



**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ &
ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ**

**ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**

**ΨΗΦΙΑΚΟ ΣΥΣΤΗΜΑ ΑΝΑΛΥΣΗΣ ΒΙΒΛΙΟΓΡΑΦΙΚΩΝ
ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΗΝ ΠΗΓΗ GOOGLE SCHOLAR**

ΦΟΙΤΗΤΗΣ : ΠΑΠΑΧΡΗΣΤΟΥ ΑΡΓΥΡΗΣ

ΑΜ : 45387

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ : ΣΤΑΥΡΑΚΑΣ ΗΛΙΑΣ

Ψηφιακό σύστημα ανάλυσης βιβλιογραφικών δεδομένων από την πηγή Google Scholar

Η Διπλωματική Εργασία έγινε αποδεκτή και βαθμολογήθηκε από την εξής τριμελή επιτροπή:

(Όνοματεπώνυμο), (βαθμίδα)	(Όνοματεπώνυμο), (βαθμίδα)	(Όνοματεπώνυμο), (βαθμίδα)
(Υπογραφή)	(Υπογραφή)	(Υπογραφή)

Copyright © Με επιφύλαξη παντός δικαιώματος. All rights reserved.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ και (Παπαχρήστου Αργύρης),

Μάρτιος 2024

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα του και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις θέσεις του επιβλέποντος, της επιτροπής εξέτασης ή τις επίσημες θέσεις του Τμήματος και του Ιδρύματος.

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

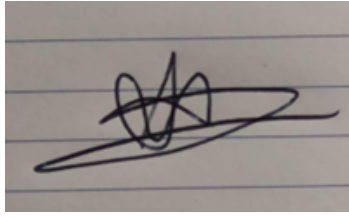
Ο κάτωθι υπογεγραμμένος Παπαχρήστου Αργύρης του Βασιλείου, με αριθμό μητρώου 45387 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής ΜΗΧΑΝΙΚΩΝ του Τμήματος ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ,

δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου.

Ο Δηλών
Παπαχρήστου Αργύρης

A handwritten signature in black ink on lined paper. The signature is stylized and appears to be 'Χρηστος Παπαχρηστου'.

ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία παρέχει μια εις βάθος ανάλυση της χρησιμότητας του Google Scholar ως πηγή επιστημονικής πληροφόρησης. Η εργασία διαρθρώνεται σε πέντε κεφάλαια, με το πρώτο κεφάλαιο να εισάγει την ανάπτυξη του Google Scholar και τον τρόπο λειτουργίας του. Το δεύτερο κεφάλαιο αναλύει τα πλεονεκτήματα και τα μειονεκτήματα του Google Scholar, ενώ το τρίτο κεφάλαιο παρουσιάζει μια βιβλιογραφική ανασκόπηση της καταλληλότητας του Google Scholar ως πηγής επιστημονικής πληροφόρησης σε σχέση με άλλες βάσεις δεδομένων. Το τέταρτο κεφάλαιο είναι η ενότητα υλοποίησης ενός λογισμικού αυτοματοποιημένης άντλησης δεδομένων από την πηγή google scholar για την εξαγωγή του έργου ενός ή και μίας ομάδας συγγραφέων. Στο πλαίσιο αυτό χρησιμοποιείται η υπηρεσία serpari.com για την εξαγωγή δεδομένων από το Google Scholar και η ίδια η εφαρμογή υλοποιήθηκε σε περιβάλλον Flask. Τέλος, στο πέμπτο κεφάλαιο παρουσιάζονται τα συμπεράσματα της εργασίας, τα οποία αναδεικνύουν τα πλεονεκτήματα της χρήσης του serpari.com για την εξαγωγή δεδομένων από το Google Scholar και τη χρησιμότητα της εφαρμογής Flask. Η χρήση του serpari.com είναι μια καινοτόμος προσέγγιση που λύνει το πρόβλημα των περιορισμών που αντιμετωπίζονται κατά τη χρήση Javascript.

Λέξεις Κλειδιά: Google Scholar, Μηχανική Μάθηση, Serpari

ABSTRACT

This work provides an in-depth analysis of the usefulness of Google Scholar as a source of scholarly information. The paper is structured in five chapters, with the first chapter introducing the development of Google Scholar and how it works. The second chapter discusses the advantages and disadvantages of Google Scholar, while the third chapter presents a literature review of the suitability of Google Scholar as a source of scholarly information relative to other databases. The fourth chapter is the implementation section, where the serpapi.com service is used to extract data from Google Scholar and a Flask application is created. Finally, the fifth chapter presents the conclusions of the work, which highlight the advantages of using serpapi.com to extract data from Google Scholar and the usefulness of the Flask application. Using serpapi.com to extract data from Google Scholar is an innovative approach that solves several problems that arise while using Javascript.

Keywords: Google Scholar, Machine Learning, Serpapi

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	5
ABSTRACT	6
ΑΛΦΑΒΗΤΙΚΟ ΕΥΡΕΤΗΡΙΟ	11
ΕΙΣΑΓΩΓΗ	12
Αντικείμενο της διπλωματικής	12
Σκοπός και στόχοι	12
Μεθοδολογία	13
Καινοτομία	13
Δομή	13
Κεφάλαιο 1 . Ανάπτυξη και Τρόπος Λειτουργία του Google Scholar	14
1.1. Η Έννοια «Ακαδημαϊκός»	16
1.2. Ποιότητα, Ακρίβεια και Λογοδοσία	17
1.3. Παραπομπές, Μετρήσεις και Ζητήματα Ένταξης	19
1.4. Βελτιστοποίηση Μηχανών Αναζήτησης Ακαδημαϊκών Άρθρων	21
Κεφάλαιο 2 . Πλεονεκτήματα και Μειονεκτήματα του Google Scholar	22
2.1. Πλεονεκτήματα του Google Scholar	22
2.1.1. Κίνητρα για Ποιότητα	22
2.1.2. Ακαδημαϊκός Συντονισμός και Ανοικτή Πρόσβαση στην Έρευνα	23
2.1.3. Συνέπεια στην Αξιολόγηση της Έρευνας	23
2.2. Μειονεκτήματα του Google Scholar	24
2.2.1. Πρωτοτυπία και Καινοτομία	26
2.2.2. Ερευνητές Πρώιμης Καριέρας	27
2.2.3. Μέγεθος Κοινότητας	27
Κεφάλαιο 3 . Βιβλιογραφική Ανασκόπηση για την Καταλληλότητα του Google Scholar ως Πηγή Επιστημονικών Πληροφοριών σε Σχέση με Άλλες Βάσεις	28
3.1. Σύγκριση του Google Scholar με Άλλες Ακαδημαϊκές Μηχανές Αναζήτησης	35

3.2. Επιλογές για την Παρακολούθηση Παραπομπών στο Google Scholar και σε Άλλες Μηχανές Αναζήτησης Ακαδημαϊκών Μελετών	38
3.3. Μελέτες Σχετικές με τον Αλγόριθμο Κατάταξης του Google Scholar	40
Κεφάλαιο 4. Υλοποίηση και άντληση δεδομένων από την βάση του Google Scholar	43
4.1. Χρήση API	46
4.2. Ανάπτυξη ιστοσελίδας με Flask	47
4.2.1. Προαπαιτούμενα	47
4.2.2. Εφαρμογή flask	48
Κεφάλαιο 5 . Συμπεράσματα	52
ΒΙΒΛΙΟΓΡΑΦΙΑ	64

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1. Η διαδικασία που ακολουθείται κατά την διατύπωση ενός ερωτήματος στο Google Scholar με στόχο την επεξεργασία και την παρουσίαση του τελικού αποτελέσματος 44

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1. Η διαδικασία για την μεταφορά απο το Google Scholar στην επιλογή Metrics	14
Εικόνα 2. Η λειτουργία του δείκτη h5	15
Εικόνα 3. Αποτελέσματα κλήσης για το ερώτημα «Machine Learning».	46
Εικόνα 4. Η αρχική σελίδα της φόρμας που ο χρήστης δίνει το ερώτημα προς αναζήτηση και προαιρετικά επιλέγει έτη (από-μέχρι).	48
Εικόνα 5. Παράδειγμα χρήσης της φόρμας.	50
Εικόνα 6. Το αποτέλεσμα μετά την εισαγωγή για αναζήτηση του λεκτικού “Machine Learning” για έτη μετά το 2018.	50
Εικόνα 7. Το αποτέλεσμα μετά την εισαγωγή για αναζήτηση του λεκτικού “Machine Learning” για έτη μετά το 2018.	51

ΑΛΦΑΒΗΤΙΚΟ ΕΥΡΕΤΗΡΙΟ

ASEBD - Academic search engines and bibliographic databases

ASEO - Academic Search Engine Optimization

GS - Google Scholar

JCR - Journal Citation Reports

RFID - Radio frequency identification

SEO - Search engine optimization

ΕΙΣΑΓΩΓΗ

Αντικείμενο της διπλωματικής

Το αντικείμενο της παρούσας εργασίας είναι η χρήση του Google Scholar ως πηγή επιστημονικής πληροφόρησης και η δημιουργία μιας εφαρμογής εξαγωγής δεδομένων χρησιμοποιώντας την υπηρεσία serpari.com. Η εργασία περιγράφει επίσης τα πλεονεκτήματα και τα μειονεκτήματα της χρήσης του Google Scholar ως πηγής επιστημονικών πληροφοριών. Παράλληλα υπάρχει η ανάλογη σύγκριση όσο αφορά την καταλληλότητα του Google Scholar σε σχέση με άλλες βάσεις δεδομένων. Το αντικείμενο της εργασίας είναι διεπιστημονικό, καλύπτοντας θέματα στην επιστήμη των υπολογιστών, την ανάκτηση πληροφοριών και την εκπαίδευση. Η παρούσα διατριβή στοχεύει να συμβάλει στην πρόοδο της εκπαίδευσης της μηχανικής μάθησης και να παρέχει χρήσιμες πληροφορίες και καθοδήγηση σε ερευνητές και φορείς σχετικά με τη χρήση του Google Scholar.

Σκοπός και στόχοι

Ο σκοπός της εργασίας είναι η μελέτη της αξιοπιστίας της χρήσης του Google Scholar ως πηγή επιστημονικής πληροφόρησης και η παρουσίαση μια μεθόδου εξαγωγής πληροφοριών από αυτό. Οι στόχοι της εργασίας είναι να παρέχει μια επισκόπηση της ανάπτυξης και λειτουργίας του Google Scholar, να περιγράψει τα πλεονεκτήματα και τα μειονεκτήματά του ως πηγή επιστημονικών πληροφοριών, να εξετάσει την καταλληλότητα του Google Scholar σε σχέση με άλλες βάσεις δεδομένων, να παρουσιάσει μια στρατηγική υλοποίησης για εξαγωγή δεδομένων από το Google Scholar χρησιμοποιώντας την υπηρεσία serpari.com και να επειδιξει την δημιουργία μιας εφαρμογής Flask.

Μεθοδολογία

Το αντικείμενο της εργασίας συνδυάζει τη βιβλιογραφική ανασκόπηση και την ανάπτυξη λογισμικού. Η ανασκόπηση της βιβλιογραφίας διεξάγεται ώστε να αναλυθεί ο τρόπος λειτουργίας και η επάρκεια του Google Scholar ως πηγής επιστημονικών πληροφοριών σε σχέση με άλλες βάσεις δεδομένων. Ο σχεδιασμός και η ανάπτυξη του λογισμικού στηρίζεται στην υπηρεσία serpari.com που χρησιμοποιείται για τη λήψη δεδομένων από το Google Scholar και στην Python για την εξαγωγή και την εμφάνιση των δεδομένων.

Καινοτομία

Η καινοτομία της εργασίας έγκειται στον συνδυασμό δύο διαφορετικών πτυχών: τη χρήση της υπηρεσίας serpari.com για την εξαγωγή δεδομένων από το Google Scholar. Παρόλο που η χρήση API για την εξαγωγή δεδομένων από πλατφόρμες ιστού δεν είναι μια νέα ιδέα, η συγκεκριμένη εφαρμογή του serpari.com και η λεπτομερής μέθοδος που παρουσιάζεται στην εργασία μπορούν να θεωρηθούν μια καινοτόμος προσέγγιση.

