



**Τμήμα Αρχειονομίας, Βιβλιοθηκονομίας και Συστημάτων
Πληροφόρησης. Σχολή Διοικητικών, Οικονομικών και
Κοινωνικών Επιστημών**

**Διπλωματική Εργασία
Αξιολόγηση προτύπων μεταδεδομένων ψηφιακών
αποθετηρίων**

Συγγραφέας: Μαρίλια Αποστολάκη

Επιβλέπων: Αλέξανδρος Κουλούρης

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

Λόγοι επιλογής του θέματος

- Η προσέγγιση του θέματος παρουσιάζει μεγάλο ενδιαφέρον
- Η σημαντικότητα της έρευνας για την επιστημονική κοινότητα
- Προσωπικό ενδιαφέρον για τα πρότυπα μεταδεδομένων και την εφαρμογή τους

Δομή της εργασίας

- Εισαγωγικό πλαίσιο
- Προβληματική
- Ερευνητικές υποθέσεις
- Βιβλιογραφική επισκόπηση
- Μεθοδολογία έρευνας
- Παρουσίαση αποτελεσμάτων
- Συμπεράσματα

Εισαγωγή (1/2)

Διαδίκτυο

- Εξέλιξη της Τεχνολογίας των Επικοινωνιών και Παγκόσμιου Ιστού
- Αλλαγή στον τρόπο πραγματοποίησης & διάθεσης της επιστημονικής έρευνας

Μηχανές αναζήτησης

- Σύγχρονες δυναμικές μηχανές επιστημονικής αναζήτησης
- Συσσωρευτές επιστημονικής γνώσης

Αύξηση χρήσης

- Συμπεριφορά των χρηστών
- Πρώτη σε χρήση το Google Scholar (GS) (Orduña-Malea & Delgado López-Cózar, 2015; Rieger, 2009)

Εισαγωγή (2/2)

**Βιβλιοθήκες /
Οργανισμοί
πληροφόρησης**

- Ανάπτυξη υπηρεσιών όπως ΨΑ
- Χρήση DC ως το πιο διαδεδομένο πρότυπο μεταδεδομένων
- Ανάγκη παροχής πρόσβασης & προβολής του έργου των δημιουργών και των φορέων που αυτοί αντιπροσωπεύουν

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

Οφέλη ευρετηρίασης του περιεχομένου των ΨΑ από τις μηχανές αναζήτησης όπως το GS

- Μεγαλύτερη προβολή, ανακάλυψη και ανάκτηση ψηφιακού περιεχομένου – Προώθηση της ΑΠ
- Λήψη περισσότερων ετεροαναφορών των συγγραφέων
- Αύξηση αναγνώρισης και κύρους των συγγραφέων (απήχηση)
- Αύξηση φήμης/κύρους των φορέων/ακαδημαϊκών ιδρυμάτων - Άνοδος στη διεθνή κατάταξη (ranking)
- Πιθανότητες ανάπτυξης συνεργασιών
- Προσέλκυση και εξασφάλιση χρηματοδότησης της ερευνητικής εργασίας

Προβληματική

- Η ευρετηρίαση του περιεχομένου των ΨΑ δεν πραγματοποιείται σε ικανοποιητικό βαθμό από τις μηχανές (ερευνητικής) αναζήτησης όπως το GS
- Εκτιμάται ότι το πρόβλημα αυτό οφείλεται μεταξύ άλλων στην ποιότητα των μεταδεδομένων, όπως και στο πρότυπο μεταδεδομένων που έχει υιοθετηθεί από το εκάστοτε ψηφιακό αποθετήριο (Arlitsch & O'Brien, 2012)

Ερευνητικές Υποθέσεις (1/3)

- Μπορεί το πρότυπο μεταδεδομένων που χρησιμοποιεί ένα ψηφιακό αποθετήριο να επηρεάζει την ανακάλυψη των ψηφιακών πόρων που περιέχονται σε αυτό, από τις σύγχρονες και δυναμικές μηχανές ερευνητικής αναζήτησης όπως το GS;

Ερευνητικές Υποθέσεις (2/3)

- Το Dublin Core (DC) είναι το πρότυπο και σχήμα συνάμα μεταδεδομένων που χρησιμοποιείται από την πλειοψηφία των ψηφιακών αποθετηρίων παγκοσμίως. Μπορούμε όμως να πετύχουμε μεγαλύτερη ευρετηρίαση και ορατότητα του ψηφιακού περιεχομένου που περιγράφεται με το συγκεκριμένο πρότυπο από τις μηχανές αναζήτησης όπως το GS;

Ερευνητικές Υποθέσεις (3/3)

- Μήπως η διάθεση/παροχή με έγκυρο τρόπο συγκεκριμένων στοιχείων μεταδεδομένων επηρεάζει την ανακάλυψη και ευρετηρίαση των ψηφιακών πόρων των αποθετηρίων από τις δυναμικές μηχανές αναζήτησης όπως το GS;

DC – Πλεονεκτήματα (1/2)

- Ευρέως χρησιμοποιούμενο διεθνές πρότυπο
- Πολύ απλό στο σύνολο των στοιχείων του
- Επεκτάσιμο
- Συμβατό με τις απαιτήσεις του πρωτοκόλλου συγκομιδής OAI-PMH για τη συλλογή μεταδεδομένων
- Με διαθέσιμη την κωδικοποίησή του για τον Ιστό, σε γλώσσες XML (eXtensible Markup Language) και RDF (Resource Description Framework)
- Υποστηρίζει ανοικτά πρότυπα και δημιουργεί δυναμική για αυτά

DC – Πλεονεκτήματα (2/2)

- Υποστηρίζεται και ενημερώνεται από το φορέα Dublin Core Metadata Initiative, ο οποίος είναι υπεύθυνος για την ανάπτυξη και τη βελτίωσή του
- Παρέχει αντιστοιχίσεις (mapping) από και προς άλλα σχήματα μεταδεδομένων
- Προωθεί την ανάκτηση της πληροφορίας από τις μηχανές αναζήτησης
- Παρέχεται σε αρκετές γλώσσες (διεθνοποίηση)

Google Scholar - Πλεονεκτήματα

- Παρέχει δωρεάν τον πιο ολοκληρωμένο κατάλογο ερευνητικών εργασιών που είναι διαθέσιμες στον Ιστό
- Εξάγει μεταδεδομένα από τις πιο σχετικές και αναγνωρισμένες επιστημονικές πηγές στο διαδίκτυο μέσω μαζικής χρήσης ανιχνευτών και εργαλείων συλλογής
- Διαθέτει απλή και εύχρηστη διεπαφή χρηστών
- Παρέχει αποτελέσματα σε σύντομο χρόνο και με απευθείας ανάκτηση του πλήρους κειμένου των ψηφιακών πόρων πολλές φορές
- Αναπτύχθηκε υπό την Google και χρησιμοποίησε μέρος της τεχνολογίας της

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

Google Scholar - Μειονεκτήματα

- Ισχυρή σύσταση για χρήση μετά-ετικετών DC ως έσχατη επιλογή
- Παρατηρούνται πολλά προβλήματα που σχετίζονται κυρίως με τα μεταδεδομένα του GS. Π.χ. δημιουργία μη υπαρκτών συγγραφέων, χρονολογιών έκδοσης κ.ά., προκαλώντας ασυνέπειες στα ονόματα των συγγραφέων και εσφαλμένη απόδοση παραπομπών (Citations)
- Έχει πάψει να υποστηρίζει το πρωτόκολλο OAI-PMH
- Η συμπερίληψη των αυτοαναφορών προσθέτει θόρυβο στα βιβλιομετρικά αποτελέσματα συγγραφέων και τίτλων
- Η μυστικότητα που διατηρείται γύρω από το επιστημονικό πεδίο που καλύπτεται και του αριθμού των πληροφοριακών πόρων που ευρετηριάζονται από αυτό

