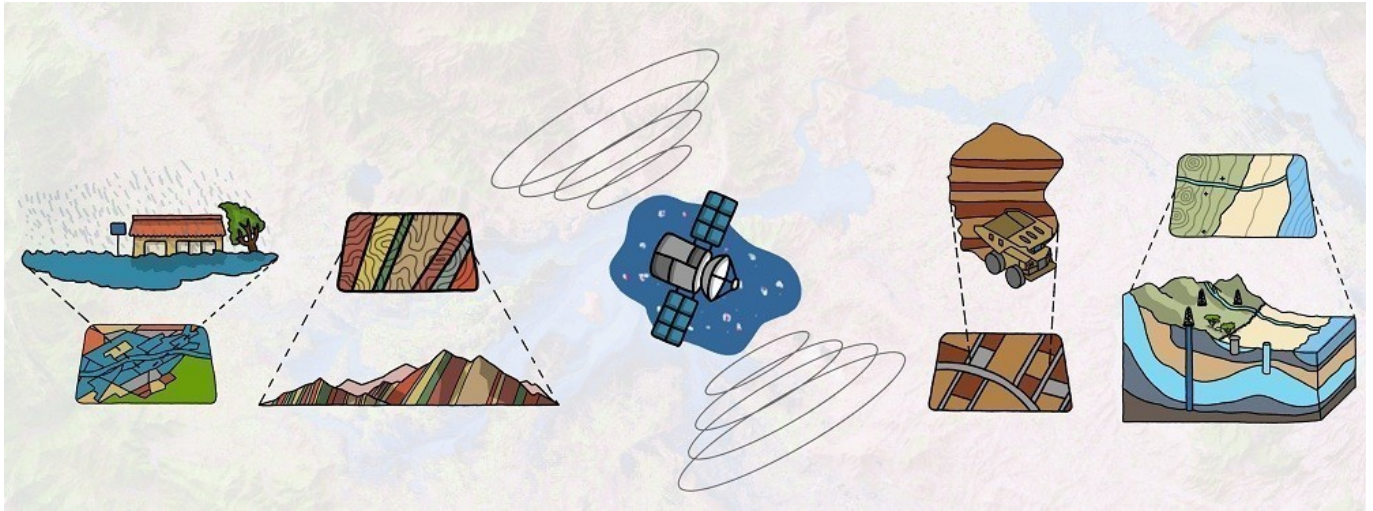


AI: Εφαρμογή αναγνωρίζει και ψηφιοποιεί ιστορικά βιβλία γραμμένα σε πολυτονικό

[/ Γενικά Θέματα](#) / [Ειδήσεις και Ανακοινώσεις](#) / [Επιστήμες, Τέχνες & Πολιτισμός](#)



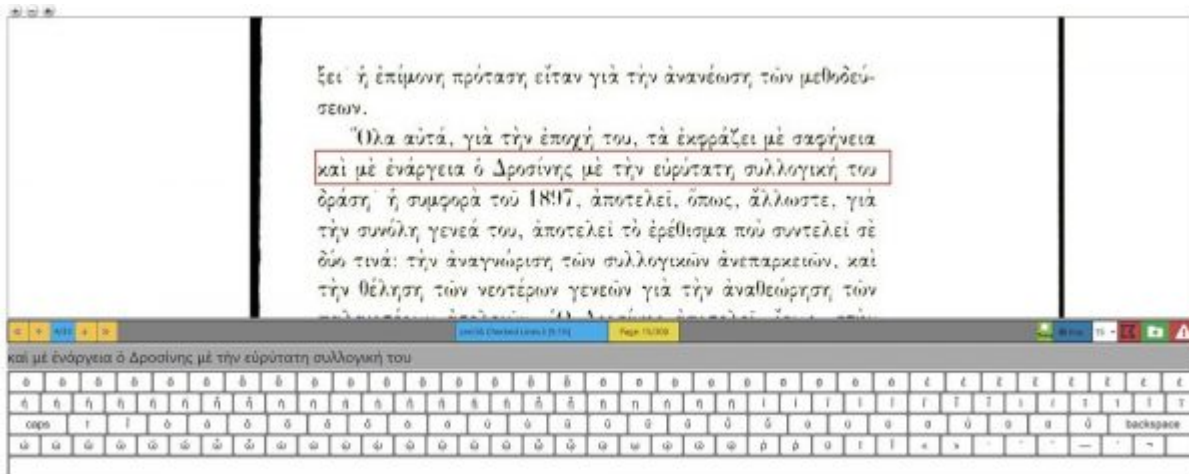
Αυτό που δεν είχε γίνει στο παρελθόν είναι κάποιο εργαλείο που να μπορεί να αναγνωρίσει τα παλιά χειρόγραφα