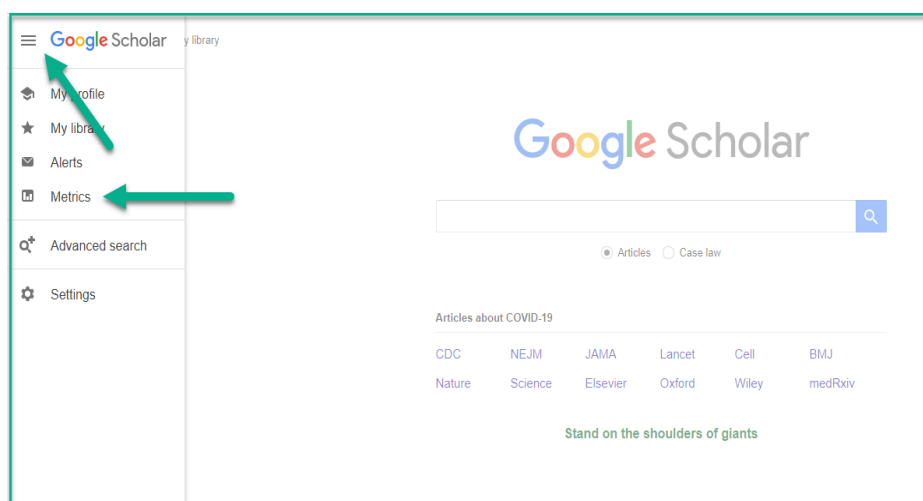
Δομή

Η παρούσα εργασία αποτελείται από την εισαγωγή και πέντε επιμέρους κεφάλαια. Η εισαγωγή θέτει τη βάση για την εργασία παρέχοντας μια σύντομη επισκόπηση της εργασίας, όπου αναλύεται το αντικείμενο, ο σκοπός και οι στόχοι της, η μεθοδολογία που ακολουθήθηκε και η καινοτομία της. Το Κεφάλαιο 1 εστιάζει στην ανάπτυξη και τη λειτουργία του Google Scholar. Το Κεφάλαιο 2 εξετάζει τα πλεονεκτήματα και τα μειονεκτήματα της χρήσης του Google Scholar ως πηγής επιστημονικών πληροφοριών. Το Κεφάλαιο 3 παρέχει μια βιβλιογραφική ανασκόπηση που συγκρίνει το Google Scholar με άλλες βάσεις δεδομένων, επισημαίνοντας τα δυνατά και τα αδύνατα σημεία του. Το Κεφάλαιο 4 παρουσιάζει τις λεπτομέρειες υλοποίησης της εργασίας και τέλος, το Κεφάλαιο 5 ολοκληρώνει την εργασία συνοψίζοντας τα κύρια ευρήματα, συζητώντας τις επιπτώσεις της εργασίας και προτείνοντας μελλοντικές κατευθύνσεις έρευνας.

Κεφάλαιο 1. Ανάπτυξη και Τρόπος Λειτουργία του Google Scholar

Το Google Scholar (www.scholar.google.com) δημιουργήθηκε το 2004 ως μηχανή αναζήτησης για ακαδημαϊκούς. Όπως και η γενική μηχανή αναζήτησης της Google, παράγει αποτελέσματα με βάση τη δύναμη της σύνδεσης μεταξύ των όρων αναζήτησης και το πόσο συχνά και πόσο πρόσφατα έχει αναφερθεί ένα έργο. Το Google Scholar ευρετηριάζει σχεδόν τα πάντα που είναι διαθέσιμα στο διαδίκτυο σε οποιαδήποτε γλώσσα, συμπεριλαμβανομένων άρθρων σε περιοδικά, επιστημονικά βιβλία, καθώς και υλικού που δεν έχει αξιολογηθεί από κριτές, όπως πρακτικά συνεδρίων, διατριβές και διδακτορικές διατριβές.

Το Google Scholar συμπεριλαμβάνει επίσης τα 20 κορυφαία περιοδικά σε έναν επιστημονικό κλάδο. Στο πάνω μέρος της αρχικής σελίδας του Google Scholar, στην επιλογή «Metrics» (βλέπε Εικόνα 1) παρέχει συνδέσμους προς τις κατατάξεις ανά επιστημονικό κλάδο και γλώσσα, χρησιμοποιώντας έναν «δείκτη H5» δηλαδή τον αριθμό X άρθρων που έχουν τουλάχιστον X αναφορές τα προηγούμενα πέντε χρόνια. Από το 2012, η Google επιτρέπει στους μελετητές να δημιουργούν επεξεργάσιμες «σελίδες προφίλ» του Google Scholar, τις οποίες συμπληρώνει αυτόματα με συνδέσμους που παραπέμπουν σε υλικό που έχει γράψει ένας συγγραφέας και είναι αρχειοθετημένο στο διαδίκτυο. Στην κεντρική σελίδα (προφίλ) επιπλέον αναφέρεται πόσες φορές έχει αναφερθεί (cited) κάθε άρθρο από τον ίδιο ή από άλλους ερευνητές.



Εικόνα 1. Η παρουσίαση της αρχικής σελίδα του Google Scholar και το πως μπορεί ο χρήστης να μεταφερθεί στην επιλογή Metrics.

Google Scholar

Top publications

Categories ▾ English ▾

h5-median for a publication is the median number of citations for the articles that make up its h5-index. [hide](#)

Publication	h5-index	h5-median
1. Nature	376	552
2. The New England Journal of Medicine	365	639
3. Science	356	526

Εικόνα 2. Η αναφορά στον δείκτη h5 όπου υπολογίζεται με βάση τα άρθρα που δημοσιεύτηκαν από περιοδικό τα τελευταία πέντε χρόνια.

Το Google Scholar ως έγκυρη και κατάλληλη πηγή για επιστημονικές πληροφορίες είναι φυσικά το θέμα πολλών συζητήσεων. Σε μια μελέτη του 2005, το 24% των ερευνητικών βιβλιοθηκών της Βόρειας Αμερικής συμπεριέλαβαν το Google Scholar στους καταλόγους των βάσεων δεδομένων τους και λιγότερο από το 20% των ακαδημαϊκών βιβλιοθηκών το συνιστούσαν ως μηχανή αναζήτησης για έρευνα (Mullen & Hartman, 2006). Οι ερευνητές, οι ειδικοί των επιστημονικών δεικτών και βιβλιοθηκών έχουν την υποχρέωση να γνωρίζουν την αξιοπιστία των πηγών πληροφόρησης άρα και τα δυνατά και τα αδύνατα σημεία του Google Scholar ως εργαλείο επιστημονικής πληροφόρησης. Τα εμπειρικά δεδομένα - είτε μέσω πειραματισμού είτε μέσω παρατήρησης - και οι επακόλουθες αναλύσεις τους χρησιμεύουν ως θεμέλιο της επιστημονικής γνώσης. Τα κριτήρια και τα μοντέλα για τη συλλογή δεδομένων είναι σαφώς καθορισμένα και έχουν διαδοθεί στην επιστημονική κοινότητα μέσω της επιστημονικής βιβλιογραφίας. Τα πληροφοριακά συστήματα που υποστηρίζουν τους εκδοτικούς οίκους επιστημονικών δημοσιεύσεων ενημερώνουν τους χρήστες για τη σημασία της αναγνώρισης έγκυρων επιστημονικών πηγών. Το περιεχόμενο που βρίσκεται μέσω υπηρεσιών καταλόγων, σε ακαδημαϊκά αποθετήρια και σε κυβερνητικούς δικτυακούς τόπους υποβάλλεται σε κάποιο επίπεδο ελέγχου, είτε μέσω αξιολόγησης, είτε μέσω συντακτικής διαδικασίας (editing), εξασφαλίζοντας ένα επίπεδο ευθύνης για την ποιότητα των στοιχείων που παρουσιάζονται. Ωστόσο, με την αυξημένη δημοτικότητα των μηχανών αναζήτησης όπως το Google Scholar, που δεν συμμορφώνονται με τις διαδικασίες πιστοποίησης εγκυρότητας επιστημονικών άρθρων, η γραμμή μεταξύ της επιστημονικής βιβλιογραφίας και της ψευδοεπιστήμης δεν είναι πλέον τόσο σαφής.

1.1. Η Έννοια «Ακαδημαϊκός»

Οι ερευνητές έχουν την ικανότητα να κάνουν διακρίσεις μεταξύ των επιστημονικών πληροφοριών και εκείνων που μπορεί να είναι δημοφιλείς, άσχετες ή αναξιόπιστες. Επιπλέον αντιλαμβάνονται ότι η φύση της επιστήμης είναι καθοριστικής σημασίας, διότι συνδέεται άρρηκτα με την ανάκτηση και την αξιολόγηση των επιστημονικών πληροφοριών" (Lascar & Mendelsohn 2011a). Η ακρίβεια, η αυθεντικότητα και η επικαιρότητα, είναι τρεις από τις σημαντικότερες ιδιότητες της ακαδημαϊκής επιστημονικής βιβλιογραφίας, όπως υποστηρίζεται από το «McLean v Arkansas Documentation Project» του 2005, περιγράφοντας ότι: «Καθοδηγούνται και εξηγούνται από τον φυσικό νόμο, μπορεί να ελεγχθούν σε σχέση με τον εμπειρικό κόσμο, τα συμπεράσματα είναι προσωρινά και μπορούν να διαψευστούν». (Lascar & Mendelsohn 2011a).

Καθώς το Google Scholar δεν δίνει σαφή απάντηση για τον ορισμό του «ακαδημαϊκού», οι Lascar & Mendelsohn 2011b διαμόρφωσαν έναν ορισμό χρησιμοποιώντας έναν συνδυασμό δηλώσεων από τις ιστοσελίδες που περιγράφουν τη λειτουργία του Google Scholar. Μπορεί να υποτεθεί ότι το πρόγραμμα ανίχνευσης που χρησιμοποιείται για τον εντοπισμό επιστημονικών πληροφοριών χρησιμοποιεί απλώς τα μεταδεδομένα και την ηλεκτρονική τοποθεσία (πανεπιστήμια, αποθετήρια, περιοδικά, άλλοι δικτυακοί τόποι) για να προσδιορίσει τον επιστημονικό χαρακτήρα των ηλεκτρονικών εγγράφων (About Google Scholar). Προφανώς, αν ένα έγγραφο σε έναν προσωπικό ιστότοπο έχει τα σωστά μεταδεδομένα μπορεί να συμπεριληφθεί, είτε είναι επιστημονικό είτε όχι. Το Google Scholar δεν κάνει διάκριση μεταξύ επιστήμης και ψευδοεπιστήμης, ούτε κάνει διάκριση μεταξύ θετικών και ανθρωπιστικών επιστημών ή άλλων. Φαίνεται να χαρακτηρίζει τις επιστημονικές πληροφορίες με βάση τη θέση τους και τα σχετικά μεταδεδομένα και, ταυτόχρονα, δεν μπορεί να ελέγξει την ποιότητα των εργασιών που εντοπίζονται κατά τη χρήση του.

Μεγάλο μέρος των ερευνών που έχουν δημοσιευτεί μέχρι σήμερα σχετικά με το Google Scholar έχει επικεντρωθεί στη σύγκριση του Google Scholar με μια ποικιλία διαφορετικών επιστημονικών βάσεων δεδομένων. Συγκρίσεις μεταξύ PubMed, BIOSIS Previews, SciFinder, Web of Science και άλλων έχουν γίνει αρκετά διεξοδικά. Τα δεδομένα που συλλέχθηκαν έχουν καλύψει επαρκώς την αποτελεσματικότητα ή τις ελλείψεις του Google Scholar ως μηχανή αναζήτησης σε σύγκριση με τις υπάρχουσες μηχανές. Αυτό που δεν έχει επαρκώς μελετηθεί είναι η αξιοπιστία των δεδομένων δεδομένου ότι οι ακαδημαϊκοί και

επιστημονικοί κύκλοι έχουν προσδοκίες για διαφάνεια και πρόσβαση σε επιστημονικές πληροφορίες, τα ερευνητικά ερωτήματα που θα πρέπει να διερευνηθούν περιλαμβάνουν:

- πώς το Google Scholar ανακτά πληροφορίες,
- ποιες πληροφορίες παρουσιάζει το Google Scholar, και
- ποιες πιθανές τεχνικές, επαγγελματικές ή εμπορικές εκτιμήσεις οδηγούν τον τρόπο συλλογής και παρουσίασης της βιβλιογραφίας στους χρήστες που πραγματοποιούν αναζήτηση