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

Μεθοδολογία έρευνας

- Ποσοτική έρευνα με αποστολή ερωτηματολογίων
- Δημιουργία δύο ερωτηματολογίων (σε ελληνική και αγγλική έκδοση)
- Επιλογή 407 ΨΑ με κριτήρια – 57 ελληνικών και κυπριακών & 350 ΨΑ από το εξωτερικό (επιλέγοντας από την ιστοσελίδα Open DOAR-Greece (https://v2.sherpa.ac.uk/view/repository_by_country/Greece.html) και Open DOAR - Cyprus (https://v2.sherpa.ac.uk/view/repository_by_country/Cyprus.html) και από την ιστοσελίδα Ranking Web of Repositories (<https://repositories.webometrics.info/en/node/30>) 9th ed. April 2020)
- Απαντήθηκαν 98 ερωτηματολόγια (29 από ελληνικά και κυπριακά ΨΑ και 69 από ΨΑ του εξωτερικού)
- Ποσοστό ανταπόκρισης 25%

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

Είδη ερωτήσεων που χρησιμοποιήθηκαν

- Ερωτήσεις πολλαπλής επιλογής (multiple choice) με αποδεκτή μία μοναδική απάντηση
- Ερωτήσεις πολλαπλής επιλογής (multiple choice) με αποδεκτές μίας ή περισσότερων από τις προτεινόμενες απαντήσεις
- Ερωτήσεις που αντιστοιχούν σε τετράβαθμη κλίμακα Likert

Στατιστική ανάλυση

- Η επεξεργασία των δεδομένων πραγματοποιήθηκε κάνοντας χρήση του λογισμικού στατιστικής ανάλυσης “IBM SPSS STATISTICS” version 20
- Προτιμήθηκε η δημιουργία ενός ενιαίου φύλλου εργασίας στο εν λόγω λογισμικό & καταχωρήθηκαν όλα τα δεδομένα από τα ερωτηματολόγια
- Χρησιμοποιήθηκε ο συντελεστής ανεξαρτησίας Pearson Chi-square (χ^2) για να ελεγχθεί η ύπαρξη πιθανής συσχέτισης μεταξύ δύο ή περισσότερων κατηγορικών μεταβλητών (μη διατάξιμων (nominal))

Παρουσίαση αποτελεσμάτων (1/3)

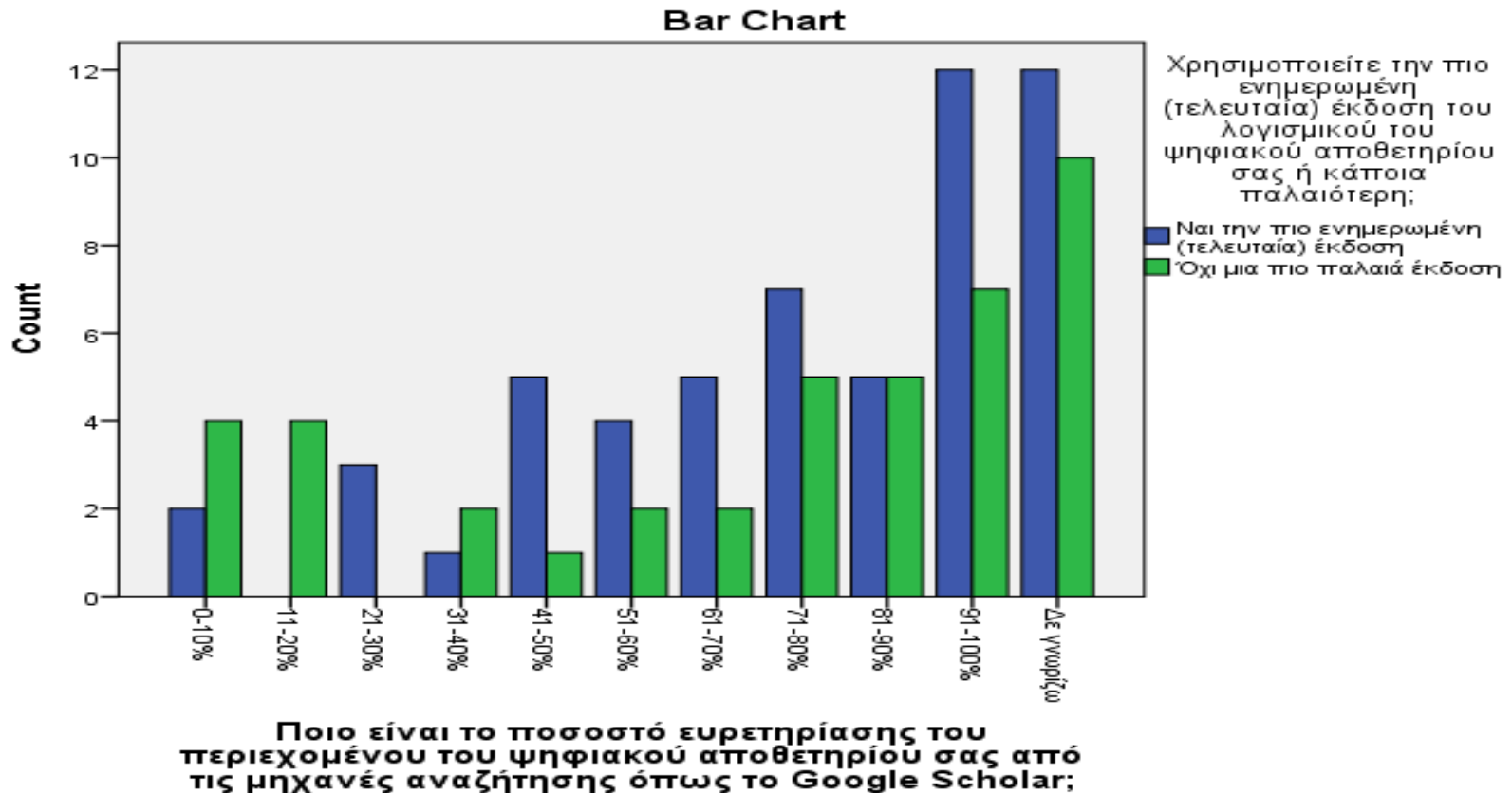
- Η πλειοψηφία των ερωτώμενων είναι γυναίκες (46,94%)
- Η πλειοψηφία ανήκει στην ηλικιακή ομάδα μεταξύ των 36-45 ετών (30,2%)
- Η ιδιότητα των περισσότερων είναι διαχειριστές ΨΑ σε ποσοστό 40,23%
- Η πλειοψηφία των ερωτώμενων εργάζεται σε Ιδρυματικά ΨΑ με ποσοστό 80,6%

Παρουσίαση αποτελεσμάτων (2/3)

- Το περιεχόμενο των ΨΑ είναι ιδιαίτερο (μη εμπορικό), συνήθως δυσεύρετο (γκρίζα βιβλιογραφία, τεχνικές εκθέσεις, εργασίες συνεδρίων, κ.ά.)
- Η απόθεση γίνεται κατά δήλωση της πλειοψηφίας με διάφορους συνδυασμούς πέραν των προτεινόμενων
- Η διαχείριση ωστόσο της αρχειοθέτησης τουλάχιστον με συμβολή από εξειδικευμένο επαγγελματία (βιβλιοθηκονόμο) δείχνει την τάση των ΨΑ για έλεγχο των μεταδεδομένων πριν την οριστική υποβολή των πόρων στο ΨΑ
- Πρώτο σε επιλογή και με διαφορά έρχεται το ανοικτό λογισμικό Dspace με ποσοστό 50%, με το λογισμικό Eprints να ακολουθεί με ποσοστό 18,4%
- Οι περισσότεροι από τους συμμετέχοντες δηλώνουν τη χρήση της πιο ενημερωμένης έκδοσης λογισμικού του ΨΑ τους