Μια έξυπνη πλατφόρμα τεχνητής νοημοσύνης που στοχεύει στην ψηφιακή απεικόνιση και διαχείριση κειμένων ιστορικών ελληνικών βιβλίων, γραμμένων σε πολυτονικό σύστημα, ανέπτυξε η ομάδα Ψηφιακής Επεξεργασίας Εγγράφων του Εργαστηρίου Υπολογιστικής Ευφυΐας του Εθνικού Κέντρου Έρευνας Φυσικών Επιστημών «Δημόκριτος».

Η εφαρμογή παρουσιάστηκε στο Athens Science Festival, στην Τεχνόπολη του Δήμου Αθηναίων.

Η πλατφόρμα αναπτύχθηκε στο πλαίσιο του έργου «reBook», που υλοποιείται σε συνεργασία με τον Σύλλογο Προς Διάδοσιν Ωφελίμων Βιβλίων (ΣΩΒ) και την εταιρεία Innews, στο πλαίσιο του ΕΣΠΑ 2014-2020.

Στόχος του έργου είναι η ανάπτυξη νέων τεχνικών και μεθοδολογιών αναγνώρισης κειμένων, κυρίως γραμμένων σε πολυτονικό σύστημα, για την επιστημονική τεκμηρίωση της πολιτιστικής κληρονομιάς.



Εικόνα της εφαρμογής τεχνητής νοημοσύνης / Εργαστήριο Υπολογιστικής Ευφυΐας του Δημόκριτου

Βιβλία από τις αρχές του 20ού αιώνα

Με τη βοήθεια της εφαρμογής ερευνητές του Δημόκριτου ψηφιοποιούν και επανεκδίδουν ψηφιακά, περίπου 100 βιβλία από τα αρχεία του ΣΩΒ, τα οποία χρονολογούνται από τις αρχές του 20ου αιώνα και μετά.

Ανάμεσα σε αυτά βρίσκονται το σύγγραμμα του Αδαμάντιου Κοραή «Επιστολαί προς τον Πρωτοψάλτην», που εκδόθηκε το 1911 και επανεκδόθηκε το 1959. Επίσης, το βιβλίο «Η ελληνικότης των νομών Προύσης και Σμύρνης» του Παντελή Κοντογιάννη, που πρωτοεκδόθηκε το 1919.

Στην εφαρμογή ανεβαίνουν εικόνες σκαναρισμένων βιβλίων και στη συνέχεια με τη βοήθεια της τεχνητής νοημοσύνης αναγνωρίζεται το κείμενο της κάθε εικόνας, ακόμα και αν είναι γραμμένο σε πολυτονικό σύστημα.

Επίσης, στο στάδιο της προεπεξεργασίας της κάθε εικόνας, η εφαρμογή διορθώνει τα προβλήματα που μπορεί να έχει η εικόνα του βιβλίου, από το στραβό σκανάρισμα μέχρι τη διόρθωση ξεθωριασμένων σελίδων.

ΤΡΑΓΩΔΙΑ ΣΑΙΚΣΠΕΙΡΟΥ

ΟΘΕΛΛΟΣ

ΜΕΤΑΦΡΑΣΘΕΙΣΑ ΕΚ ΤΟΥ ΑΓΓΛΙΚΟΥ

ΥΠΟ

ΔΗΜΗΤΡΙΟΥ ΒΙΚΕΛΑ



ΕΚΔΟΣΙΣ ΝΕΑ ΕΠΙΔΙΟΡΘΩΜΕΝΗ



ΕΝ ΑΘΗΝΑΙΣ

ΕΚ ΤΩΝ ΚΑΤΑΣΤΗΜΑΤΩΝ «ΑΝΔΡΕΟΥ ΚΟΡΟΜΗΛΑ» ΚΑΙ «ΚΟΡΑΗ»
ΑΝΕΣΤΗ ΚΩΝΣΤΑΝΤΙΝΙΔΟΥ

1885

Η σημασία της αναγνώρισης του πολυτονικού από την εφαρμογή

«Υπάρχει ένας τεράστιος όγκος βιβλίων που δεν είναι διαθέσιμα ψηφιακά και θέλουμε να τα κάνουμε διαθέσιμα, να βγάλουμε στην επιφάνεια τεκμήρια, τα οποία είναι στα ντουλάπια και τα ράφια βιβλιοθηκών. Οπότε ο στόχος μας είναι ιστορικά βιβλία να φτάσουν στο ευρύ κοινό και τους ερευνητές», εξηγεί στο ΑΠΕ-ΜΠΕ η Κατερίνα Χριστοπούλου, υποψήφια διδάκτορας Οικολογίας Τοπίου και επιστημονική συνεργάτιδα του «Δημόκριτου».