1.2. Ποιότητα, Ακρίβεια και Λογοδοσία

Όπως αναφέρουν οι Mayr και Walter (2008), το Google Scholar είναι μια ελεύθερα διαθέσιμη υπηρεσία με ένα φιλικό περιβάλλον εργασίας παρόμοιο με το Google Web Search. Τα αποτελέσματα που ανακτώνται από μια αναζήτηση στο Google Scholar είναι μια σημαντική πτυχή του ερευνητικού έργου που πρέπει να αξιολογηθεί και να ληφθεί υπόψη. Σύμφωνα με προηγούμενες μελέτες και με προσπάθειες των συγγραφέων φαίνεται ότι η Google δεν αποκαλύπτει συγκεκριμένες πηγές που χρησιμοποιούνται για τη δημιουργία του ευρετηρίου του Google Scholar. Στη σελίδα βοήθειας του Google Scholar αναφέρεται ότι το Google Scholar περιλαμβάνει άρθρα σε περιοδικά και συνέδρια, διατριβές και διδακτορικές διατριβές, ακαδημαϊκά βιβλία, προεκτυπώσεις (preprints), περιλήψεις, τεχνικές εκθέσεις και άλλη επιστημονική βιβλιογραφία από όλους τους ευρείς τομείς της έρευνας (Scholar Help). Αναφέρει επίσης ότι περιλαμβάνει έργα από μια μεγάλη ποικιλία εκδοτικών οίκων, εταιρειών και πανεπιστημιακών αποθετηρίων, καθώς και επιστημονικά άρθρα που είναι διαθέσιμα οπουδήποτε στο διαδίκτυο, δικαστικές αποφάσεις και διπλώματα ευρεσιτεχνίας. Επαναλαμβανόμενα μηνύματα ηλεκτρονικού ταχυδρομείου προς το Google Scholar με αίτημα να δοθεί κατάλογος των περιοδικών και άλλων πηγών που ανιχνεύθηκαν δεν έλαβαν σαφείς απαντήσεις και δόθηκαν μόνο σύνδεσμοι προς την τεκμηρίωση βοήθειας του Google Scholar και τις σελίδες About Google Scholar. Αν και οι διαθέσιμες πληροφορίες είναι χρήσιμες, ο σκοπός της επικοινωνίας με το Google Scholar ήταν να ληφθούν πιο εμπειριστατωμένες και ακριβείς πληροφορίες σχετικά με τη μηχανή αναζήτησης και τις πηγές που παρουσιάζει στα αποτελέσματα. Το Google Scholar έχει προσθέσει έναν κατάλογο με τίτλο "Κορυφαίες δημοσιεύσεις στα αγγλικά" ως μέρος της σελίδας Google Scholar Metrics (Top Publications), αλλά δεν υπάρχει πλήρης κατάλογος εργασιών και δεν αναφέρεται αν αυτός πρόκειται να δοθεί. Τα αιτήματα για πληροφορίες σχετικά με τα κριτήρια κατάταξης απαντήθηκαν επίσης μη ικανοποιητικά. Η ασάφεια των κριτηρίων επιλογής των

επιστημονικών άρθρων δημιουργεί ανησυχίες σχετικά με το πόσο ολοκληρωμένα είναι αυτά ή ακόμα και τον βαθμό χειραγώγησής τους.

Ένας λόγος που το Google Scholar δεν παρέχει περαιτέρω βοήθεια ή πλήρη κατάλογο των πηγών ήταν ότι πρόκειται για μια δωρεάν υπηρεσία. Όταν ρωτήθηκε: «Υπάρχουν τυποποιημένες πηγές για αυτά τα άρθρα που ανιχνεύονται σε τακτική βάση; Αν ναι, ποιες είναι αυτές οι πηγές; (Hauser [2012])», το Google Scholar απάντησε: «Αυτό καλύπτεται στις σελίδες βοήθειας και τα περισσότερα είναι προφανή αν κάνετε μερικές αναζητήσεις. Λυπούμαστε, δεν είμαστε σε θέση να σας βοηθήσουμε με μεγάλη λεπτομέρεια, πρόκειται για μια δωρεάν υπηρεσία (Google Scholar Team [2012])». Η στάση του Google Scholar φαίνεται να είναι ότι οι χρήστες πρέπει απλώς να είναι ευχαριστημένοι που η υπηρεσία αυτή διατίθεται δωρεάν και οι χρήστες δεν δικαιούνται άμεσες απαντήσεις σχετικά με τις πηγές του.

Το Google Scholar έχει επίσης διατυπώσει μια ενδιαφέρουσα δήλωση σχετικά με την κάλυψη των δημοσιεύσεων, η οποία εγείρει και πάλι ανησυχίες σχετικά με την έλλειψη οποιασδήποτε ουσιαστικής περιγραφής των πηγών του. Στη σελίδα Google Scholar Metrics αναφέρεται: "δεδομένου ότι το Google Scholar καταχωρεί σε ευρετήριο άρθρα από μεγάλο αριθμό δικτυακών τόπων, δεν μπορούμε πάντα να γνωρίζουμε πού (ή αν!) έχει δημοσιευθεί ένα συγκεκριμένο άρθρο". Αυτή η πρόταση έχει ένα αναδεικνύει ένα σημαντικό πρόβλημα: Ποιοτικό επιστημονικό υλικό είναι αυτό που έχει ελεγχθεί με κάποιο τρόπο για να διασφαλιστεί η ποιότητα, η ακρίβεια και το κύρος του. Παρόλο που ένα μεγάλο μέρος από αυτό που ανακτά μια αναζήτηση στο Google Scholar είναι επιστημονικά, υπάρχει ακόμα η πιθανότητα μια αναζήτηση στο Google Scholar να επιστρέψει κάποιο υλικό που δεν ανταποκρίνεται στο κύρος και την ποιότητα που αναμένουν οι επιστήμονες και οι ερευνητές. Το Google Scholar προσπαθεί να ελαχιστοποιήσει τα σφάλματα στον εντοπισμό δημοσιεύσεων ελέγχοντας ποιοι τύποι εγγράφων περιλαμβάνονται για τους αλγόριθμους μέτρησης, αλλά η διαδικασία αυτή εξακολουθεί να είναι ελαττωματική. Αν αυτό το εργαλείο δεν μπορεί να πει πού ή αν έχει δημοσιευτεί ένα άρθρο, πώς μπορεί κανείς να γνωρίζει αν το υλικό που βρίσκεται στο Google Scholar είναι ακριβές, έγκυρο ή έχει κύρος σε έναν τομέα έρευνας;

1.3. Παραπομπές, Μετρήσεις και Ζητήματα Ένταξης

Οι αναφορές του Google Scholar είναι ένας άλλος παράγοντας που πρέπει να ληφθεί υπόψη. Οι Egghe και Rousseau υποστηρίζουν ότι ένας υψηλός αριθμός παραπομπών όχι μόνο υποδηλώνει τη χρήση του εν λόγω δημοσιευμένου άρθρου, αλλά επίσης προσθέτει αξιοπιστία στην εργασία και καταδεικνύει την αποδοχή σε ένα συγκεκριμένο ερευνητικό πεδίο. Επομένως, η συγκέντρωση αυτών των αριθμών παραπομπών από το Google Scholar είναι σημαντική, καθώς αυτοί οι αριθμοί παραπομπών μπορούν να καθορίσουν ποια άρθρα επιλέγονται και περιλαμβάνονται στην πλαισίωση της εισαγωγικής ιστορικής αναδρομής από τους ερευνητές.

Οι μετρήσεις παραπομπών στο Google Scholar περιλαμβάνουν μόνο υλικό που έχει δημοσιευθεί τα τελευταία πέντε χρόνια και περιλαμβάνουν μόνο άρθρα περιοδικών που ακολουθούν τις κατευθυντήριες γραμμές ένταξης του Google Scholar και έναν μικρό αριθμό χειροκίνητα εντοπισμένων εγγράφων συνεδρίων και (preprints). Το Google Scholar αναφέρει επίσης στον ιστότοπό του ότι αποκλείει ορισμένα στοιχεία, συγκεκριμένα "δικαστικές αποφάσεις, διπλώματα ευρεσιτεχνίας, βιβλία και διατριβές - δημοσιεύσεις με λιγότερα από 100 άρθρα που δημοσιεύτηκαν μεταξύ 2007 και 2011 - και δημοσιεύσεις που δεν έλαβαν καμία αναφορά σε άρθρα που δημοσιεύτηκαν μεταξύ 2007 και 2011" (Google Scholar Metrics). Το Google Scholar δεν δίνει συγκεκριμένους λόγους για τις αποφάσεις του να συμπεριλάβει ή να αποκλείσει ορισμένα στοιχεία, εκτός από την "αποφυγή εσφαλμένης ταυτοποίησης δημοσιεύσεων". Αυτές οι αποφάσεις συμπερίληψης/αποκλεισμού μπορούν να επηρεάσουν τον αριθμό των αναφορών, δεδομένου του ότι περιλαμβάνεται στις μετρήσεις επαφίεται στο Google Scholar και στο τι αποφασίζει ότι είναι σε θέση να συμπεριληφθεί.

Ο αριθμός των αναφορών στο Google Scholar μπορεί να επηρεαστεί με διάφορους τρόπους. Ορισμένοι εκδότες μπορεί να επιλέξουν να μην επιτρέψουν στο Google Scholar να ανιχνεύσει το υλικό τους: "Η ACS, ως ένας από τους κυρίαρχους εκδότες στον τομέα της χημείας, δεν συνεργάζεται με το Google, γεγονός που προκαλεί σημαντική απώλεια αναφορών και απαγορεύει τη χρήση του Google Scholar στην ανάλυση των αναφορών" (Bornmann και συνεργάτες, 2009). Άλλοι εκδότες που επιτρέπουν στο Google Scholar να ανιχνεύει το υλικό τους μπορεί να έχουν ένα ξεχωριστό πλεονέκτημα ή μειονέκτημα για την ανακάλυψη μέσω του Google Scholar με βάση τους πόρους τους. Οι οδηγίες ένταξης, ακόμη

και όπως περιγράφονται από το Google Scholar, είναι μακροσκελείς και πυκνές. Οι εκδοτικοί οίκοι με ισχυρό τεχνικό προσωπικό και πόρους είναι πιθανότερο να πληρούν αυτές τις τεχνικές κατευθυντήριες γραμμές ένταξης και, κατά συνέπεια, να αυξάνουν την ένταξη και την κατάταξη του περιεχομένου τους στο Google Scholar. Το περιεχόμενο που δημοσιεύτηκε πριν από τις κατευθυντήριες γραμμές συμπερίληψης του Google Scholar ή από μικρότερους εκδότες με λιγότερους πόρους μπορεί να αποκλειστεί ή να καταταχθεί χαμηλότερα από τους μεγαλύτερους εκδότες. Αυτό θα μπορούσε να οδηγήσει σε απώλεια της πληρότητας των επιστημονικών αναζητήσεων.

Παρόλο που το Google Scholar αποκλείει κάποιο υλικό σε μια προσπάθεια να διατηρήσει ακριβείς τις μετρήσεις παραπομπών, εξακολουθεί να ανιχνεύει στοιχεία όπως εγχειρίδια φοιτητών, οδηγούς βιβλιοθήκης και εκδοτικές σημειώσεις (Harzing 2010) για χρήση στις μετρήσεις παραπομπών, εφόσον τα έγγραφα αυτά ακολουθούν τις οδηγίες συμπερίληψης του Google Scholar. Αυτά τα υλικά ενδέχεται να μην θεωρούνται επιστημονικά ή άξια πραγματικής αναφοράς. Αυτό μπορεί να οδηγήσει σε διογκωμένες τιμές για τις μετρήσεις παραπομπών και, με τη σειρά του, να υπερτιμήσει την ακαδημαϊκή αξία του άρθρου και να αλλάξει τον τρόπο κατάταξής του στις λίστες αποτελεσμάτων του Google Scholar.

Η σελίδα Google Scholar Metrics παρέχει ορισμένες πληροφορίες σχετικά με τον τρόπο με τον οποίο χρησιμοποιούνται οι μετρήσεις των παραπομπών για την κατάταξη των περιοδικών (Google Scholar Metrics). \ Η επικαιρότητα είναι συχνά μείζονος σημασίας και προβληματισμού για τους επιστήμονες και οι μετρήσεις του Google Scholar δεν είναι απαραίτητα ενημερωμένες. Ενώ η επικαιρότητα της κατάταξης των περιοδικών του Google Scholar και οι αριθμοί παραπομπών δεν αποτελούν την πρωταρχική ανησυχία των περισσότερων, οι υπερβολικοί ή ανακριβείς αριθμοί παραπομπών μπορεί να είναι παραπλανητικοί.

1.4. Βελτιστοποίηση Μηχανών Αναζήτησης Ακαδημαϊκών Άρθρων

Υπάρχει μεγάλο ενδιαφέρον στην ακαδημαϊκή κοινότητα προκειμένου να μπορούν οι ερευνητές να διασφαλίζουν ότι τα άρθρα τους μπορούν να ευρετηριάζονται από ακαδημαϊκές μηχανές αναζήτησης, όπως το Google Scholar, το IEEE Xplore, το PubMed και το SciPlore.org, αφού με τον τρόπο αυτό τα άρθρα τους καθίστανται διαθέσιμα στην ακαδημαϊκή κοινότητα. Ενδιαφέρον μεταξύ των συγγραφέων αποτελεί όχι μόνο αν τα άρθρα τους είναι ευρετηριασμένα αλλά και σε ποια θέση στην κατάταξη εμφανίζονται. Όπως και σε κάθε κατηγορία αποτελεσμάτων αναζήτησης κατάταξης, μεγαλύτερη πιθανότητα να διαβαστούν έχουν τα άρθρα που εμφανίζονται σε κορυφαίες θέσεις.

