



ΤΜΗΜΑ ΑΡΧΕΙΟΝΟΜΙΑΣ, ΒΙΒΛΙΟΘΗΚΟΝΟΜΙΑΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ

ΠΛΗΡΟΦΟΡΗΣΗΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΤΙΚΩΝ, ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

Πρόγραμμα Μεταπτυχιακών Σπουδών: «Διαχείριση
Πληροφοριών σε Βιβλιοθήκες, Αρχεία, Μουσεία»

Διπλωματική Εργασία

**«Συγκέντρωση και καταγραφή ελληνικών
γλωσσικών πόρων του ΠαΔΑ»**

Φοιτήτρια
Αγγελική Μπαμνιώτη (ΑΜ: mslam206682018)

Επιβλέπων: Δρ. Σαράντος Καπιδάκης

Αθήνα, Δεκέμβριος 2022



Σκοπός της έρευνας

- Διερεύνηση, καταγραφή και συγκέντρωση των γλωσσικών πόρων που έχουν παραχθεί στα πλαίσια της λειτουργίας και του εκπαιδευτικού και ερευνητικού έργου του Πανεπιστημίου Δυτικής Αττικής
- Διασύνδεση και τεκμηρίωση των γλωσσικών πόρων με τη γλωσσική υποδομή του Clarin:el
- Εμπλουτισμός και οργάνωση του ψηφιακού αποθετηρίου του Πανεπιστημίου Δυτικής Αττικής στην υποδομή του Clarin:el

Ορισμοί – Βιβλιογραφική ανασκόπηση

Clarin
Clarin:el

Παρεμφερείς υποδομές
(*Dariah. Europeana,*
Απολλωνίς κτλ)

Γλωσσικοί πόροι
Γλωσσικές τεχνολογίες
Υπηρεσίες γλωσσικής επεξεργασίας

Ψηφιακά αποθετήρια
Ψηφιακές ανθρωπιστικές επιστήμες
Υπολογιστική γλωσσολογία
Τεχνητή νοημοσύνη

Μεθοδολογία

Διασύνδεση με το Clarin:el

- Χρήση ακαδημαϊκού λογαριασμού
- Εμπλουτισμός του αποθετηρίου του ΠαΔΑ που υπήρχε αλλά ήταν άδειο

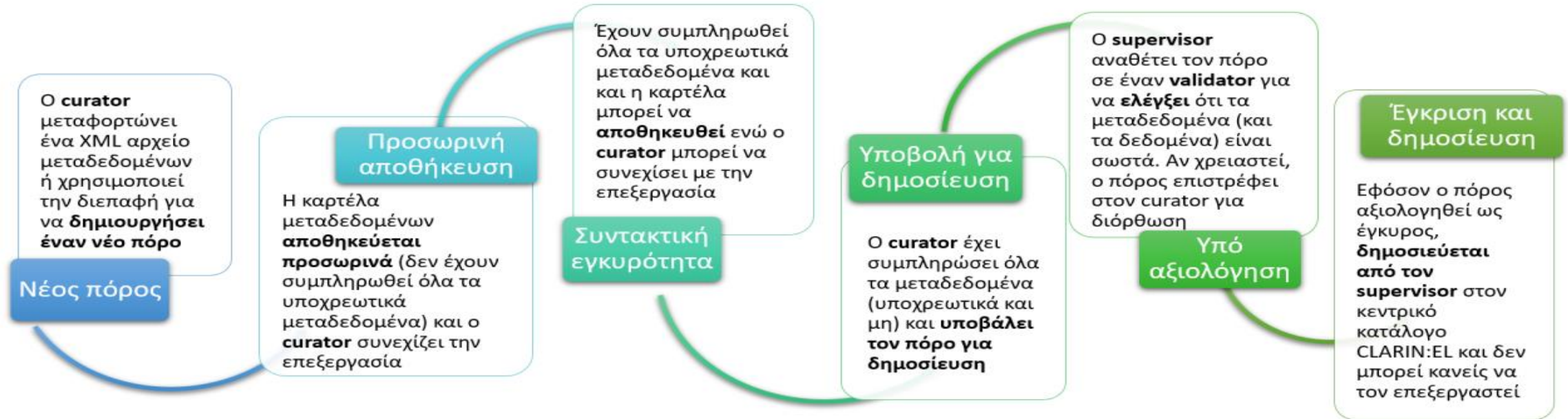
Αναζήτηση γλωσσικών πόρων

- Επικοινωνία με τους πιθανούς δημιουργούς γλωσσικών πόρων (μέσω email, συνάντηση μέσω teams ή δια ζώσης)
- Σύνταξη συχνών ερωταπαντήσεων για καλύτερη πληροφόρηση των παραγωγών
- Γλωσσικό υλικό που διατίθεται με ανοιχτές άδειες χρήσης Creative Commons