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

Χρήση της τελευταίας έκδοσης λογισμικού ΨΑ



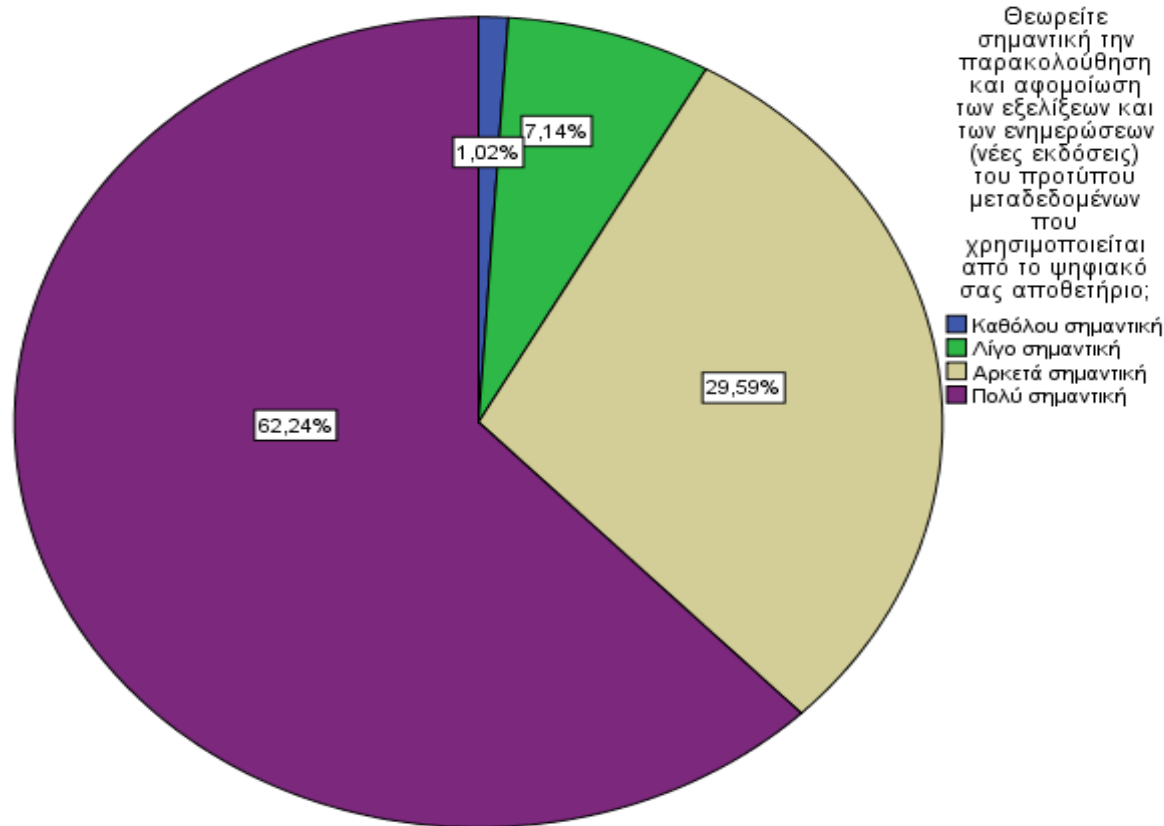
Πρόγραμμα Μεταπτυχιακών Σπουδών: «Διαχείριση Πληροφοριών σε Βιβλιοθήκες, Αρχεία, Μουσεία»

Παρουσίαση αποτελεσμάτων (3/3)

- Η περιγραφή των πόρων στην πλειοψηφία των ΨΑ γίνεται σύμφωνα με το πρότυπο μεταδεδομένων εξειδικευμένο (Qualified) Dublin Core σε ποσοστό 54,1% και αμέσως μετά ακολουθεί η επιλογή «Άλλο» σε ποσοστό 17,3% ενώ το μη εξειδικευμένο Dublin Core (Unqualified) έρχεται στην τρίτη θέση με ποσοστό 10,2%
- Τα στοιχεία μεταδεδομένων που καταχωρούνται με σειρά προτεραιότητας είναι τα εξής: τίτλος, συγγραφέας , χρονολογία έκδοσης/δημοσίευσης, εκδότης, γλώσσα, θέμα, λέξεις-κλειδιά, αριθμός τόμου, παραπομπή, αριθμός τεύχους, αριθμός σελίδων, αναγνωριστικό και «άλλο»

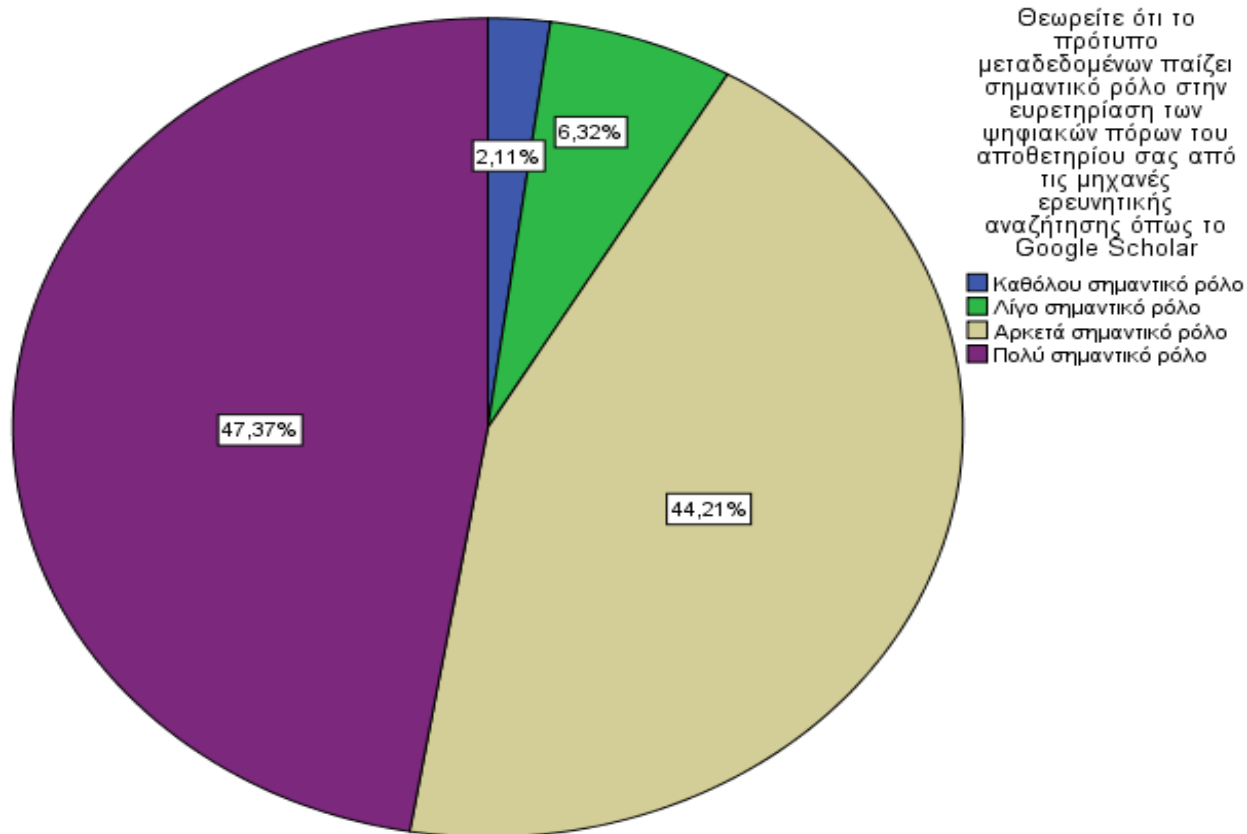
Πρόγραμμα Μεταπτυχιακών Σπουδών: «Διαχείριση Πληροφοριών σε Βιβλιοθήκες, Αρχεία, Μουσεία»

Θεωρείτε σημαντική την παρακολούθηση και αφομοίωση των εξελίξεων και ενημερώσεων του προτύπου μεταδεδομένων που χρησιμοποιείται από το ΨΑ σας;