Στο Web, η βελτιστοποίηση μηχανών αναζήτησης (search engine optimization-SEO) για τοποθεσίες Web αποτελεί μια κοινή διαδικασία. Η διαδικασία SEO περιλαμβάνει τη δημιουργία ή την τροποποίηση ενός ιστότοπου με τρόπο που καθιστά «ευκολότερο για τις μηχανές αναζήτησης τόσο την ανίχνευση όσο και την ευρετηρίαση του περιεχομένου»¹. Υπάρχει μια τεράστια κοινότητα που συζητά τις τελευταίες τάσεις στο SEO και παρέχει συμβουλές για Webmasters σε φόρουμ, ιστολόγια και ομάδες συζητήσεων. Επιπλέον υπάρχει πλήθος ερευνητικών άρθρων και βιβλίων σχετικά με το SEO (Evans, 2007; Couzin & Grappone, 2006). Όταν ξεκίνησε το SEO, πολλοί εξέφρασαν τις ανησυχίες τους για το ότι θα προωθούσε ανεπιθύμητα μηνύματα. Στην πραγματικότητα βέβαια είναι γεγονός ότι τα ανεπιθύμητα μηνύματα στις μηχανές αναζήτησης αποτελούν ένα σοβαρό ζήτημα (Geng et al., 2008). Σήμερα, ωστόσο, το SEO είναι μια κοινή και ευρέως αποδεκτή διαδικασία και συνολικά, οι μηχανές αναζήτησης καταφέρνουν να εντοπίζουν αρκετά καλά τα ανεπιθύμητα μηνύματα. Πιθανώς το ισχυρότερο επιχείρημα για το SEO είναι το γεγονός ότι οι ίδιες οι μηχανές αναζήτησης δημοσιεύουν οδηγίες σχετικά με τον τρόπο βελτιστοποίησης ιστοσελίδων για μηχανές αναζήτησης. Αντίστοιχες πληροφορίες σχετικά με τη βελτιστοποίηση της επιστημονικής βιβλιογραφίας για τις ακαδημαϊκές μηχανές αναζήτησης δεν υπάρχουν, από ότι είναι έως τώρα γνωστό.

Κεφάλαιο 2. Πλεονεκτήματα και Μειονεκτήματα του Google Scholar

2.1. Πλεονεκτήματα του Google Scholar

Τα πλεονεκτήματα του Google Scholar εστιάζουν κυρίως στην ευκολία χρήσης του. Η σελίδα προφίλ παρέχει μια γρήγορη και φιλική επισκόπηση των δημοσιεύσεων ενός ερευνητή, με σειρά κατάταξης ανάλογα με τον αριθμό των αναφορών κάθε δημοσίευσης. Μπορεί κανείς να κάνει κλικ στους υπερσυνδέσμους κάθε δημοσίευσης για να δει περιλήψεις και να αποκτήσει πρόσβαση σε δημόσια διαθέσιμα άρθρα.

2.1.1. Κίνητρα για Ποιότητα

Η ανάγκη των ερευνητών για παραγωγή δημοσιεύσεων γίνεται ολοένα μεγαλύτερη «να δημοσιεύσουν ή να χαθούν». Στις Ηνωμένες Πολιτείες, το σύστημα μονιμοποίησης ασκεί συνεχή πίεση στο διδακτικό προσωπικό, ιδίως στο νεαρό διδακτικό προσωπικό. Στο Ηνωμένο Βασίλειο, η πίεση αυτή θεσμοθετήθηκε μέσω του συστήματος Πλαισίου Ερευνητικής Αριστείας, σύμφωνα με το οποίο η δημόσια χρηματοδότηση των πανεπιστημίων εξαρτάται από τις δημοσιεύσεις και τον «αντίκτυπο». Το Ευρωπαϊκό Συμβούλιο Έρευνας απαιτεί από όλους τους αιτούντες επιχορήγηση να παρέχουν πληροφορίες σχετικά με το ιστορικό τους, συμπεριλαμβανομένου του αριθμού των δημοσιεύσεων και των αναφορών. Σε αυτό το πλαίσιο, η εύκολη πρόσβαση στο Google Scholar μπορεί να παρέχει κίνητρα στους μελετητές να δίνουν έμφαση στην ποιότητα έναντι της ποσότητας. Μπορεί να θέλουν να παράγουν καλύτερες δημοσιεύσεις που περιλαμβάνουν περισσότερες αναφορές αντί για περισσότερες δημοσιεύσεις με λιγότερες αναφορές. Η εστίαση στον αριθμό των αναφορών μπορεί επίσης να εμπνεύσει τους συγγραφείς να δημοσιοποιούν περισσότερο το έργο τους, όπως μέσω των μέσων κοινωνικής δικτύωσης.

2.1.2. Ακαδημαϊκός Συντονισμός και Ανοικτή Πρόσβαση στην Έρευνα

Μια αναζήτηση στο Google Scholar για μια λέξη-κλειδί μπορεί να αναδείξει έναν αναγνώστη σε έναν νέο ερευνητή, ο οποίος μπορεί εύκολα να δει το υπόλοιπο έργο του ερευνητή αυτού μέσα από το προφίλ του. Ο αριθμός των αναφορών δείχνει ποια από τα άρθρα και τα βιβλία του συγγραφέα είναι τα πιο δημοφιλή. Επιπλέον, οι αυτόματες ειδοποιήσεις ηλεκτρονικού ταχυδρομείου του Google Scholar μπορούν να ενθαρρύνουν αυτή την δικτύωση. Οι ενδιαφερόμενοι αναγνώστες μπορούν να εγγραφούν για να λαμβάνουν αυτόματα μηνύματα ηλεκτρονικού ταχυδρομείου όταν οι ερευνητές δημοσιεύουν νέες εργασίες, και οι ερευνητές μπορούν να ειδοποιούνται όταν κάποιος τους κάνει αναφορά. Το Google Scholar ωθεί τους ερευνητές προς πρακτικές ανοικτής πρόσβασης (Hendrix 2016).

2.1.3. Συνέπεια στην Αξιολόγηση της Έρευνας

Το προφίλ του Google Scholar, με τον πλήρη κατάλογο της ακαδημαϊκής παραγωγής, τον αριθμό των αναφορών και τον δείκτη h, παρέχει απλές μετρήσεις της επιστημονικής ποιότητας και του αντίκτυπου. Τα μέτρα αυτά είναι εξαιρετικά αξιόπιστα (δηλαδή, όποιος τα αναζητά λαμβάνει την ίδια τιμή) και συνεπή (δηλαδή, το ίδιο σύστημα κωδικοποίησης ισχύει για όλους τους μελετητές).

Τέλος, η εύκολη διαθεσιμότητα των μετρήσεων των παραπομπών μπορεί να βοηθήσει τους ερευνητές να λάβουν αναγνώριση για βιβλία και άρθρα που δεν δημοσιεύονται στα πιο διάσημα έντυπα, αλλά παρόλα αυτά είναι σημαντικά στο πεδίο έρευνας τους. Οι μετρήσεις αναφορών μπορούν να ενημερώσουν την ερευνητική κοινότητα για την επίδραση της δημοσίευσης ανεξάρτητα από το κύρος του εκδοτικού οίκου.

2.2. Μειονεκτήματα του Google Scholar

Οι περισσότεροι θα συμφωνούσαν ότι οι καλές πρακτικές αναφορών περιλαμβάνουν την αναγνώριση προηγούμενων εργασιών που βοήθησαν στη δημιουργία ιδεών και την εξήγηση του τρόπου με τον οποίο οι ισχυρισμοί κάποιου εντάσσονται στην βιβλιογραφία. Ωστόσο, πρόσφατες μελέτες δείχνουν ότι αυτές οι κοινές πρακτικές αναφοράς θέτουν σε μειονεκτική θέση τους πρωτοπόρους ερευνητές, αυτούς που βρίσκονται στο ξεκίνημα της σταδιοδρομίας τους, αλλά και αυτούς τους ερευνητές που ανήκουν σε μικρότερες ερευνητικές κοινότητες, τις γυναίκες και τους μεμονωμένους συγγραφείς. Επιπλέον, πολλοί δημοσιεύουν άρθρα για στρατηγικούς λόγους και όχι για λόγους αρχών. Κατά συνέπεια, οι μετρήσεις των αναφορών δεν παρέχουν ακριβή εκτίμηση του επιστημονικού αντίκτυπου και ενδέχεται να επιδεινώσουν τις υφιστάμενες κοινωνικές ιεραρχίες και ανισότητες.

2.2.1. Πρωτοτυπία και Καινοτομία

Σε γενικές γραμμές, ο αριθμός των αναφορών που λαμβάνει ένα άρθρο ή βιβλίο κατά τα πέντε περίπου έτη μετά τη δημοσίευση αποκαλύπτει ελάχιστα τη μακροπρόθεσμη επίδοσή του. Οι Wang, Song και Barabasi (2013) εξέτασαν ένα δείγμα άρθρων φυσικής και διαπίστωσαν ότι η ύπαρξη 50 αναφορών τα πρώτα πέντε χρόνια μετά τη δημοσίευση δεν σχετίζεται με περισσότερες αναφορές μετά από 20 χρόνια. Στην πραγματικότητα, οι εργασίες με τις περισσότερες αναφορές στα 30 χρόνια έτειναν να έχουν σχετικά λίγες αναφορές στην αρχή. Οι Stephan, Veugelers και Wang (2017) εξέτασαν 660.000 ερευνητικά άρθρα στη βάση δεδομένων Web of Science και διαπίστωσαν ότι τα πολύ πρωτότυπα άρθρα είχαν λιγότερες πιθανότητες να έχουν υψηλές αναφορές εντός τριών ετών από τη δημοσίευση, αλλά περισσότερες πιθανότητες να έχουν υψηλές αναφορές τρία ή περισσότερα χρόνια μετά τη δημοσίευση.

2.2.2. Ερευνητές Πρώιμης Καριέρας

Επειδή οι αποφάσεις πρόσληψης και προαγωγής λαμβάνονται αρκετά νωρίς στην καριέρα ενός ερευνητή, ο αριθμός των αναφορών μπορεί να είναι πιο σημαντικός ακριβώς όταν είναι λιγότερο κατατοπιστικός. Οι εργασίες και τα βιβλία που αναφέρονται πολλά χρόνια μετά τη δημοσίευση είναι αναμφισβήτητα πιο σημαντικά από τις εργασίες και τα βιβλία που αναφέρονται μόνο λίγο μετά τη δημοσίευση. Το πρόβλημα της χρονικής υστέρησης θέτει συνεπώς ιδιαίτερες προκλήσεις στην αξιολόγηση νεότερων επιστημόνων.

Η χρήση του Google Scholar για την αξιολόγηση των επιστημόνων που βρίσκονται στο ξεκίνημα της σταδιοδρομίας τους δημιουργεί στρεβλά κίνητρα. Από αριθμητική άποψη, είναι προτιμότερο να δημοσιεύονται σταδιακά εργασίες σε θέματα στα οποία υπάρχει μια μεγάλη, ενεργή υποομάδα μελετητών που παραπέμπουν η μία στην άλλη παρά να ανοίγεται ένα νέο πεδίο έρευνας. Οι βραχυπρόθεσμες, επικεντρωμένες στις αναφορές αξιολογήσεις αποθαρρύνουν την τόλμη και την καινοτομία, ιδίως μεταξύ των επιστημόνων που βρίσκονται στα πρώτα στάδια της σταδιοδρομίας τους. Πολλές σημαντικές επιστημονικές εργασίες αρχικά αψηφούν την εύκολη πλαισίωση και εντάσσονται ελάχιστα στην υπάρχουσα βιβλιογραφία, όπως δείχνει το παράδειγμα της εργασίας του John Nash. Οι αποφάσεις πρόσληψης προσωπικού που βασίζονται μόνο ή σε μεγάλο βαθμό στις αναφορές που συγκεντρώνονται εντός πέντε ετών από τη δημοσίευση επιβραβεύουν την ικανότητα έναντι της διαρκούς σημασίας.

2.2.3. Μέγεθος Κοινότητας

Οι ερευνητές σε μεγαλύτερες ερευνητικές κοινότητες έχουν ένα πλεονέκτημα σε σχέση με τους ερευνητές σε μικρότερα πεδία: υπάρχει μια μεγαλύτερη ομάδα ερευνητών οι οποίοι μπορούν να κάνουν αναφορά σε αυτούς και τις δημοσιεύσεις τους. Μια εργασία που αναφέρεται σε 16 από τα 100 άρθρα που δημοσιεύονται σε ένα δεδομένο έτος για το Κογκρέσο των ΗΠΑ έχει πιθανόν μικρότερη απήχηση από μια εργασία που αναφέρεται σε οκτώ από τα 10 άρθρα που δημοσιεύονται εκείνο το έτος για το Πακιστάν, αλλά έχει διπλάσιο αριθμό. Οι ερευνητές που παράγουν σταδιακές βελτιώσεις σε καλά εδραιωμένα επιστημονικά πεδία τείνουν έτσι να έχουν μεγαλύτερο αριθμό αναφορών από εκείνους τους ερευνητές που παράγουν νέες ιδέες σε μικρά, αλλά σημαντικά ή αναπτυσσόμενα πεδία.