Περιγραφή υλοποίησης

- Προετοιμασία του υλικού
- Συλλογή των κατάλληλων δεδομένων
- Κατηγοριοποίηση και οργάνωση με συνάφεια
- Συμπύεση χρησιμοποιώντας τους κατάλληλους μορφότευπους
- Τεκμηρίωση με τα κατάλληλα μεταδεδομένα και περιγραφή των γλωσσικών πόρων

Μεταφόρτωση μεταδεδομένων και τεκμηρίωση



Curator: τεκμηριωτής

Supervisor: υπεύθυνος αποθετηρίου

Validator: ελεγκτής

Αρχές FAIR

Το Clarin:el και το σχήμα μεταδεδομένων που υποστηρίζει ανταποκρίνονται στις αρχές FAIR:

Findability (ευρεσιμότητα)

Accessibility (προσβασιμότητα)

Interoperability (διαλειτουργικότητα)

Reuse (επαναχρησιμοποίηση)

Υποχρεωτικά μεταδεδομένα για την τεκμηρίωση γλωσσικών πόρων σε μορφή σώματος κειμένου

LANGUAGE RESOURCE/ TECHNOLOGY	CORPUS	PART	DISTRIBUTION	DATA
IDENTITY <ul style="list-style-type: none">Resource NameDescriptionVersion	TECHNICAL <ul style="list-style-type: none">Corpus subclassPersonal DataSensitive DataAnonymized (*)	MEDIA PART <ul style="list-style-type: none">Corpus PartLinguality type (text, audio, video, image)Multilinguality type (text, audio, video)Language (text, audio, video, image)Type of content (video, image, textNumerical)	TECHNICAL <ul style="list-style-type: none">Dataset DistributionDataset Distribution FormDistribution Location (*)Download Location (*)Access Location (*)Distribution Medium Features (*)Data FormatSizeLicence Terms	DATA
CATEGORIES <ul style="list-style-type: none">Keyword				
CONTACT <ul style="list-style-type: none">Additional Information				
DOCUMENTATION				
RELATED LRTs				

Υποχρεωτικά μεταδεδομένα για την τεκμηρίωση γλωσσικών πόρων σε μορφή λεξικό- εννοιολογικών πόρων

LANGUAGE RESOURCE/ TECHNOLOGY	CORPUS	PART	DISTRIBUTION	DATA
IDENTITY <ul style="list-style-type: none">Resource NameDescriptionVersion	TECHNICAL <ul style="list-style-type: none">Encoding LevelPersonal DataSensitive DataAnonymized (*)	MEDIA PART <ul style="list-style-type: none">Lexical/Conceptual Resource PartLinguality type (text, audio, video, image)Language (text, audio, video, image)Type of content (video, image)	TECHNICAL <ul style="list-style-type: none">Dataset DistributionDataset Distribution FormDistribution Location (*)Download Location (*)Access Location (*)Distribution Medium Features (*)Data FormatSizeLicence Terms	DATA
CATEGORIES <ul style="list-style-type: none">Keyword				
CONTACT <ul style="list-style-type: none">Additional Information				
DOCUMENTATION				
RELATED LRTs				

Προαιρετικά πεδία που χρησιμοποιήθηκαν στην τεκμηρίωση

- Δημιουργός γλωσσικού πόρου (*LRT creator*)
- Μεταφόρτωση λογότυπου (Upload a logo)
- Ανωνυμοποίηση (Anonymized)
- Τομέας (Domain)
- Τύπος κειμένου (text type)
- Τύπος περιεχομένου (Content type)
- Γεωγραφική κάλυψη (Geographic coverage)
- Χρονική κάλυψη (Time coverage)

Άδειες χρήσης και πνευματικά δικαιώματα

- Αρχική άδεια που είχε ο πόρος όταν πρωτοδημοσιεύτηκε
- Άδειες CC-BY-NC-SA και CC-BY-NC-ND

[\(https://creativecommons.org/\)](https://creativecommons.org/)

