

Περιγραφή της πλατφόρμας Transkribus



Βασίλης Γάτος

Διευθυντής Ερευνών



Εργαστήριο Υπολογιστικής Ευφυΐας
Ινστιτούτο Πληροφορικής & Τηλεπικοινωνιών
Εθνικό Κέντρο Έρευνας Φυσικών Επιστημών
«Δημόκριτος»

Αγία Παρασκευή, Αθήνα

<http://www.iit.demokritos.gr/~bgat>



Institute of Informatics and Telecommunications – NCSR “DEMOKRITOS”

Η πλατφόρμα Transkribus – Εισαγωγή

- Το Transkribus είναι ένα expert tool. Είναι λογισμικό με πλούσια λειτουργικότητα και πολλά χαρακτηριστικά, έχει σχεδιαστεί για τις ανάγκες των χρηστών που “ξέρουν τι θέλουν να κάνουν και πώς να το κάνουν”
- Σκοπός του είναι να υποστηρίξει τον χρήστη να δημιουργήσει υψηλού επιπέδου επιστημονική μεταγραφή των χειρόγραφων (και τυπωμένων) εγγράφων χρησιμοποιώντας τεχνολογίες αιχμής με τον πιο αποτελεσματικό και φιλικό προς τον χρήστη τρόπο.
- Είναι ένα εργαλείο σε εξέλιξη (ενσωματώνει ήδη πολλά χαρακτηριστικά, όμως είναι πολλά ακόμα για να γίνουν)
- Χρειάζεται χρόνος και υπομονή για να το εξερευνήσετε πλήρως.
- Μπορείτε να βοηθήσετε και εσείς (bug report, feature request buttons)
- Φιλοξενείται από το Digitisation and Digital Preservation group (DEA) at the University of Innsbruck
- Χρηματοδότηση: European Commission, H2020 Project READ (2015-2019) και FP7 Project tranScriptorium (2013-2015)

Technical Infrastructure

- Server hosting & DB administration: ZID
- New HP Blade center since Feb 2017
 - 16 machines, 16 cores & 64 GB RAM each
- Leo 3 cluster
 - HPC (1944 cores, 24 GB RAM per node, 3 nodes with Nvidia Tesla M2090 GC)

Stats (6.3.2017)

users: 5595

documents: 10879

images 300352

tags: abbrev: 23910, date: 8332

person: 7498, place: 5624

jobs started: 48838

CITlab HTR models trained (since middle of 2016): 40

Η πλατφόρμα Transkribus – Εγκατάσταση

- **Βήμα 1:** Registration στο <https://transkribus.eu/Transkribus/> (θα χρειαστεί προσωπικό email)
- **Βήμα 2:** Ενεργοποίηση του λογαριασμού από το email που θα λάβετε
- **Βήμα 3:** Download και αποσυμπίεση (unzip) της εγκατάσταση εφαρμογής (για Windows, Linux ή Mac) από την ίδια σελίδα.
- **Βήμα 4:** Έναρξη της εφαρμογής: .exe (Windows), .command (Mac) ή .sh (Linux)

Η πλατφόρμα Transkribus – Είσοδος

Transkribus v1.1.2 (27_03_2017_14:23)

Search current document...

Server Overview Layout Metadata Tools

Login

Jobs on server... Versions of page...

Manage collections... User activity...

Recent documents...

Collections

0 / 0

ID	Title	N-Pages
----	-------	---------

Items per page: 100

Line based Region: 0 / 0 HTR suggestions CATTI

Η πλατφόρμα Transkribus – Είσοδος

Transkribus v1.1.2 (27_03_2017_14:23)

Server Overview Layout Metadata Tools

Logout bgat.ncsr@gmail.com

Jobs on server... Versions of page...

Manage collections... User activity...

Recent documents...

Collections

bgat.ncsr@gmail.com Collection (5566, 0)

ID	Title	N-Pages
17928	Test Document	3

Items per page: 100

Συλλογή για δοκιμές
με 1 έγγραφο 3 σελίδων

Η πλατφόρμα Transkribus – Άσκηση 1

- 1.1 Δείτε όλες τις σελίδες του εγγράφου της δοκιμαστικής συλλογής
- 1.2 Κάντε zoom-in, zoom-out, scroll σε κάθε σελίδα
- 1.3 Δείτε την ιεραρχία των περιοχών από το Layout Panel
- 1.4 Δείτε το αναγνωρισμένο κείμενο, ψάξτε και διορθώστε λάθη
- 1.5 Αναζητήστε κάποια λέξη στο έγγραφο (π.χ. is*)

Η πλατφόρμα Transkribus – Άσκηση 2

- 2.1 Δημιουργία νέας συλλογής
- 2.2 Εισαγωγή των αρχείων materials/unit7/bentham* σε νέο έγγραφο της συλλογής που δημιουργήσατε ([Jeremy Bentham](#))
- 2.3 Αυτόματος εντοπισμός περιοχών κειμένου και γραμμών από το Tool Panel
- 2.4 Αναγνώριση χαρακτήρων με χρήση της βάσης «English Writing M1»
- 2.5 Οπτικός έλεγχος της κατάτμησης και αναγνώρισης

Η πλατφόρμα Transkribus – Άσκηση 3

- 3.1 Εισαγωγή των αρχείων materials/unit7/fek* σε νέο έγγραφο της νέας συλλογής ([ΦΕΚ](#))
- 3.2 Αναγνώριση τυπωμένου κειμένου (Run OCR από το Tool Panel)
- 3.3 Οπτικός έλεγχος της κατάτμησης και αναγνώρισης
- 3.4 Αναζήτηση 3 λέξεων που θα επιλέξετε βλέποντας τις εικόνες
- 3.5 Παροχή πρόσβασης σε άλλους χρήστες στην συλλογή που δημιουργήσατε

Η πλατφόρμα Transkribus – Άσκηση 4

- 4.1 Εισαγωγή των αρχείων materials/unit7/tsirkas* σε νέο έγγραφο της νέας συλλογής ([Επιστολή Τσίρκας - Παπαιωάνου, 17/8/1955](#))
- 4.2 Αυτόματος εντοπισμός περιοχών κειμένου και γραμμών από το Tool Panel
- 4.3 Διόρθωση λαθών στις εντοπισμένες γραμμές κειμένου
- 4.4 Πληκτρολόγηση κειμένου
- 4.5 Εξαγωγή του εγγράφου στον υπολογιστή σας