Κεφάλαιο 3. Βιβλιογραφική Ανασκόπηση για την Καταλληλότητα του Google Scholar ως Πηγή Επιστημονικών Πληροφοριών σε Σχέση με Άλλες Βάσεις

Σε αντίθεση με άλλες βάσεις δεδομένων, το πραγματικό μέγεθος του Google Scholar δεν αποκαλύφθηκε ποτέ από την Google και δεν υπάρχει επίσημη τεκμηρίωση για τον αριθμό των εγγραφών και των πηγών που περιλαμβάνει ή για τη συχνότητα και τις κατευθυντήριες γραμμές των ενημερώσεων της. Αρκετές μελέτες προσπάθησαν να αναπτύξουν μια μεθοδολογία που θα επιτρέψει την εκτίμηση του μεγέθους της βάσης δεδομένων του Google Scholar.

Οι προσεγγίσεις για μια τέτοια εκτίμηση κυμαίνονταν από τη συγκριτική αξιολόγηση του Google Scholar με ελεγχόμενες βάσεις δεδομένων, όπως η Microsoft Academic Search (Khabsa & Giles, 2014) ή η Web of Science (Orduna-Malea, et al., 2014). Αυτές οι μελέτες χρησιμοποίησαν τυχαία δείγματα ερωτημάτων και εξέτασαν το επίπεδο επικάλυψης μεταξύ του Google Scholar και των άλλων βάσεων δεδομένων. Βάσει των αποτελεσμάτων τους εκτίμησαν ότι ο αριθμός των αγγλικών επιστημονικών δημοσιεύσεων στο διαδίκτυο ανέρχεται σε περίπου 144 εκατομμύρια, εκ των οποίων το Google Scholar κάλυπτε κατ' εκτίμηση 100 εκατομμύρια. Σε σύγκριση με την Web of Science, διαπιστώθηκε ότι το Google Scholar έχει τριπλασιάσει τον αριθμό των εγγράφων που περιλαμβάνουν μη αγγλικές δημοσιεύσεις. Αυτές οι εκτιμήσεις πρέπει να λαμβάνονται υπόψη ως σημαντικός παράγοντας κατά τη σύγκριση του Google Scholar με άλλες βάσεις δεδομένων.

Λαμβάνοντας υπόψη τις κύριες διαφορές μεταξύ των βάσεων δεδομένων και του Google Scholar, οι συγκρίσεις κάλυψης έχουν αποτελέσει το επίκεντρο πολλών μελετών στον τομέα αυτό. Αρκετές μελέτες σύγκριναν την κάλυψη του Google Scholar με εκείνη του Scopus, του Web of Science, του PubMed και άλλων. Στα πρώτα χρόνια της λειτουργίας του το Google Scholar αποτέλεσε αντικείμενο αρκετών μελετών που το σύγκριναν με βάσεις δεδομένων ανάκτησης πληροφοριών σε διάφορους κλάδους.

Παρά τις δυνατότητες ανίχνευσης και το γεγονός ότι πολλοί εκδότες επέτρεψαν στο Google Scholar να ανιχνεύσει τους ιστότοπους τους κατά τα πρώτα αυτά χρόνια, διαπιστώθηκε ότι το Google Scholar παρέλειπε σχετικές δημοσιεύσεις που βρίσκονταν στους ιστότοπους των εκδοτών και είχε αξιοσημείωτη καθυστέρηση στην ενημέρωση των δημοσιευμένων άρθρων από τις πληρωμένες πλατφόρμες. Αυτό πιθανόν οφείλεται στα εμπόδια που αντιμετώπισε το

Google Scholar κατά τις διαπραγματεύσεις με τους εκδότες. Ένα παράδειγμα είναι η American Chemical Society (ACS) η οποία μήνυσε το Google Scholar για τη χρήση της λέξης "Scholar" στο όνομά του. Η ACS το είδε αυτό ως άμεσο ανταγωνισμό προς το δικό της SciFinder Scholar και κατέθεσε αγωγή νωρίς κατά την έναρξη λειτουργίας του Google Scholar. Ωστόσο, οι πηγές ανοικτής πρόσβασης βρέθηκαν να καλύπτονται επαρκώς στο Google Scholar ακόμη και τα πρώτα χρόνια μετά την έναρξη λειτουργίας του.

Ωστόσο, διαπιστώθηκε ότι το Google Scholar κάλυπτε καλύτερα τις θεματικές πηγές σε σύγκριση με τις καθιερωμένες βάσεις δεδομένων όπως το Web of Science και το Scopus (Walters, 2007). Οι μελέτες αυτές διαπίστωσαν ότι οι διεπιστημονικές θεματικές περιοχές και οι κοινωνικές επιστήμες καλύπτονταν καλύτερα από το Google Scholar σε σχέση με το Web of Science ή το Scopus. Ωστόσο, παρά το πλεονέκτημα της κάλυψης των κοινωνικών επιστημών, όλες οι παραπάνω πρώτες μελέτες διαπίστωσαν ότι οι ενημερώσεις του Google Scholar παρέμεναν προβληματικές με τη βάση δεδομένων να παρουσιάζει ακανόνιστες δυνατότητες ενημέρωσης. Πρέπει να σημειωθεί ότι οι μελέτες αυτές διεξήχθησαν στα πρώτα χρόνια του Google Scholar και εντός 2-4 ετών μετά την ίδρυσή του. Ως εκ τούτου, δεν θα πρέπει να προκαλεί έκπληξη το γεγονός πως οι τακτικές ενημερώσεις ήταν κάπως προβληματικές.

Εκτός από την απλή εξέταση της κάλυψης, οι βιβλιομετρικές μελέτες αποσκοπούσαν επίσης στη μέτρηση των δυνατοτήτων παρακολούθησης των αναφορών του Google Scholar σε σύγκριση με το Web of Science και το Scopus.

Ο κύριος λόγος για το αυξανόμενο ενδιαφέρον για τις δυνατότητες εντοπισμού αναφορών του Google Scholar απορρέει από το γεγονός ότι σε αντίθεση με το Scopus ή το Web of Science δεν έχει καθορισμένη πολιτική κάλυψης. Όπως αναφέρθηκε παραπάνω, οι περισσότερες βάσεις δεδομένων περιορίζονται σε συγκεκριμένο κατάλογο πηγών, θεμάτων, τύπων εγγράφων και άλλων περιορισμών. Τα Web of Science και Scopus περιλαμβάνουν σε ένα ευρετήριο μόνο επιστημονική βιβλιογραφία, η οποία έχει περάσει πρώτα από μια κριτική διαδικασία. Το Google Scholar, ωστόσο, έχει καταχωρημένες σε ευρετήριο όλες τις εκδόσεις, ανεξάρτητα από το αν είναι ή όχι αξιολογημένες από επιστήμονες. Αυτό σημαίνει ότι ακόμη και πηγές όπως ερευνητικές εκθέσεις, παρουσιάσεις, ιστολόγια, διατριβές (διδακτορικές ή μεταπτυχιακές), αναφορές σε περιοδικά καταγράφονται ως αναφορές στο Google Scholar. Αυτό, φυσικά, έχει τεράστιο αντίκτυπο στους συγγραφείς. Όταν μετράται η απήχηση, το Google Scholar θα εμφανίζει πολύ περισσότερες αναφορές από οποιαδήποτε άλλη βάση

δεδομένων που περιορίζει τους τύπους παραγωγής που καλύπτει. Αρκετές μελέτες συνέστησαν, επομένως, ότι το Google Scholar θα χρησιμοποιείται για την παρακολούθηση των αναφορών σε συνδυασμό με ελεγχόμενες βάσεις δεδομένων (Levine-Clark & Gil, 2008). Το Google Scholar παρουσίασε καθυστερήσεις στην καταχώρηση σε ευρετήριο περιοδικών, που έχουν πρωτότερα κριθεί και που πληρώνονται από εκδότες, ωστόσο κατάφερε να παρακολουθήσει περισσότερες αναφορές σε διάφορα θεματικά πεδία που δεν καλύπτονταν καλά από το Web of Science ή το Scopus.

Ένας σημαντικός τομέας στον οποίο διαπιστώθηκε ότι αυτό ισχύει είναι η μηχανική. Μια σύγκριση των Google Scholar και Compendex πραγματοποιήθηκε από τον Cusker (2013). Η μελέτη αυτή χρησιμοποίησε μια μεθοδολογία που χρησιμοποιήθηκε προηγουμένως (Meier & Conkling, 2008) για να εξετάσει κατά πόσον το Google Scholar προχώρησε με τρόπους που θα μπορούσαν να το καταστήσουν το κύριο εργαλείο χρήσης ευρετηρίων για τη βιβλιογραφία της μηχανικής. Επαναλαμβάνοντας τις σειρές των ερωτημάτων που χρησιμοποιήθηκαν στη μελέτη του 2008, επιλέχθηκαν τα 25 κορυφαία άρθρα με τη χρήση της λειτουργίας κατάταξης συνάφειας (Relevance ranking) εντός του Compendex. Τα άρθρα αυτά αναζητήθηκαν στη συνέχεια στο Google Scholar. Σε αντίθεση με την προηγούμενη μελέτη, οι συγγραφείς διαπίστωσαν ότι παρά το γεγονός ότι το Google Scholar δεν διαθέτει τα προηγμένα χαρακτηριστικά που προσφέρουν ορισμένα από τα άλλα λογισμικά υποδοχής, όπως επιλογές κατάταξης, ταξινόμησης και φιλτραρίσματος, είχε συνολικά καλύτερες επιδόσεις έναντι του Compendex σε αυτή τη μελέτη από ότι σε εκείνη την προηγούμενη εργασία. Με βάση το παραπάνω οι συγγραφείς πιστεύουν ότι το Google Scholar μπορεί να χρησιμοποιηθεί ως πρωτογενής μηχανή αναζήτησης βιβλιογραφίας για τη Μηχανική. Ο κύριος λόγος για το συμπέρασμα είναι το κόστος του Compendex σε σύγκριση με το Google Scholar που είναι ένα δωρεάν εργαλείο. Οι ελλείψεις του Google Scholar δεν κρίθηκαν αρκετά σημαντικές σε σύγκριση με το γεγονός ότι είναι δωρεάν και καλύπτει τη βιβλιογραφία. Επομένως, η χρήση του αριθμού των παραπομπών από το Scopus και η σύγκρισή τους με το Google Scholar φαίνεται να αποτυπώνει καλύτερα το σύνολο των αναφορών της επιστημονικής δουλειάς

Αντίστοιχη φαίνεται να είναι η συμπεριφορά και σε διαφορετικά θεματικά πεδία, όπως την ιατρική (Mastrangelo et al., 2010), την ψυχολογία (Garcia-Perez, 2010), τις γεωεπιστήμες (Mikki, 2010), τις επιστήμες της ζωής (Beckmann & Wehrden, 2012), αλλά και την εκπαίδευση (van Aalst, 2010), με τους κύριους λόγους για τη χρήση του Google Scholar, εκτός από τις θεματικές ή ελεγχόμενες βάσεις δεδομένων, να είναι η έλλειψη σαφήνειας

όσον αφορά τον ποιοτικό έλεγχο και το εύρος της δημιουργίας των ευρετηρίων του, που καθιστούν το Google Scholar ακατάλληλο να χρησιμοποιηθεί ως ενιαία πηγή παρακολούθησης των αναφορών.

Ανεξάρτητα από τα θεματικά πεδία που μελετώνται σε αυτές τις εργασίες, η συντριπτική πλειοψηφία των μελετών επισημαίνει διάφορες εγγενείς προκλήσεις για τη χρήση του Google Scholar ως μοναδικής πηγής βιβλιογραφίας, μεταξύ των οποίων είναι: η έλλειψη ποιοτικού ελέγχου της κάλυψης και η γενική έλλειψη σαφήνειας σχετικά με το ποιες πηγές εντάσσονται στο ευρετήριο του. Η ανάκτηση από το Google Scholar μη αξιολογημένων πηγών, τα κενά στην κάλυψη του περιεχομένου των εκδοτών και η αδυναμία του να κατατάξει τα αποτελέσματα με ουσιαστικό τρόπο ήταν μεταξύ των παραγόντων που οδήγησαν στη γενική σύσταση να χρησιμοποιείται το Scholar σε συνδυασμό με θεματικές, ελεγχόμενες βάσεις δεδομένων (Shariff και συνεργάτες, 2013).