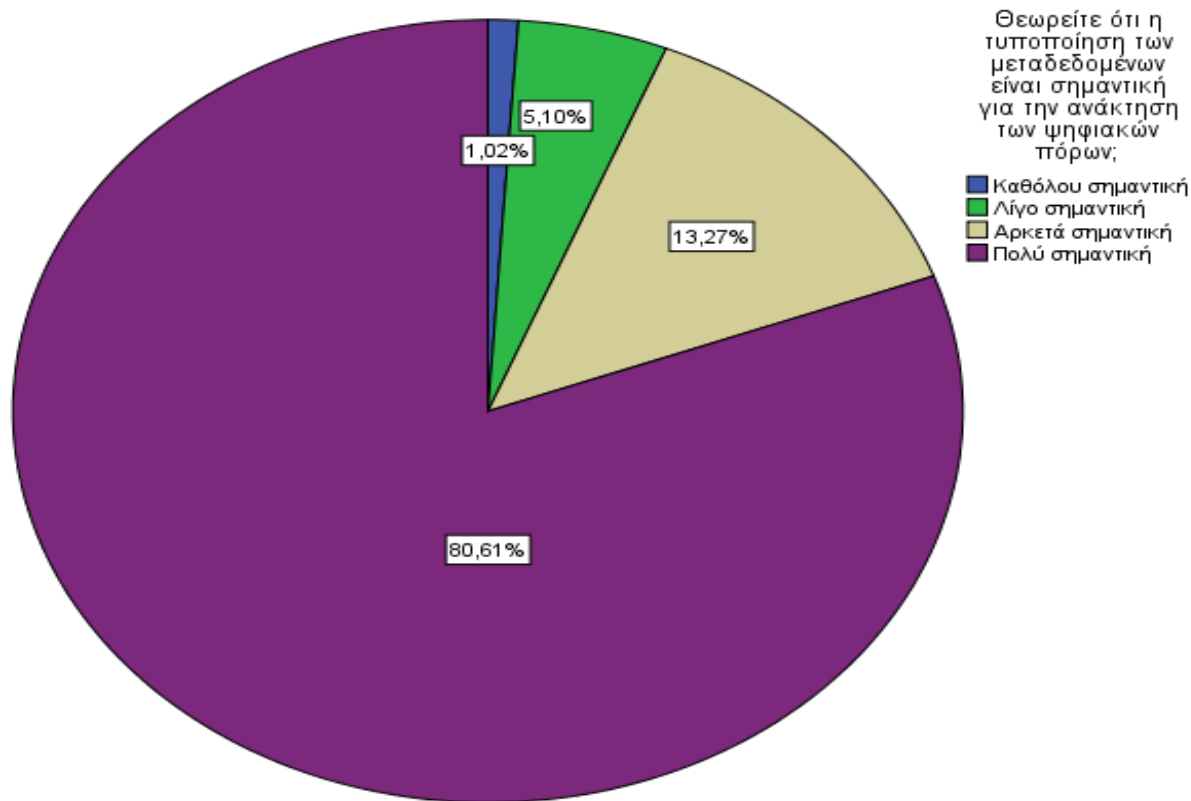
Πρόγραμμα Μεταπτυχιακών Σπουδών: «Διαχείριση Πληροφοριών σε Βιβλιοθήκες, Αρχεία, Μουσεία»

Θεωρείται ότι το πρότυπο μεταδεδομένων παίζει σημαντικό ρόλο στην ευρετηρίαση των ψηφιακών πόρων του ΨΑ σας από τις μηχανές αναζήτησης όπως το GS;



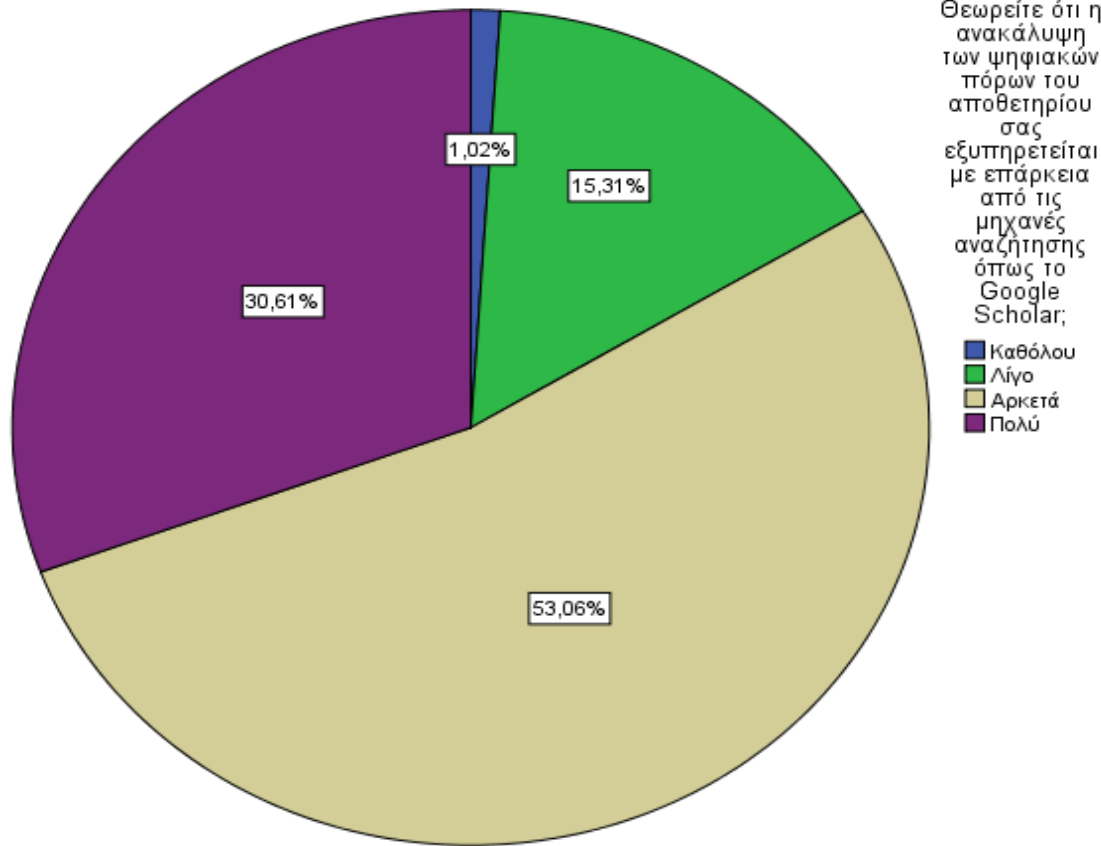
Πρόγραμμα Μεταπτυχιακών Σπουδών: «Διαχείριση Πληροφοριών σε Βιβλιοθήκες, Αρχεία, Μουσεία»

Θεωρείτε ότι η τυποποίηση των μεταδεδομένων είναι σημαντική για την ανάκτηση των πόρων του ΨΑ σας;