Εξηγώντας την αξία της εφαρμογής η κ. Χριστοπούλου επισημαίνει ότι «δεν βλέπουμε απλά ένα pdf με την εικόνα της σελίδας, αλλά η εικόνα έχει από πίσω την οπτική αναγνώριση χαρακτήρων (OCR), οπότε ο αναγνώστης μπορεί να χρησιμοποιήσει τμήματα του αρχείου ή να κάνει αναζήτηση μέσα σε αυτό».

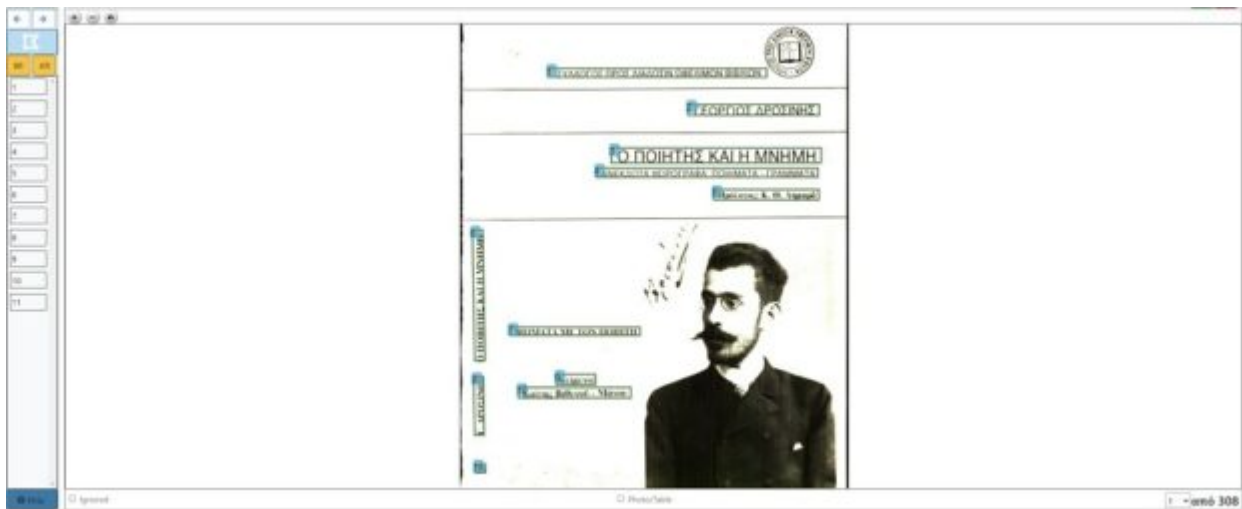
Η μεγάλη διαφορά όμως της εφαρμογής έγκειται «στην ανάγνωση του πολυτονικού συστήματος».

Ψηφιοποίηση 150 εκδόσεων του Σέξπιρ

Δεν είναι η πρώτη φορά που το Εργαστήριο Υπολογιστικής Ευφυΐας του Ινστιτούτου Πληροφορικής και Τηλεπικοινωνιών του Δημόκριτου ασχολείται με την επεξεργασία και αναγνώριση ιστορικών εγγράφων.

Σε αντίστοιχο έργο που υλοποίησε σε συνεργασία με το Πανεπιστήμιο Κύπρου προχώρησε στην ψηφιοποίηση με την ίδια μέθοδο πολυτονικών κειμένων από περίπου 150 εκδόσεις έργων του Σέξπιρ στα ελληνικά.

Πρόκειται για μεταφράσεις που υπέγραψαν σπουδαίοι λογοτέχνες, όπως ο Κωνσταντίνος Καβάφης, ο Κωνσταντίνος Θεοτόκης και ο Δημήτριος Βικέλας, και οι οποίες προέρχονται μεταξύ άλλων από τις συλλογές της Βιβλιοθήκης της Βουλής, της Εθνικής Βιβλιοθήκης και του Ελληνικού Λογοτεχνικού και Ιστορικού Αρχείου.