Κάποιοι περιορισμοί του Clarin:el

- Δύσχρηστη επικύρωση των γλωσσικών πόρων (πολλά κλικ...)
- Αδυναμία να πραγματοποιήσει κάποιος διορθώσεις μόνος του στα μεταδεδομένα του πόρου του, εφόσον τον έχει μεταφορτώσει ήδη στο Clarin:el
- Δυσκολίες στην εύκολη ανεύρεση των πόρων στην υποδομή
- Παρανοήσεις και λάθη στην περιγραφή των γλωσσικών πόρων λόγω μη λεπτομερούς περιγραφής μεταδεδομένων που μοιάζουν μεταξύ τους
- Την εξαγωγή των μεταδεδομένων όλων των γλωσσικών πόρων συγκεντρωτικά κάποιου αποθετηρίου δεν μπορεί να την πραγματοποιήσει ο χρήστης αλλά μόνο η ομάδα του Clarin

Αποτελέσματα – Ευρήματα (1)

Συνολικά διασυνδέθηκαν 193 γλωσσικοί πόροι με το Clarin:el

η	Τύποςη		Μορφήη	Γλώσσαη			Γλωσσικός·τύποςη	
	Σώμα· κειμένουη	Λεξικό· εννοιολογικός· πόροςη	Κείμενο· (text)η	Ελληνικάη	Αγγλικάη	Μεσσηνιακή· διάλεκτοςη	Μονόγλωσσοη	Δίγλωσσοη
Ανοιχτά· ακαδημαϊκά· μαθήματαη	148η	η	148η	148η	η	η	148η	η
Υλικό·που· στάλθηκε·από· τους·διδάσκοντεςη	19η	5η	24η	22η	6η	1η	20η	4η
Εργασίες· φοιτητώνη	8η	η	8η	7η	1η	η	8η	η
Συγγράμματα·από· τον·Κάλλιπποη	7η	η	7η	7η	η	η	7η	η
Περιγραφές· μαθημάτωνη	6η	η	6η	6η	η	η	6η	η

Πίνακας·16··Συγκεντρωτικός·πίνακας·του·υλικού·σε·αριθμητικές·τιμέςη

Αποτελέσματα – Ευρήματα (2)

Προέλευση πόρου ανά ακαδημαϊκό ίδρυμα	Αριθμός γλωσσικών πόρων
ΤΕΙ Πειραιά	80
ΤΕΙ Αθηνών	68
Πανεπιστήμιο Δυτικής Αττικής	45

Είδος γλωσσικού πόρου	Αριθμός γλωσσικών πόρων
Ανοιχτά Ακαδημαϊκά Μαθήματα	148
Υλικό που στάλθηκε απευθείας από τους διδάσκοντες του ΠαΔΑ	24
Εργασίες φοιτητών του ΠαΔΑ	8
Κάλλιπος	7
Περιγραφές μαθημάτων	6

Αποτελέσματα – Ευρήματα (3)

Υλικό ανά σχολή του ΤΕΙ Αθηνών	Αριθμός γλωσσικών πόρων
Σχολή Τεχνολογικών Εφαρμογών	36
Σχολή Καλλιτεχνικών Σπουδών	11
Σχολή Επαγγελματιών Υγείας και Πρόνοιας	9
Σχολή Τεχνολογίας Τροφίμων και Διατροφή	7
Σχολή Διοίκησης και Οικονομίας	5

Υλικό αριθμητικά ανά τμήμα του ΤΕΙ Αθηνών

- Ναυπηγών Μηχανικών 19
- Οινολογίας και Τεχνολογίας Ποτών 6
- Μηχανικών Πληροφορικής 6
- Φωτογραφίας και Οπτικοακουστικών Τεχνών 6
- Πολιτικών Μηχανικών και Μηχανικών Τοπογραφίας και Γεωπληροφορικής 5
- Βιβλιοθηκονομίας και Συστημάτων Πληροφόρησης 4
- Μηχανικών Ενεργειακής Τεχνολογίας 4
- Γραφιστικής 3
- Αισθητικής και Κοσμετολογίας 2
- Οπτικής και Οπτομετρίας 2
- Ραδιολογίας και Ακτινολογίας 2
- Μηχανικών Βιοιατρικής Τεχνολογίας 2
- Συντήρησης Αρχαιοτήτων και Έργων Τέχνης 1
- Δημόσιας Υγείας και Κοινωνικής Υγείας 1
- Οδοντικής Τεχνολογίας 1
- Ιατρικών Εργαστηρίων 1
- Εσωτερικής Αρχιτεκτονικής, Διακόσμησης και Σχεδιασμού Αντικειμένων 1
- Τεχνολογίας Τροφίμων 1
- Διοίκησης Επιχειρήσεων - Τουριστικών Επιχειρήσεων και Επιχειρήσεων Φιλοξενίας 1

Αποτελέσματα – Ευρήματα (4)