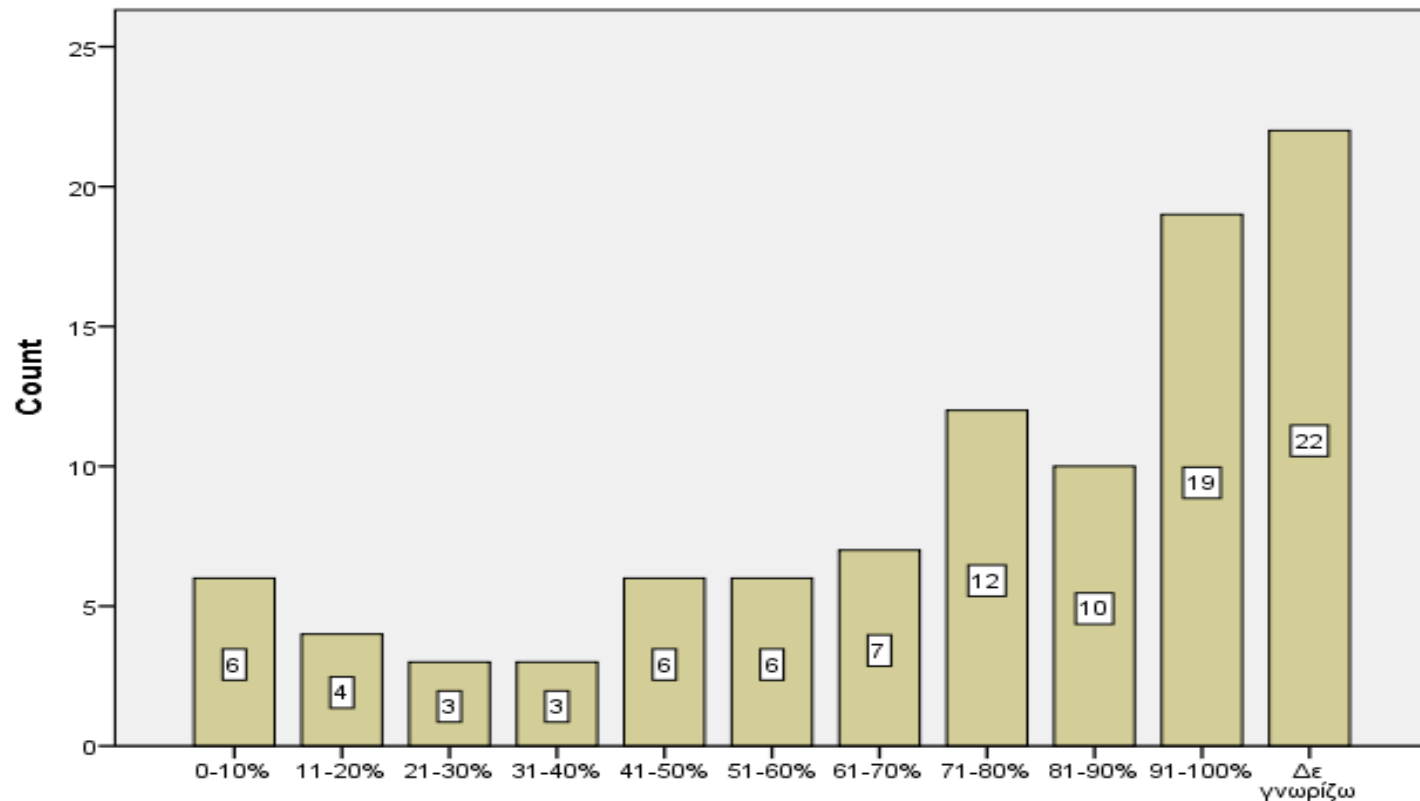
Πρόγραμμα Μεταπτυχιακών Σπουδών: «Διαχείριση Πληροφοριών σε Βιβλιοθήκες, Αρχεία, Μουσεία»

Θεωρείτε ότι η ανακάλυψη των ψηφιακών πόρων του ΨΑ σας εξυπηρετείται με επάρκεια από τις μηχανές αναζήτησης όπως το GS;



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

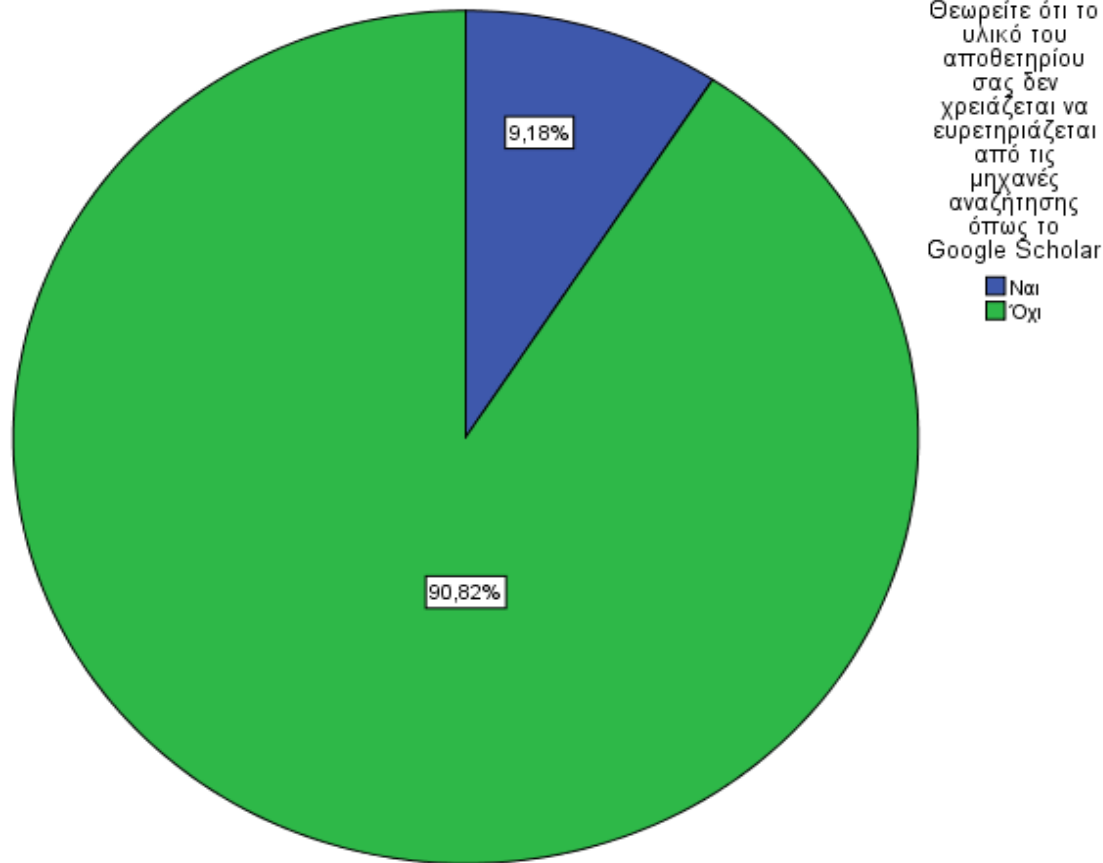
Ποιο είναι το ποσοστό ευρετηρίασης του περιεχομένου του ΨΑ σας από τις μηχανές αναζήτησης όπως το GS;



Ποιο είναι το ποσοστό ευρετηρίασης του περιεχομένου του ψηφιακού αποθετηρίου σας από τις μηχανές αναζήτησης όπως το Google Scholar;

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

Θεωρείτε ότι το υλικό του αποθετηρίου σας δεν
χρειάζεται να ευρετηριάζεται από τις μηχανές
αναζήτησης όπως το GS;



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

Ενέργειες στις οποίες έχει προχωρήσει κάθε φορέας για αύξηση της ευρετηρίασης του ΨΑ από τις μηχανές αναζήτησης όπως το GS

- Μετατροπή του προτύπου μεταδεδομένων σε ένα περισσότερο συμβατό πρότυπο – 15,2%
- Ενσωμάτωση ενός μικρού block κώδικα του προτύπου μεταδεδομένων στο HTML κάθε ψηφιακού πόρου σε μορφή μεταετικετών (meta tags) – 13,5%
- Συμμόρφωση με τα Inclusion Guidelines for Webmasters – 20,2%
- Παρακολούθηση των ενημερώσεων του προτύπου μεταδεδομένων του ΨΑ τους και αφομοίωση των προτεινόμενων αλλαγών – 10,7%
- Εγκατάσταση της τελευταίας, πιο ενημερωμένης έκδοσης του λογισμικού – 16,3%
- Ένα μικρό ποσοστό 5,6% έχει προχωρήσει σε κάποια άλλη ενέργεια
- Το 18,5% των συμμετεχόντων δεν έχει προχωρήσει σε καμία ενέργεια

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

Προβλήματα που πιστεύεται ότι επηρεάζουν την ευρετηρίαση του περιεχομένου των ΨΑ από τις μηχανές αναζήτησης όπως το Google Scholar