Το 2010, τέθηκαν διάφορα ερωτήματα σχετικά με την ικανότητα του Google Scholar να διακρίνει τις "ανεπιθύμητες δημοσιεύσεις" από την πραγματική επιστημονική πληροφόρηση (Beel & Gipp, 2010). Σε αυτές τις μελέτες, ψεύτικα έγγραφα ανέβηκαν στον ιστό και τα δέχτηκε το Google Scholar ως έγκυρα, αποδεικνύοντας ότι παρά το μέγεθός του, το Google Scholar δεν διαθέτει πολιτικές ελέγχου ποιότητας. Το 2012 αναφέρθηκε μια γνωστή μελέτη (Lopez- Cozar, et al., 2012) που περιγράφει τη δημιουργία έξι πλαστών εγγράφων γραμμένων από έναν πλαστό συγγραφέα. Σε αυτή τη μελέτη οι συγγραφείς ανέβασαν τα πλαστά έγγραφα σε έναν ιστότοπο που δημιούργησαν οι ίδιοι και αποδόθηκαν στους ίδιους. Οι συγγραφείς είδαν μια αύξηση 774 αναφορών που αναφέρθηκαν σε 129 έγγραφα, οι οποίες με τη σειρά τους αύξησαν το δείκτη H των συγγραφέων και των περιοδικών. Παρά τις προσπάθειές τους να ειδοποιήσουν το Google Scholar ακόμη και μετά την αφαίρεση αυτών των εγγράφων από τον ιστότοπο, οι αναφορές εξακολουθούσαν να παρακολουθούνται και να διατηρούνται από το Google Scholar. Η μελέτη αυτή κατέδειξε πόσο εύκολα μπορούν να χειραγωγηθούν τα αποτελέσματα του Google Scholar από οποιονδήποτε επιθυμεί να το κάνει λόγω της έλλειψης ποιοτικού ελέγχου ή επιλεκτικότητας στην κάλυψη του.

Τα αποτελέσματα του Google Scholar, που περιείχαν λάθη στα μεταδεδομένα και τις διπλές εγγραφές, παρακολουθήθηκαν επίσης από συγκριτικές μελέτες, οι οποίες διαπίστωσαν ότι το Google Scholar δεν περιείχε μόνο διπλά στοιχεία αλλά και αντίγραφα άρθρων με λάθη στα ονόματα των συγγραφέων, τις ημερομηνίες και τα θέματα (Tobeg, 2011). Χρησιμοποιώντας μέτρα ανάκλησης, ακρίβειας και κατάταξης, οι μελέτες σημείωσαν επίσης ότι σε σύγκριση

με το Web of Science και το Scopus, το Google Scholar δεν διαθέτει επαρκή εργαλεία αναζήτησης που θα μπορούσαν να βοηθήσουν σε πιο εξειδικευμένες αναζητήσεις, με αποτέλεσμα να συνιστάται ότι το Google Scholar θα χρησιμοποιείται για γενικές αναζητήσεις και όχι για εις βάθος βιβλιογραφικές αναζητήσεις.

Η κάλυψη των αποθετηρίων από το Google Scholar εξετάστηκε από τους Arlitsch και O'Brien (2012), οι οποίοι μελέτησαν διάφορα ιδρυματικά αποθετήρια και το περιεχόμενό τους σε σύγκριση με το Google Scholar. Κατά τη στιγμή της μελέτης, τα εν λόγω αποθετήρια δεν καλύπτονταν καλά από το Google Scholar λόγω προβλημάτων μεταδεδομένων και αδυναμίας του να τα ανακτήσει. Πιο πρόσφατες μελέτες δείχνουν σημαντική βελτίωση της κάλυψης του Google Scholar σε σύγκριση με τα πρώτα χρόνια της λειτουργίας του. Οι DeGraff, DeGraff και Romesburg (2013) απέδειξαν ότι η αύξηση του αριθμού των περιοδικών ανοικτής πρόσβασης και των ιδρυματικών αποθετηρίων αυξάνει τον αριθμό των άρθρων που είναι άμεσα διαθέσιμα μέσω του Google Scholar στον τομέα των γεωεπιστημών. Ωστόσο, όταν εξετάζεται η κάλυψη των περιοδικών ανοικτής πρόσβασης (ΑΠ) που βρίσκονται σε ευρετήριο στα αποθετήρια της Λατινικής Αμερικής, το Google Scholar και η Google παρουσίασαν χαμηλά ποσοστά χρήσης τέτοιων ευρετηρίων και, ως εκ τούτου, δεν είναι αντιπροσωπευτικά για το εν λόγω σώμα βιβλιογραφίας (Ordu.na-Malea & Lopez-Cezar, 2015)

Οι Moed, et al. (2016) διαπίστωσαν σημαντικές διαφορές στην κάλυψη των πηγών μεταξύ των θεματικών πεδίων. Για άρθρα σε ένα επιλεγμένο σύνολο άρθρων που περιλαμβάνονται σε περιοδικά, διαπίστωσαν ότι στη χημεία και την βιολογία, τα ποσοστά παραπομπών στο Google Scholar και το Scopus είναι στατιστικά παρόμοια, ενώ στις πολιτικές επιστήμες τα ποσοστά παραπομπών στο Google Scholar είναι σχεδόν 4 φορές μεγαλύτερα από αυτά στο Scopus. Μια σημαντική πτυχή της κάλυψης των πηγών και της παρακολούθησης των αναφορών είναι η ταχύτητα δημιουργίας των ευρετηρίων. Στη μελέτη τους που συνέκρινε το Google Scholar με το Scopus, οι Moed et al. (2016) διαπίστωσαν ότι η μέση καθυστέρηση της δημιουργίας των ευρετηρίων αυτών στο Scopus είναι δύο μήνες σε σύγκριση με το Google Scholar. Πρόσθεσαν ότι η καθυστέρηση αυτή οφείλεται εν μέρει στη μη αναφορά των αναφορών που αναφέρονται σε άρθρα υπό έκδοση στο Scopus.

Οι συγκρίσεις σε επίπεδο τύπου πηγής επικεντρώνονται κυρίως στα περιοδικά. Οι περισσότερες μελέτες στον τομέα αυτό εξετάζουν τον βαθμό στον οποίο το Google Scholar καλύπτει το περιεχόμενο των περιοδικών και τον ρυθμό με τον οποίο τοποθετείται στο

ευρετήριο το νέο περιεχόμενο συγκριτικά με άλλες βάσεις δεδομένων. Η σημασία αυτών των μελετών έγκειται στον εγγενή περιορισμό των μεγάλων βιβλιογραφικών βάσεων δεδομένων. Η Scopus, η Web of Science και άλλες χρησιμοποιούν προσεκτικά κριτήρια επιλογής των πηγών προς εισαγωγή στο ευρετήριο τους. Τα περιοδικά, τα πρακτικά συνεδρίων και τα βιβλία αξιολογούνται πριν από την εισαγωγή αυτή και έτσι δημιουργείται ένας εγγενής περιορισμός στο εύρος και το βάθος της βάσης δεδομένων. Ορισμένες βάσεις δεδομένων έχουν σχεδιαστεί για να καλύπτουν ορισμένα θεματικά πεδία (π.χ. Compendex) και ως εκ τούτου έχουν περιορισμένο πεδίο εφαρμογής. Ωστόσο, ακόμη και μεγαλύτερες βάσεις δεδομένων, όπως οι Scopus και Web of Science, έχει αποδειχθεί ότι έχουν περιορισμένη κάλυψη ορισμένων επιστημονικών κλάδων (Norris & Oppenheim, 2007).

Εκτός από την κάλυψη, οι μελέτες εξέτασαν επίσης τις βαθμολογίες κατάταξης των περιοδικών, όπως αυτές υπολογίζονται με τη χρήση διαφορετικών βάσεων δεδομένων. Δεδομένου ότι η κατάταξη των περιοδικών συνδέεται στενά με τη συνολική αξιολόγηση του έργου των ερευνητών, ο υπολογισμός των ποσοστών αναφορών είναι ζωτικής σημασίας (Delgado et al., 2012). Αυτό καθίσταται ακόμη πιο κρίσιμο σε κλάδους όπως οι Τέχνες και οι Ανθρωπιστικές και οι Κοινωνικές Επιστήμες (Jacobs, 2011), οι οποίοι ενδέχεται να καλύπτονται λιγότερο από γενικές βιβλιογραφικές βάσεις δεδομένων όπως το Scopus και το Web of Science.

Ένα παράδειγμα μπορεί να βρεθεί σε μια μελέτη σχετικά με την κάλυψη των αναφορών των δημοσιεύσεων 360 ερευνητών (Minasny et al., 2013), η οποία συνέκρινε το Google Scholar με το Web of Science και το Scopus. Οι συγγραφείς κατέληξαν στο συμπέρασμα ότι ο αριθμός των δημοσιεύσεων ανά ερευνητή ήταν 2,3 φορές υψηλότερος στο Google Scholar σε σχέση με τις άλλες δύο βάσεις δεδομένων. Ο κύριος ισχυρισμός τους είναι ότι ο δείκτης h-index του Google Scholar είναι ένα καλό μέτρο για τα περιοδικά σε αυτό το πεδίο μελέτης και αποτυπώνει την απόδοση του περιοδικού σε αναφορές καλύτερα από το Impact Factor, επειδή περιλαμβάνει περισσότερες αναφορές από αυτές που καλύπτουν οι άλλες βάσεις δεδομένων. Η παρούσα μελέτη συμφωνεί με τους Harzing και Van Der Wal, (2009) οι οποίοι συνέστησαν τη χρήση του δείκτη Google Scholar h-index για τη μέτρηση των περιοδικών στον τομέα της διοίκησης και των διεθνών επιχειρηματικών περιοδικών.

Οι αναφορές βιβλίων στο Google Scholar μελετήθηκαν από τους (Kousha, Thelwall, & Rezaie, 2011). Η μελέτη αυτή συνέκρινε τις αναφορές σε 1000 βιβλία με τη χρήση Google Scholar και Scopus. Η μελέτη διαπίστωσε ότι οι αναφορές βιβλίων στο Google Scholar ήταν

σχεδόν 3 φορές μεγαλύτερες από εκείνες στο Scopus. Οι συγγραφείς καταλήγουν στο συμπέρασμα ότι στους κλάδους που είναι προσανατολισμένοι στα βιβλία στις κοινωνικές, καλλιτεχνικές και ανθρωπιστικές επιστήμες, οι ηλεκτρονικές αναφορές βιβλίων μπορεί να είναι αρκετά πολλές ώστε να υποστηρίζουν την αξιολόγηση της έρευνας από συναδέλφους με τη χρήση του Google Scholar. Ωστόσο, λαμβάνοντας υπόψη ότι το Scopus δεν κάλυπτε ολοκληρωμένα τα βιβλία εκείνη την εποχή, το συμπέρασμα αυτό είναι κάπως προβληματικό.

Οι τελευταίες μελέτες που επικεντρώνονται σε συγκρίσεις του Google Scholar με το Web Of Science και το Scopus φαίνεται ότι εξακολουθούν να εξετάζουν τα ζητήματα της κάλυψης (Jacobs, 2016) και των αναφορών (Levay et al., 2016) σε διάφορους κλάδους.

Έχοντας ειπωθεί τα παραπάνω, η ακρίβεια και η έλλειψη προηγμένων επιλογών αναζήτησης παραμένουν μια πρόκληση και εμποδίζουν το Google Scholar να γίνει το μοναδικό εργαλείο για επιστημονικές αναζητήσεις. Πρόσφατες μελέτες σχετικά με την παρακολούθηση των αναφορών καταλήγουν στο συμπέρασμα ότι το Google Scholar ανακτά πράγματι περισσότερες αναφορές ανά άρθρο, αλλά προειδοποιεί για την έλλειψη ποιοτικού ελέγχου, τις επικαλύψεις και τις μετρήσεις αναφορών διαφορετικών εκδόσεων, οι οποίες διογκώνουν τις βαθμολογίες και είναι αναξιόπιστες ως μοναδική πηγή αξιολόγησης της έρευνας. Οι Moed και συνεργάτες (2016) διαπίστωσαν μια τεράστια διασπορά στους τίτλους πηγών στο Google Scholar. Στη μελέτη τους που συγκρίνει την κάλυψη των πηγών του Google Scholar με εκείνη του Scopus, οι μοναδικές πηγές στο Google Scholar που δεν καλύπτονται στο Scopus προέρχονται από το Google Books ή/και από μεγάλους εκδοτικούς οίκους βιβλίων, καθώς και από μεγάλα επιστημονικά και ιδρυματικά αποθετήρια, που παρουσιάζουν έναν κατάλογο των σημαντικότερων πηγών του Google Scholar που δεν καλύπτονται από το Scopus.

3.1. Σύγκριση του Google Scholar με Άλλες Ακαδημαϊκές Μηχανές Αναζήτησης

Οι ακαδημαϊκές μηχανές αναζήτησης και οι βιβλιογραφικές βάσεις δεδομένων (ASEBD) αποτελούν πλέον το τυπικό μέρος από όπου μπορεί κανείς να έχει πρόσβαση σε ενημερωμένες επιστημονικές δημοσιεύσεις. Αυτές οι υπηρεσίες καθιστούν ένα διαρκώς αυξανόμενο απόθεμα επιστημονικής γνώσης, προσβάσιμο στους επιστήμονες φιλτράροντας τις πιο σχετικές πληροφορίες.