Υλικό ανά σχολή του ΤΕΙ Πειραιά	Αριθμός γλωσσικών πόρων
Τεχνολογικών Εφαρμογών	57
Διοίκησης και Οικονομίας	23

Υλικό ανά τμήμα του ΤΕΙ Πειραιά	Αριθμός γλωσσικών πόρων
Διοίκηση Επιχειρήσεων	16
Μηχανολόγων Μηχανικών	14
Μηχανικών Αυτοματισμού	13
Ηλεκτρονικών Μηχανικών	12
Μηχανικών Ηλεκτρονικών Υπολογιστών Συστημάτων	8
Λογιστικής και χρηματοοικονομικής	7
Ηλεκτρολόγων Μηχανικών	6
Πολιτικών Μηχανικών	4

Αποτελέσματα – Ευρήματα (5)

Είδος υλικού από το ΠαΔΑ	Αριθμός γλωσσικών πόρων
Υλικό που δόθηκε απευθείας από τους διδάσκοντες του ΠαΔΑ	24
Συγγράμματα από τον Κάλλιππο	7
Διπλωματικές εργασίες	7
Περιγραφές μαθημάτων	6
Πτυχιακή εργασία	1



Αποτελέσματα – Ευρήματα (6)

Υλικό ανά σχολή του ΠαΔΑ	Αριθμός γλωσσικών πόρων
Διοικητικών Οικονομικών και Κοινωνικών Επιστημών	35
Μηχανικών	8
Εφαρμοσμένων Τεχνών και Πολιτισμού	1
Επιστημών Υγείας και Πρόνοιας	1

Υλικό ανά τμήμα του ΠαΔΑ	Αριθμός γλωσσικών πόρων
Αρχειονομίας, Βιβλιοθηκονομίας και Συστημάτων Πληροφόρησης	29
Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής	8
Αγωγής και Φροντίδας στην Πρώιμη Παιδική Ηλικία	6
Γραφιστικής και Οπτικής Τεχνολογίας	1
Βιοιατρικών Επιστημών	1

Περιορισμοί της έρευνας

Τα email απαντήθηκαν από περιορισμένο αριθμό πιθανών παραγωγών γλωσσικών πόρων, λόγω έλλειψης χρόνου και πολυάσχολου ακαδημαϊκού προγράμματος

Άγνοια σχετικά με την έρευνα και τους σκοπούς της και κάποια «καχυποψία» ως προς το «που θα χρησιμοποιηθεί το υλικό», «τι είναι η συγκεκριμένη εφαρμογή» κτλ

Η έρευνα εστίασε περισσότερο στους διδασκόντων στο τμήμα *Αρχειονομίας, Βιβλιοθηκονομίας και Συστημάτων Πληροφόρησης* και σε μικρότερο ποσοστό σε διδάσκοντες άλλων σχολών και τμημάτων

Χρονικός περιορισμός ενώ η αναζήτηση και τεκμηρίωση του υλικού είναι εξαιρετικά χρονοβόρα διαδικασία

Υλικό χρήσιμο αλλά σε μορφότυπους μη υποστηριζόμενους από την υποδομή

Υλικό σε μορφή που δεν υποστηρίζεται πλέον (πχ το adobe flash player)

Πνευματικά δικαιώματα ή δικαιώματα διάθεσης του υλικού

Ήδη διασυνδεδεμένο υλικό από διδάσκοντες του ΠαΔΑ με άλλο ιδρυματικό αποθετήριο

Περιορισμοί του Clarin:el

Συμπέρασμα

- Μέσω του Clarin:el, γίνεται προσπάθεια συγκέντρωσης του γλωσσικού υλικού του ΠαΔΑ σε ένα αποθετήριο

Αξιοποίηση

- Το ΠαΔΑ γίνεται μέλος μιας διεθνούς κοινότητας διαμοιρασμού πόρων και ανταλλαγής γνώσης και τεχνογνωσίας
- Οι δημιουργοί των πόρων κερδίζουν προβολή και πιθανόν περισσότερη αναγνώριση του έργου τους

Μελλοντικές προεκτάσεις

- Επιπλέον γλωσσικό υλικό θα μπορούσε να συλλεχθεί και διασυνδεθεί με την υποδομή
- Αξιοποίηση υλικού σε μορφότυπους που δεν υποστηρίζονται στην παρούσα φάση από το Clarin:el αλλά θα μπορούσαν να υποστηριχθούν στο μέλλον
- Μελλοντική παραγωγή και διασύνδεση γλωσσικών πόρων του ΠαΔΑ, στα πλαίσια της ερευνητικής του λειτουργίας

Ευχαριστώ για την προσοχή σας!