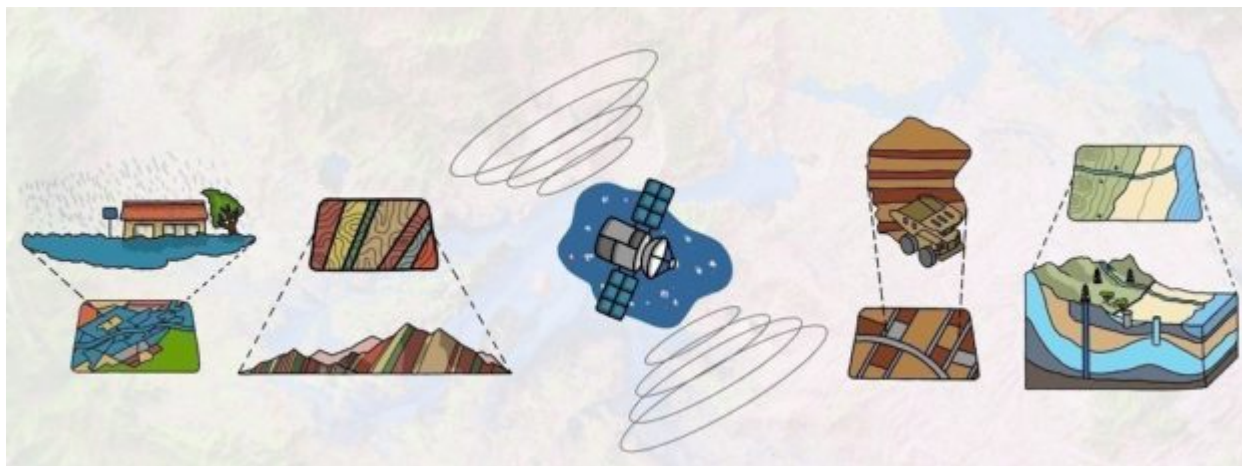
Χειρόγραφα, το μεγάλο στοίχημα

Ένα από τα στοιχήματα που έχει βάλει το Εργαστήριο είναι η ψηφιακή απεικόνιση χειρογράφων.

«Η αναγνώριση χειρογράφων στα σύγχρονα κείμενα έχει προχωρήσει πολύ. Αυτό που δεν είχε γίνει στο παρελθόν είναι κάποιο εργαλείο που να μπορεί να αναγνωρίσει τα παλιά χειρόγραφα και ιδίως τα ελληνικά πολυτονικά», εξηγεί στο ΑΠΕ-ΜΠΕ ο υπεύθυνος του Εργαστηρίου, Βασίλης Γάτος.

Αυτή την περίοδο βρίσκεται σε εξέλιξη έργο σε συνεργασία με την Τράπεζα της Ελλάδας για την αναγνώριση και επεξεργασία, και πάλι με τη βοήθεια τεχνητής νοημοσύνης, των χειρογράφων πρακτικών του ΔΣ της Τράπεζας της περιόδου 1928-1988. Το αρχείο αριθμεί περίπου 30.000 σελίδες.

«Είναι πολύ δύσκολο το πρόβλημα των χειρογράφων, αλλά κάτι που μας βοηθάει σε αυτή την περίπτωση είναι ότι οι γραφείς είναι συγκεκριμένοι όλα αυτά τα χρόνια, οπότε για κάθε γραφέα έχουμε χιλιάδες σελίδες και αυτό μας βοηθάει στο θέμα της εκπαίδευσης του συστήματος», τονίζει ο κ. Γάτος.



Σε παλιότερο έργο της, η ομάδα συνεργάστηκε με το Ίδρυμα της Μονής του Όρους Σινά με σκοπό την ανάπτυξη τεχνολογιών για αναζήτηση πληροφορίας απευθείας από τις εικόνες των χειρογράφων της Μονής. Στο πλαίσιο του έργου αναλύθηκαν και αναγνωρίστηκαν περισσότερες από 100.000 σελίδες ιστορικών χειρογράφων.

Σημειώνεται ότι στο «Athens Science Festival» θα παρουσιαστεί η δουλειά μιας ακόμα ερευνητικής ομάδας του ίδιου Εργαστηρίου του Δημόκριτου.

Η ομάδα «AI4GEO» θα εξηγήσει πώς η παρατήρηση της Γης μέσω εφαρμογών τεχνητής νοημοσύνης «μεταμορφώνεται» σε ένα γιγάντιο άγρυπνο μάτι που μας βοηθά να ανακαλύψουμε πετρώματα και κοιτάσματα, να καταγράψουμε φυσικές καταστροφές ή να παρακολουθήσουμε τις εξελισσόμενες ανθρωπιστικές κρίσεις.

ot.gr