- Συμβατότητας του προτύπου μεταδεδομένων του ΨΑ με τις απαιτήσεις των μηχανών αναζήτησης – 12,8%
- Δυσκολία ενσωμάτωσης στο HTML των ιστοσελίδων των ΨΑ των μεταδεδομένων των ψηφιακών πόρων σε μορφή μετα-ετικετών (meta-tags) – 8,3%
- Δυσκολία στην αντιστοίχιση του προτύπου μεταδεδομένων του ΨΑ σε κάποιο άλλο πρότυπο περισσότερο συμβατό – 11,3%
- Μη ύπαρξης σαφών οδηγιών καταλογογράφησης του υλικού του ΨΑ – 6,0%
- Έλλειψης εξειδικευμένου προσωπικού – 16,5%
- «Άλλο πρόβλημα» - 15,0%
- Δεν αντιμετωπίζει κανένα απολύτως πρόβλημα – 30,1%

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

Έλεγχοι ανεξαρτησίας μεταβλητών 1/2

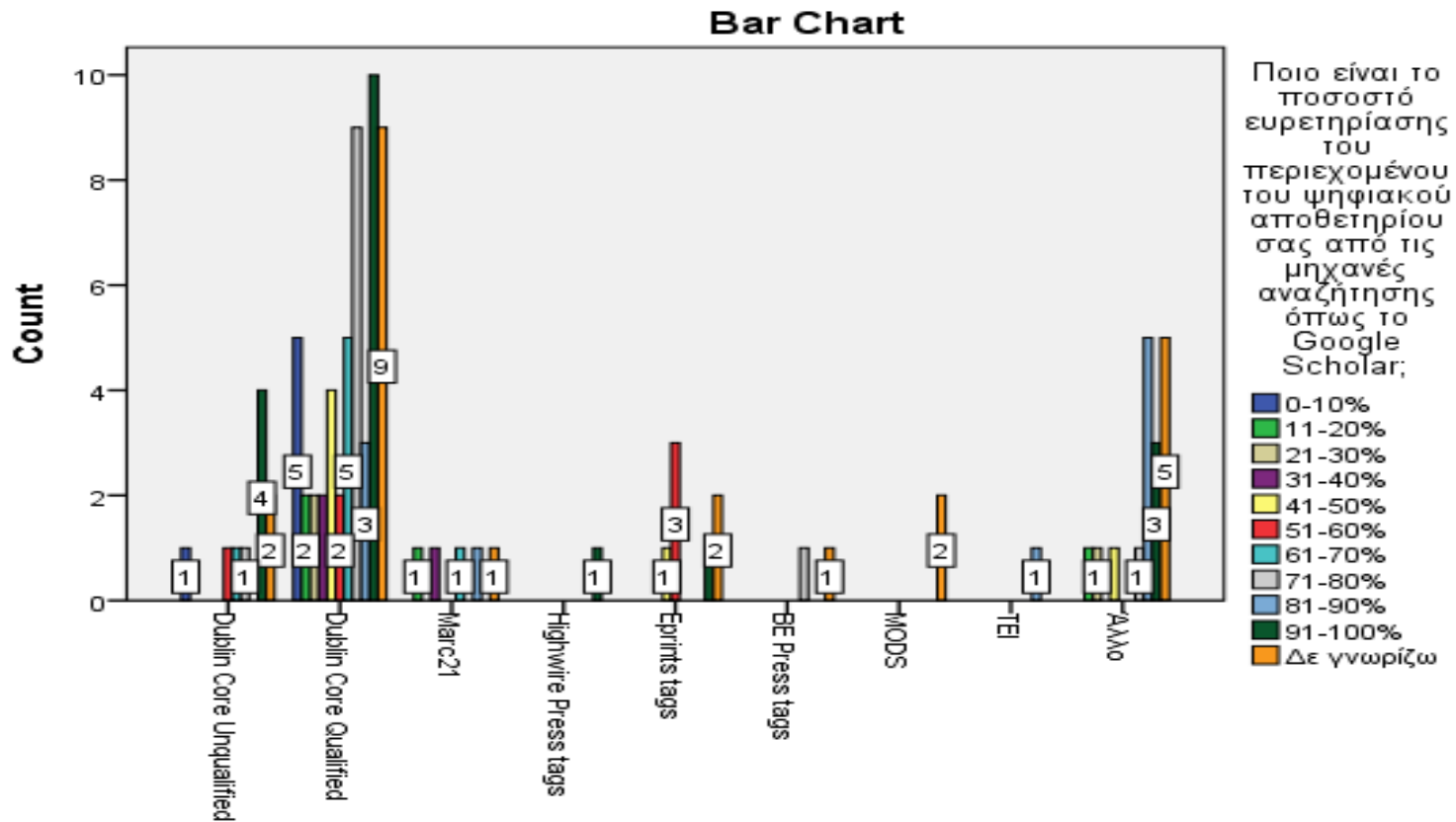
1. Έλεγχος συσχέτισης μεταξύ του προτύπου μεταδεδομένων σε σχέση με το ποσοστό ευρετηρίασης του περιεχομένου του ΨΑ κάθε φορέα από τις μηχανές αναζήτησης όπως το GS

• Από τον έλεγχο προκύπτει: $p\text{-value (sig)} > 5\%$ (Sig.=0.548>0,05). Άρα ισχύει η H_0

2. Έλεγχος συσχέτισης μεταξύ των ενεργειών στις οποίες έχει προχωρήσει το ΨΑ σε σχέση με το ποσοστό ευρετηρίασης του περιεχομένου του ΨΑ κάθε φορέα από τις μηχανές αναζήτησης όπως το GS

• Από τον έλεγχο προκύπτει $p\text{-value (sig)} > 5\%$ (Sig.=0.212>0,05). Επομένως, ισχύει η H_0

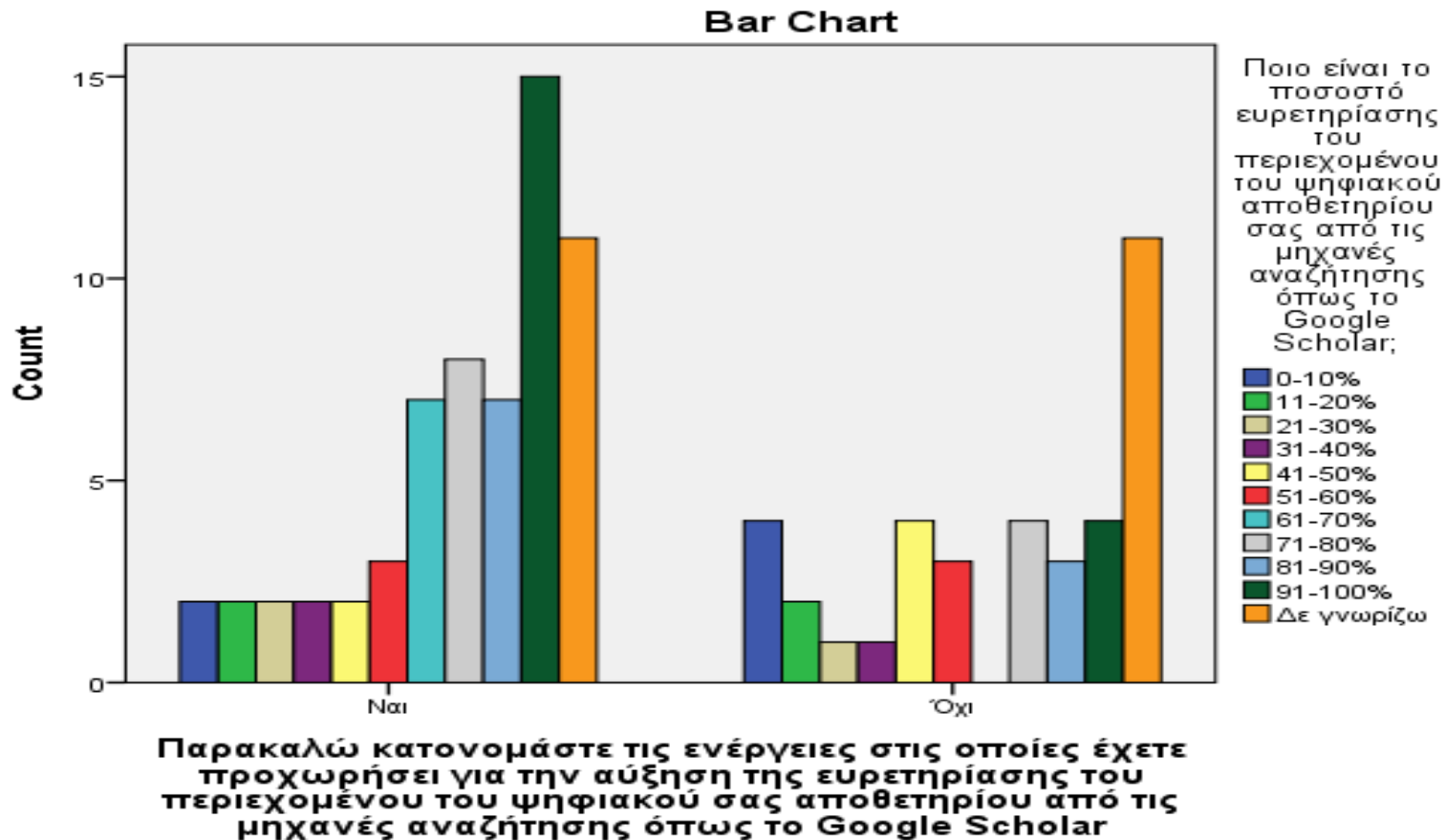
Κυριότερα ευρήματα – Συσχετίσεις (1/4)