Στα τέλη της δεκαετίας του 1990, η άνοδος του Διαδικτύου είδε τα ASEBD να αντικαθιστούν όλο και περισσότερο τα παραδοσιακά συστήματα ανάκτησης πληροφοριών εκτός σύνδεσης. Οι υφιστάμενοι πάροχοι δεδομένων και εκδότες όπως οι ProQuest, Ebsco, Thomson Reuters και Elsevier μπήκαν στο διαδικτυακό χώρο για να προσφέρουν υπηρεσίες πληροφόρησης. Ωστόσο, μόνο στις αρχές της δεκαετίας του 2000 οι καινοτομίες στην πρόσβαση σε δεδομένα μεταμόρφωσαν την πρόσβαση σε επιστημονικές πληροφορίες. Μεγάλες μηχανές αναζήτησης που βασίστηκαν σε προγράμματα ανίχνευσης, όπως το Google Scholar, το Microsoft Academic και το Scirus άρχισαν να καθιστούν τεράστιους όγκους επιστημονικών δεδομένων εύκολα προσβάσιμους σε οποιοδήποτε χωρίς κόστος. Το Google Scholar έγινε η νούμερο ένα πηγή πληροφοριών στον ακαδημαϊκό χώρο και χρησιμοποιείται συχνά λόγω της ευκολίας και της εξοικείωσης των χρηστών με το συγκεκριμένο σύστημα αναζήτησης (Georgas, 2014; Jamali & Asadi 2010). Αν και δεν ήταν όλα τα έγγραφα διαθέσιμα σε μορφή πλήρους κειμένου, το Google Scholar θα μπορούσε να δημιουργήσει έναν σημαντικό πόρο δημοσίων διαθέσιμων εγγράφων που καλύπτουν μια μεγάλη σειρά επιστημονικών κλάδων και γλωσσών. Το Google Scholar φαίνεται ασυναγώνιστο στην αποτελεσματική παροχή επιστημονικών εγγράφων στο διαδίκτυο. Ωστόσο, η Microsoft Academic, ξεκίνησε εκ νέου τη μηχανή ακαδημαϊκής αναζήτησής της το 2017 με σκοπό να μπορέσει να ανταγωνιστεί ξανά το Google Scholar (Harzing & Alakangas, 2017). Εκτός από το Google Scholar και το Microsoft Academic, υπάρχουν ωστόσο πολλές άλλες μεγαλύτερες πολυεπιστημονικές μηχανές αναζήτησης, βιβλιογραφικές βάσεις δεδομένων και άλλες υπηρεσίες πληροφοριών που προσπαθούν να πείσουν τους ακαδημαϊκούς χρήστες για την εγκυρότητα της μοναδικής προσφοράς των πληροφοριών τους.

Ενώ οι χρήστες έχουν τη δυνατότητα επιλογής της υπηρεσίας που θα χρησιμοποιήσουν, ωστόσο συχνά δεν είναι σαφές ποιο από τα συστήματα αναζήτησης τους εξυπηρετεί καλύτερα. Υπάρχουν πολλά κριτήρια για την αξιολόγηση της ποιότητας των συστημάτων

αναζήτησης, όπως η συνάφεια, η αντικειμενικότητα ή η ακρίβεια. Στην μελέτη του ο (Gusenbauer, 2018) επικεντρώθηκε σε ένα κριτήριο, το εύρος ενός συστήματος αναζήτησης ως προς το μέγεθός του, που αντικατοπτρίζει τον αριθμό των προσβάσιμων πόρων για έναν συγκεκριμένο χρήστη (Grigas et al. 2017). Τα αποτελέσματα που αποκτά ένας χρήστης με ένα ερώτημα επηρεάζονται, μεταξύ άλλων κριτηρίων ποιότητας, από τα όρια των δεδομένων που είναι διαθέσιμα στη συγκεκριμένη μηχανή αναζήτησης ή στη βιβλιογραφική βάση δεδομένων. Όταν η υπερφόρτωση πληροφοριών υπολογίζεται με συνάφεια, ένα μεγαλύτερο εύρος μπορεί να φέρει καλύτερα αποτελέσματα αναζήτησης συγκριτικά με ένα μικρότερο εύρος.

Εκτός από τους ακαδημαϊκούς χρήστες, άλλες ομάδες που ενδιαφέρονται να γνωρίζουν τα μεγέθη των συστημάτων ακαδημαϊκής αναζήτησης σε μια συγκεκριμένη χρονική στιγμή περιλαμβάνουν: ειδικούς πληροφοριών σε ερευνητικά ιδρύματα ή βιβλιοθήκες, προκειμένου να επιτρέπουν τη σύγκριση αλλά και να γνωρίζουν το μέγεθος των μεμονωμένων συστημάτων αναζήτησης σε πολλαπλά χρονικά σημεία ώστε να καθιστούν δυνατή τη διαχρονική αξιολόγηση της απόδοσης και της σταθερότητας. Επομένως, η γνώση του εύρους ενός δεδομένου συστήματος αναζήτησης δεν έχει σημασία μόνο για τους ακαδημαϊκούς χρήστες, αλλά και για τους ειδικούς της πληροφόρησης.

Ωστόσο, η ανάπτυξη της προσφοράς ASEBD από τη μία βελτίωσε τον τρόπο με τον οποίο οι ερευνητές είχαν πρόσβαση σε πληροφορίες, από την άλλη όμως δημιούργησε μειονεκτήματα στη διαφάνεια του πεδίου εφαρμογής (Halevi et al. 2017). Ιδιαίτερα το εύρος του Google Scholar παραμένει άγνωστο, ειδικά επειδή ο στόχος του Google Scholar είναι να ευρετηριάσει ολόκληρο το σύμπαν των επιστημονικών πληροφοριών, εκτιμώντας ότι το μέγεθός του έχει προσελκύσει πολυάριθμες ακαδημαϊκές εργασίες. Η γνώση του μεγέθους και της ανάπτυξης του Google Scholar μπορεί να είναι ενδεικτική του μεγέθους και της αύξησης των επιστημονικών δεδομένων στο σύνολό τους (Halevi et al. 2017): «ίσως ακόμη και το Google Scholar να μην γνωρίζει αυτόν τον αριθμό» «... ένας αριθμός που αντιπροσωπεύει περίπου την διαδικτυακή επιστημονική κληρονομιά που βρίσκεται σε κυκλοφορία αυτή τη στιγμή» (Orduña-Malea et al., 2014). Οι ερευνητές παραμένουν απογοητευμένοι για το απόρρητο του Google Scholar: «Η μυστικοπάθεια του σχετικά με κάθε πτυχή του Google Scholar είναι εφάμιλλη με αυτή της κυβέρνησης της Βόρειας Κορέας. Η βάση δεδομένων γίνεται ολοένα και μεγαλύτερη αλλά με λάθος τρόπο, μέσα από τη συσσώρευση γιγαντιαίων συλλογών άσχετου ή/και μη επιστημονικού περιεχομένου». Το Google Scholar ενθαρρύνει την επιστημονική έρευνα σχετικά με την κάλυψη του για την

αντιμετώπιση τέτοιων επικρίσεων, όπως φαίνεται στις σελίδες Συχνών Ερωτήσεων: «Όλες αυτές οι ερωτήσεις [σχετικά με την κάλυψη αναζήτησης] απαντώνται καλύτερα με την αναζήτηση ενός στατιστικού δείγματος άρθρων που έχει την ιδιότητα ενδιαφέροντος — περιοδικό, συγγραφέας, κ.λπ. Πολλές συγκρίσεις κάλυψης είναι διαθέσιμες στην περίπτωση αναζήτησης για [allintitle:"google scholar"], αλλά ορισμένες από αυτές είναι πιο έγκυρες στατιστικά από άλλες». Η πρόταση δείχνει ότι το Google Scholar αναγνωρίζει την εγκυρότητα ορισμένων από τις επιστημονικομετρικές μεθόδους που εξετάζει.

Η έρευνα για το μέγεθος του Google Scholar έχει μακρά παράδοση και θεωρείται από ορισμένους ως το «χρυσόμαλλο δέρας» (Orduña-Malea et al., 2014). Πράγματι, ακόμη και μόλις δύο χρόνια μετά την κυκλοφορία του στα τέλη του 2004, οι (Mayr & Walter, 2007) ανέλαβαν την πρόκληση να είναι οι πρώτοι που θα αξιολογήσουν την κάλυψη του. Η μελέτη κατέληξε στο συμπέρασμα ότι η κάλυψη του Google Scholar με λίστες Thomson Scientific Journal, Directory of Open Access Journals και περιοδικά από τη βάση δεδομένων SOLIS έφτανε το 78,5%. Αργότερα, ο (Aguillo, 2012) διαπίστωσε ότι το Google Scholar μπορεί να απαριθμήσει συνολικά περισσότερες από 86 εκατομμύρια εγγραφές. Χρησιμοποιώντας τη μεθοδολογία Query hit count (QHC), οι (Orduña-Malea et al., 2014) κατέληξαν στο συμπέρασμα ότι το μέγεθός του Google Scholar πρέπει να εκτείνεται πέρα από όλες τις προηγούμενες εκτιμήσεις και κατέληξαν στο συμπέρασμα ότι το Google Scholar είναι πιθανό να περιέχει 176 εκατομμύρια έγγραφα, συμπεριλαμβανομένων άρθρων, αναφορών και διπλωμάτων ευρεσιτεχνίας. Ωστόσο, λόγω της αδιαφάνειας της τεχνικής λειτουργικότητας του Google Scholar «όλες οι μέθοδοι [αξιολόγησης της κάλυψής τους] εμφανίζουν μεγάλες ασυνέπειες, περιορισμούς και αβεβαιότητες». Μπροστά σε αυτές τις προκλήσεις, το ερώτημα παραμένει εάν η ίδια η Google είναι απρόθυμη να αναφέρει το μέγεθός της ή αν τελικώς απλώς ίσως δεν είναι σε θέση να το κάνει. Με την εργασία του ο (Gusenbauer, 2018) σκόπευε να ρίξει περισσότερο φως στο πόσο μεγάλο είναι στην πραγματικότητα το Google Scholar και πώς συγκρίνεται με άλλα μεγάλα πολυεπιστημονικά ASEBD.

Ενώ το Google Scholar είναι μια από τις πιο δημοφιλείς μηχανές αναζήτησης ακαδημαϊκών, δεν είναι και η μόνη σχετική για επιστημονικές έρευνες (Orduña Malea, et al., 2014). Με αυξανόμενο αριθμό μηχανών αναζήτησης και βιβλιογραφικών βάσεων δεδομένων, η ανταγωνιστική πίεση αυξάνεται για την παροχή χρήσιμων πληροφοριών. Καθώς ο αριθμός των συστημάτων αναζήτησης αυξάνεται, τόσο οι δυνατότητες και η λειτουργικότητα που προσφέρονται για την πρόσβαση στα αποτελέσματα αναζήτησης διαφοροποιούνται.

Συγκεκριμένα, διαπιστώθηκε ότι είναι εύλογο να θεωρηθεί ότι το μέγεθος του Google Scholar έχει υποτιμηθεί μεταξύ 8% και 55% μέχρι στιγμής εξαιτίας της διαφοράς της μεθόδου, όχι του χρόνου. Είναι σίγουρα η πιο ολοκληρωμένη μηχανή αναζήτησης του ακαδημαϊκού χώρου, ωστόσο, παραμένει ασαφές γιατί το Google Scholar δεν αναφέρει το μέγεθός του. Δεδομένης της ασταθούς φύσης του QHC του Google Scholar, είναι πιθανό η ίδια η Google είτε να αντιμετωπίζει δυσκολίες στην ακριβή εκτίμηση του μεγέθους του είτε να μην θέλει να αναγνωρίσει ότι το μέγεθός του παρουσιάζει σημαντικές διακυμάνσεις. Ίσως είναι σημαντικό για την Google να θέλει να πείσει όσους αναζητούν πληροφορίες ότι προσφέρει μια δομημένη, αξιόπιστη και σταθερή πηγή γνώσης. Εάν η Google διατηρήσει την πολιτική της να μην προσφέρει πληροφορίες, η επιστημονικομετρική εκτίμηση θα πρέπει να παραμείνει η μοναδική πηγή πληροφοριών για το μέγεθός της.

