώνει την μεταγραφή αυτόματα με εντυπωσιακή ακρίβεια ανεξάρτητα από τη γλώσσα ή το στυλ γραφής!



Αυτόματη Αναγνώριση και ταυτοποίηση Γραφέα Χειρογράφων (AWI)

Η ταυτοποίηση και χρονολόγηση των χειρογράφων είναι ζωτικής σημασίας για την αξιοποίηση του περιεχομένου τους.

Οι αλγόριθμοι αναγνώρισης προτύπων για την σύγκριση των συμβόλων της αλφαβήτου βασίζονται στην κωδικοποίηση και ομαδοποίηση ενός συνόλου ειδικών και διακριτών χαρακτηριστικών τεχνοτροπίας γραφής και γραφολογικής φύσης.

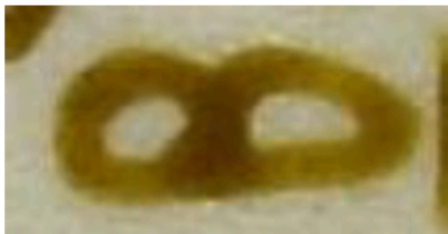
Οι μεθοδολογίες περιλαμβάνουν την:

- Απομόνωση με κατάτμηση από τη συνολική εικόνα του κειμένου (επιγραφής ή παπύρου) των μεμονωμένων υλοποιήσεων των συμβόλων της αλφαβήτου.
- Εξαγωγή των περιγραμμάτων τους.
- Υπολογισμός, βέλτιστη προσαρμογή και σύγκριση της καμπυλότητας των περιγραμμάτων κάθε συμβόλου κατά ζεύγη.

Εξαγωγή περιγραμμάτων

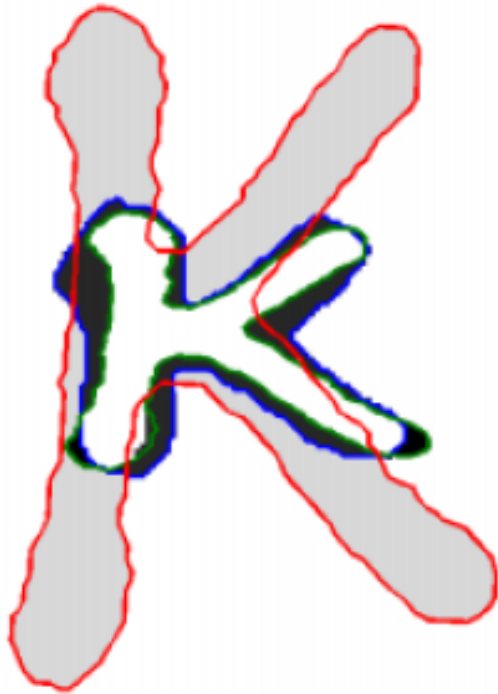
Η διαδικασία εξαγωγής περιγραμμάτων των υλοποιήσεων διαφόρων συμβόλων της αλφαβήτου περιλαμβάνει:

- Εξαγωγή της υλοποίησης από την εικόνα του κειμένου.
- Μαύρισμα του σώματος του γράμματος.
- Εξαγωγή του περιγράμματος της υλοποίησης του γράμματος.



Βέλτιστη προσαρμογή της καμπυλότητας του περιγράμματος

Παρακάτω η απεικόνιση της διαδικασίας προσαρμογής και αντιστοίχισης κατά την εφαρμογή της στα περιγράμματα δύο υλοποιήσεων του γράμματος 'κ', οι οποίες ανήκουν σε διαφορετικά κείμενα.



Αρχική βέλτιστη τοποθέτηση των περιγραμμάτων



Οι ενδιάμεσες παραμορφώσεις του περιγράμματος

ΜΠΛΕ: Η καμπύλη η οποία χρησιμοποιείται σαν αναφορά

ΚΟΚΚΙΝΟ: Η αρχική θέση της καμπύλης την οποία επιχειρούμε να προσαρμόσουμε και οι ενδιάμεσες παραμορφώσεις

ΠΡΑΣΙΝΟ: η τελική θέση βέλτιστης προσαρμογής της καμπύλης

Σύγκριση της καμπυλότητας κατά ζεύγη

Δυο υλοποιήσεις του γράμματος “λ”



Ταίριασμα

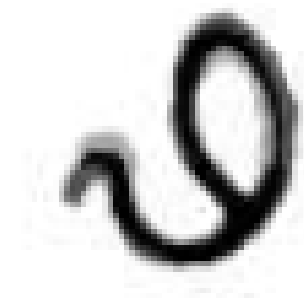


Διαφορετικός γραφέας

Δυο υλοποιήσεις του γράμματος “θ”



Ταίριασμα



Ίδιος γραφέας

Αποκατάσταση κατεστραμμένων κειμένων σε επιγραφές

Παράδειγμα: Pythia