Σύμφωνα με ποιο πρότυπο μεταδεδομένων περιγράφονται οι ψηφιακοί πόροι στο αποθετήρι...

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

Κυριότερα ευρήματα – Συσχετίσεις (2/4)



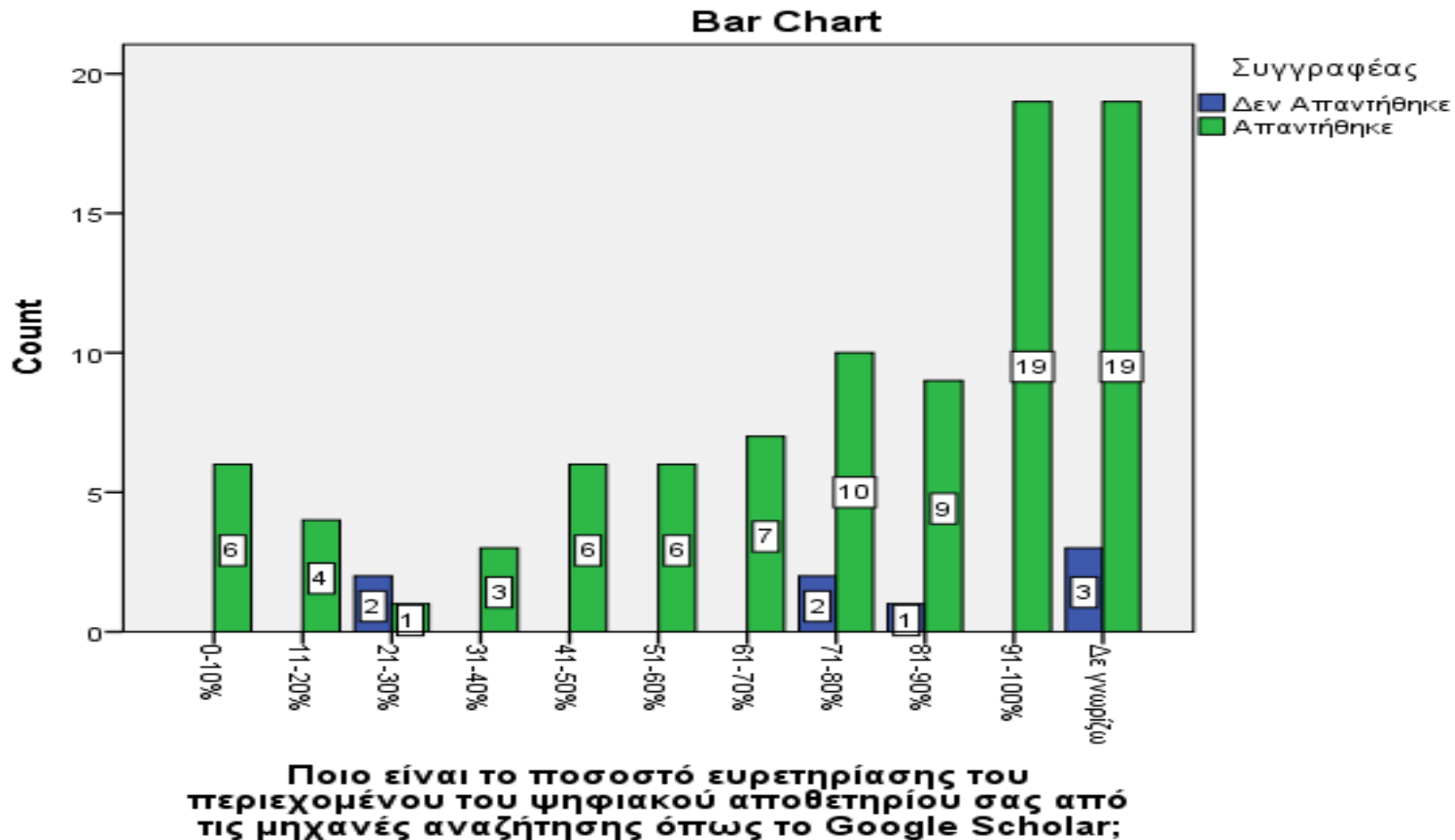
Πρόγραμμα Μεταπτυχιακών Σπουδών: «Διαχείριση Πληροφοριών σε Βιβλιοθήκες, Αρχεία, Μουσεία»

Έλεγχοι ανεξαρτησίας μεταβλητών 2/2

3. Έλεγχος συσχέτισης μεταξύ των στοιχείων/πεδίων μεταδεδομένων σε σχέση με το ποσοστό ευρετηρίασης του περιεχομένου των ΨΑ από τις μηχανές αναζήτησης όπως το GS

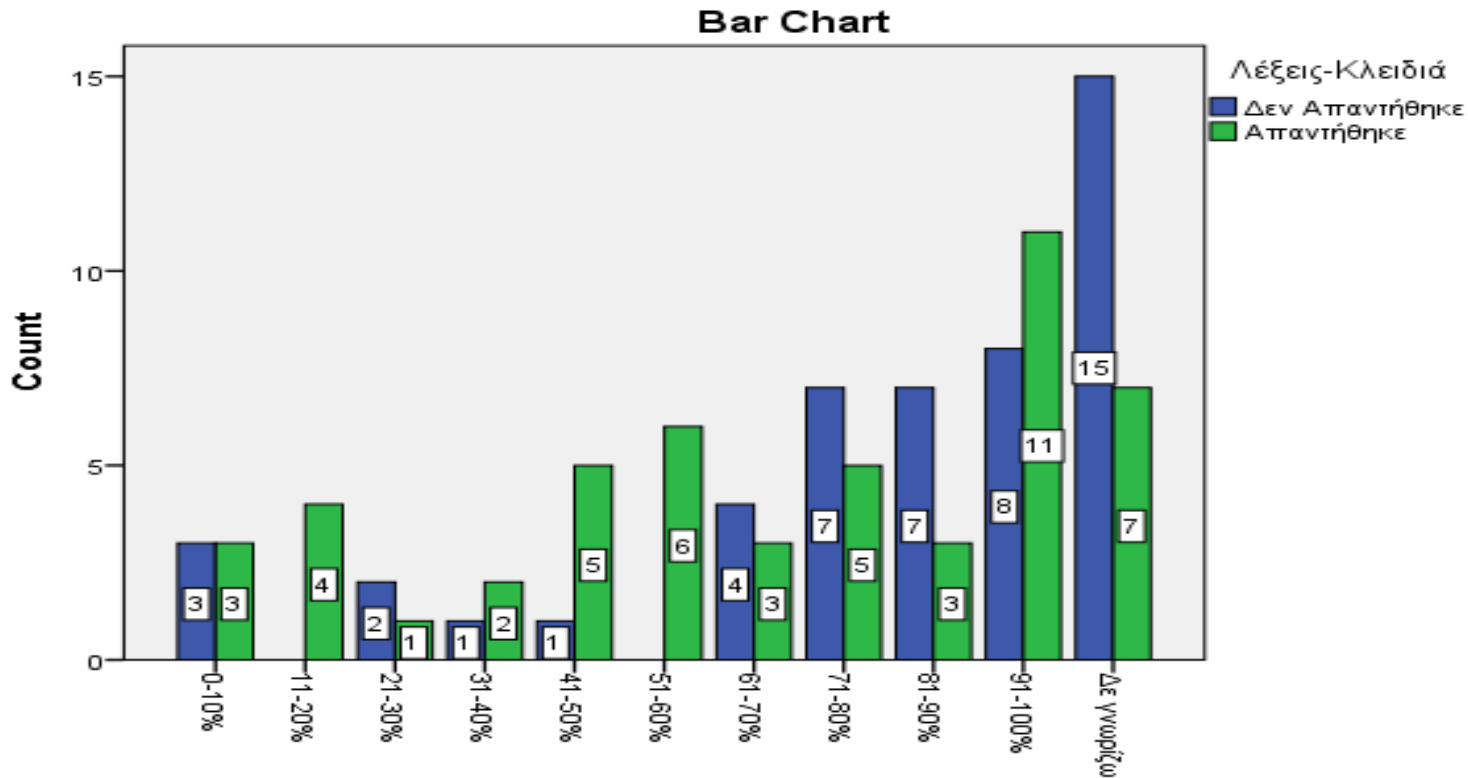
- Από τον έλεγχο προκύπτει p-value (sig) μικρότερο από 5% για την περίπτωση του Συγγραφέα (Sig.=0,026<0,05)
- Αντίστοιχα στην περίπτωση της Λέξης-κλειδιού p-value (sig) μικρότερο από 5%(Sig.=0,043<0,05). Επομένως, ισχύει η H1, ότι δηλαδή οι δυο μεταβλητές που εξετάζουμε είναι εξαρτημένες

Κυριότερα ευρήματα – Συσχετίσεις (3/4)



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

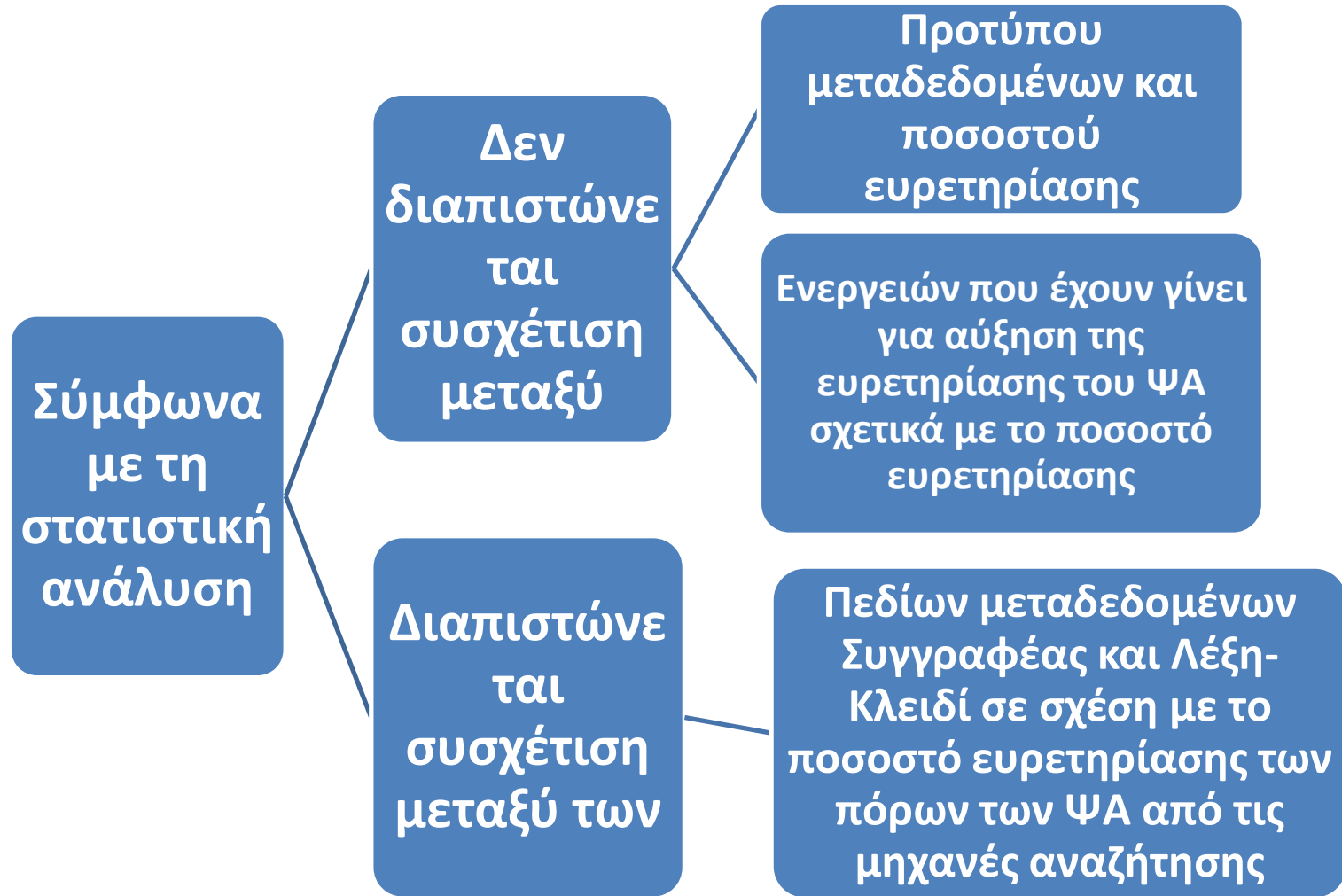
Κυριότερα ευρήματα – Συσχετίσεις (4/4)



Ποιο είναι το ποσοστό ευρηματισμού του περιεχομένου του ψηφιακού αποθετηρίου σας από τις μηχανές αναζήτησης όπως το Google Scholar;

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

Συμπεράσματα



Συνοψίζοντας

- Επαληθεύεται η υπόθεση ότι ορισμένα στοιχεία μεταδεδομένων παίζουν πολύ σημαντικό ρόλο στην ανάκτηση των πληροφοριακών πόρων των ΨΑ από τις μηχανές αναζήτησης όπως το GS
- Παρ' όλο που δεν επαληθεύονται οι άλλες δύο υποθέσεις δημιουργούνται ερεθίσματα για περαιτέρω έρευνα
- Προτείνονται διάφορες ενέργειες οι οποίες μπορούν να ληφθούν από τα ΨΑ για να πετύχουν καλύτερα αποτελέσματα στην ευρετηρίασή τους από τις μηχανές αναζήτησης όπως το GS
- Επίσης, επισημαίνονται δυσκολίες που αντιμετωπίζουν οι εργαζόμενοι στα ΨΑ σχετικά με την περιγραφή των πόρων και τη διάθεση των μεταδεδομένων. Η λύση των οποίων θα βοηθήσει στην επίτευξη καλύτερων ποσοστών ευρετηρίασης από τις μηχανές αναζήτησης όπως το GS

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

Σας ευχαριστώ πολύ για την προσοχή σας!

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