3.2. Επιλογές για την Παρακολούθηση Παραπομπών στο Google Scholar και σε Άλλες Μηχανές Αναζήτησης Ακαδημαϊκών Μελετών

Πολλοί ερευνητές ενδιαφέρονται να βρουν πληροφορίες παραπομπών για ένα δεδομένο άρθρο – τόσο όσον αφορά το πόσες φορές αναφέρεται το άρθρο όσο και το ποιος παραθέτει αυτό το άρθρο. Αυτό μπορεί να αφορά την πληρότητα μιας βιβλιογραφικής αναζήτησης ή ίσως για να εξεταστεί το πόσο συχνά αναφέρονται οι δικές τους δημοσιεύσεις. Ο Eugene Garfield κατέστησε δυνατή την ευρεία χρήση της ανάλυσης αναφορών στον ακαδημαϊκό χώρο μέσω της δημιουργίας τριών δεικτών αναφορών: των Science, Humanities , Social Science Citation, οι οποίοι συνδυάστηκαν και μετατράπηκαν σε μια ηλεκτρονική έκδοση που ονομάζεται Web of Science. Αυτοί οι δείκτες βασίστηκαν στην ιδέα ότι ένα προσεκτικά επιλεγμένο υποσύνολο περιοδικών θα παρήγαγε την πλειονότητα της σημαντικής βιβλιογραφικής αναφοράς για κάθε δεδομένο άρθρο. Για πολλούς λόγους οι καθηγητές και οι ερευνητές μπορεί να θέλουν να αποδείξουν τον αντίκτυπο της εργασίας τους και η ανάλυση των παραπομπών αποτελεί έναν τρόπο (αν και αμφιλεγόμενο (Cheek et al., 2006) για να επιτευχθεί αυτό. Για πολλά χρόνια το Web of Science είχε το μονοπώλιο στην παροχή παρακολούθησης αναφορών. Στα τέλη του 2004 εμφανίστηκαν δύο ανταγωνιστές του– το Google Scholar και το Scopus.

Ο αριθμός των φορών που αναφέρεται ένα άρθρο περιοδικού, ένα κεφάλαιο βιβλίου ή ένας ιστότοπος παίζει επίσης σημαντικό ρόλο στον αλγόριθμο κατάταξης του Google Scholar. Τα αποτελέσματα αναζήτησης εμφανίζονται έτσι ώστε τα άρθρα με περισσότερες αναφορές και

τα οποία είναι εξαιρετικά σχετικά να ανεβαίνουν στην κορυφή της λίστας. Αυτό διαφέρει από την πιο παραδοσιακή προεπιλεγμένη "αντίστροφη χρονολογική" σειρά που χρησιμοποιείται από τις περισσότερες επιστημονικές βάσεις δεδομένων. Το Google Scholar δεν παραθέτει τους τίτλους των δημοσιευμάτων που περιλαμβάνει, ούτε τις ημερομηνίες κάλυψης· αν και έχει δηλωθεί ότι υπάρχουν συμφωνίες με τους περισσότερους μεγάλους εκδότες (εκτός από την Elsevier). Ένας άλλος τομέας διαφοράς για το Google Scholar είναι ότι σε αντίθεση με τις περισσότερες βάσεις δεδομένων επιστημονικής έρευνας, «βλέπει» πέρα από τη βιβλιογραφία των δημοσιευμάτων προκειμένου να καλύψει άλλους τρόπους επιστημονικής επικοινωνίας. Άλλες πηγές που καλύπτονται στο Google Scholar περιλαμβάνουν διακομιστές preprints όπως το arXiv (φυσική) και κυβερνητικές και ακαδημαϊκές ιστοσελίδες. Το Google Scholar όμως δεν διευκρινίζει με ποιο τρόπο ένας ιστότοπος πληροί τις προϋποθέσεις ώστε να μπορεί να συμπεριλαμβάνεται στις αναζητήσεις του..

Την ίδια περίπου εποχή που δημοσιοποιήθηκε το Google Scholar, ο Elsevier παρουσίασε το Scopus, μια υπηρεσία ευρετηρίασης και αφαίρεσης που περιέχει το δικό της εργαλείο παρακολούθησης παραπομπών. Το Scopus ευρετηριάζει μεγαλύτερο αριθμό δημοσιευμάτων από το Web of Science και περιλαμβάνει περισσότερα διεθνή και ανοικτής πρόσβασης δημοσιεύματα. Ωστόσο, η κάλυψη αναφορών χρονολογείται μόνο από το 1996 (οι περιλήψεις, όχι όμως και η κάλυψη αναφορών, είναι διαθέσιμες από το 1966 για ορισμένα περιοδικά.) Το Scopus περιλαμβάνει τη δική του μηχανή αναζήτησης Ιστού, το Scirus. Τα αποτελέσματα του Scirus παρουσιάζονται χωριστά από άλλα αποτελέσματα του Scopus. Επίσης, υλικό από το Scirus δεν περιλαμβάνεται στον αριθμό αναφορών για τα αρχεία του Scopus.

Η ανάλυση παραπομπών αποτέλεσε το επίκεντρο της έρευνας και της συζήτησης για δεκαετίες. Πολλά έχουν γραφτεί για τις τεχνικές ανάλυσης παραπομπών (Kurtz et al., 2005), την εφαρμογή σε διαφορετικούς κλάδους (Holden et al., 2005), τις διαμάχες γύρω από τη χρήση της ανάλυσης παραπομπών, των παραγόντων επιρροής του δημοσιεύματος για τη μέτρηση της αξίας και του αντίκτυπου ενός δεδομένου τίτλου δημοσιεύματος ή το άρθρο ενός δεδομένου συγγραφέα (Cheek et al., 2006). Με την εισαγωγή του Scopus και του Google Scholar, υπήρξαν πολλά πρόσφατα άρθρα που περιλαμβάνουν προσεκτική ανάλυση των χαρακτηριστικών κάθε μεμονωμένου εργαλείου καθώς και συγκρίσεις μεταξύ δύο ή περισσότερων από αυτά τα εργαλεία και άλλων (για παράδειγμα, PubMed και Scirus) (Goodman, 2005). Αν και αυτά τα άρθρα συζητούν τα γενικά χαρακτηριστικά και αναφέρουν τα αποτελέσματα των δειγματοληπτικών αναζητήσεων που έχουν ολοκληρώσει οι

συγγραφείς, δεν εξετάζουν συστηματικά τις λειτουργίες ανάλυσης παραπομπών. Σε μια μελέτη του 2005 που ανέλυσε το Google Scholar, ο (Noruzi, 2005) συνέκρινε εν συντομία τους αριθμούς αναφορών για– το Google Scholar και το Web of Science – στον τομέα της ιστομετρίας «webometrics». Αρχικά, ο συγγραφέας επέλεξε το πρώτο άρθρο για να καθορίσει τη λέξη "webometrics" και παρείχε τους "χρόνους που αναφέρονται" τόσο για το Web of Science όσο και για το Google Scholar. Στη συνέχεια, ο συγγραφέας συνέκρινε τον αριθμό των μοναδικών και επικαλυπτόμενων αναφορών με αυτό το άρθρο . Ο (Noruzi, 2015) εξέτασε επίσης τους αριθμούς αναφορών για τα άρθρα με τις περισσότερες αναφορές στο πεδίο, πραγματοποιώντας μια αναζήτηση για τον όρο "webometrics ή webometric".

Υπάρχουν εγγενή προβλήματα όσον αφορά τη χρήση των αναζητήσεων θεμάτων ως μέτρο σύγκρισης, λόγω των διαφορών στον τρόπο με τον οποίο πραγματοποιούν αναζητήσεις το Web of Science, το Google Scholar και το Scopus. Για παράδειγμα, το Web of Science δεν αναζητά αυτόματα κοινές παραλλαγές λέξεων, ενώ το Scopus και το Google Scholar κάνουν τέτοιου τύπου αναζήτηση. Παρόμοιες αναζητήσεις λέξεων-κλειδιών στο Scopus και στο Web of Science συχνά επιστρέφουν σχετικά μικρά σύνολα αποτελεσμάτων (λιγότερες από εκατό εγγραφές), ενώ η ίδια αναζήτηση στο Google Scholar μπορεί να επιστρέψει εκατοντάδες αποτελέσματα. Για παράδειγμα, μια αναζήτηση για τη φράση «συμπληρωματική ιατρική» με τη λέξη «παχυσαρκία» επιστρέφει 9 αποτελέσματα στο Scopus, 6 στο Web of Science και 596 αποτελέσματα στο Google Scholar.

Η παρακολούθηση παραπομπών γνωστών άρθρων ως μέθοδος σύγκρισης αποφεύγει τις ασυνέπειες στην αναζήτηση του θέματος. Σε μια προκαταρκτική μελέτη οι (Bakkalbasi et al., 2006) εξέτασαν τον αριθμό των αναφορών για αυτά τα τρία εργαλεία σε άρθρα από το Journal of the American Society for Information Science and Technology (JASIST) που δημοσιεύτηκαν το 1985 και το 2000. Διαπίστωσαν ότι το παλαιότερο υλικό φαίνεται να καλύπτεται καλύτερα από το Web of Science, αν και αυτό δεν επιβεβαιώθηκε στατιστικά λόγω του μικρού μεγέθους του συνόλου δεδομένων. Για το πιο σύγχρονο υλικό, ο αριθμός αναφορών στο Google Scholar ήταν υψηλότερος από το Web of Science ή το Scopus, ενώ δεν υπήρχε στατιστική διαφορά μεταξύ των μετρήσεων αναφορών που αναφέρθηκαν από το Web of Science και το Scopus. Οι συγγραφείς συνέστησαν την υλοποίηση μιας μεγαλύτερης και πιο ισχυρής μελέτης.

Προσπαθώντας να παρέχει μια πιο ισχυρή μελέτη, η εργασία των (Bakkalbasi et al., 2006) εξέτασε ένα γνωστό σύνολο άρθρων και τον αριθμό των άρθρων παραπομπής καθώς και

άλλου υλικού που παρέχεται από καθένα από τα τρία εργαλεία αναζήτησης για αυτό το διακριτό σύνολο, αφαιρώντας έτσι την ασάφεια που είναι εγγενής στις αναζητήσεις θέματος. Με αυτόν τον τρόπο η μελέτη παράγει επαρκή δεδομένα ώστε να ελεγχθεί η υπόθεση που λέει ότι η διαφορετική κάλυψη επιστημονικών δημοσιεύσεων που παρέχεται από τα τρία εργαλεία αναζήτησης θα οδηγήσει σε διαφορετικούς αριθμούς αναφορών από το καθένα. Κατά την επιλογή ενός συνόλου άρθρων για εργασία, αποφασίστηκε ότι θα πρέπει επίσης να ληφθούν υπόψη οι διαφορές στις συνήθειες δημοσίευσης για διάφορους κλάδους. Έτσι, επιλέχθηκαν δύο κλάδοι για να διερευνηθεί το αν ακολουθούνται διαφορετικά πρότυπα δημοσίευσης. Ο ένας κλάδος, η φυσική, έχει αγκαλιάσει σε μεγάλο βαθμό τη χρήση διακομιστών preprints για την πρώιμη διάδοση της ερευνητικής βιβλιογραφίας, ενώ ένας δεύτερος κλάδος, η ιατρική, όχι. Τα θέματα περιορίστηκαν στη φυσική της συμπεκνωμένης ύλης και στην ογκολογία. Σύνολα γνωστών άρθρων από κάθε κλάδο επιλέχθηκαν τόσο από το 1993 (πριν από την επανέκδοση των κυρίαρχων επιστημονικών κλάδων) όσο και από το 2003

Η διαδικασία δειγματοληψίας περιλάμβανε δύο βήματα: πρώτον την επιλογή των τίτλων περιοδικών από κάθε κλάδο (ογκολογία και φυσική της συμπεκνωμένης ύλης) και δεύτερον την επιλογή άρθρων από αυτά τα περιοδικά. Στο πρώτο βήμα, ανακτήθηκαν 123 περιοδικά που περιλαμβάνονται στην κατηγορία «Ογκολογία» και 60 περιοδικά που αναφέρονται στην κατηγορία «Φυσική, Συμπεκνωμένη ύλη», χρησιμοποιώντας τη βάση δεδομένων Εκθέσεων παραπομπής σε δημοσιεύματα του 2004 (JCR). Έντεκα περιοδικά από κάθε κατηγορία επιλέχθηκαν χρησιμοποιώντας τεχνική συστηματικής δειγματοληψίας για να διασφαλιστεί ότι το δείγμα περιείχε ομοιόμορφη κατανομή των περιοδικών σε όλα τα επίπεδα των συντελεστών επιρροής. Για την κατάρτιση του δείγματος, όλοι οι τίτλοι κατατάχθηκαν από την υψηλότερη κατάταξη προς τη χαμηλότερη με βάση τους παράγοντες αντίκτυπου. Στη συνέχεια, η επιλογή τίτλου ξεκίνησε με τον πρώτο τίτλο και επεκτάθηκε για να συμπεριλάβει κάθε νιοστό επόμενο τίτλο όπου το n , το διάστημα δειγματοληψίας, υπολογίστηκε ως:

$$n = \text{Μέγεθος πληθυσμού} / \text{Μέγεθος δείγματος}$$

Άρθρα από τα έτη 1993 και 2003 επιλέχθηκαν ως ο πληθυσμός από τον οποίο θα αντληθεί το δείγμα. Το δεύτερο βήμα της διαδικασίας δειγματοληψίας ξεκίνησε με την ανάκτηση όλων των άρθρων που δημοσιεύτηκαν στους επιλεγμένους έντεκα τίτλους και για τα δύο χρόνια χρησιμοποιώντας το Web of Science, το INSPEC και το PubMed.

3.3. Μελέτες Σχετικές με τον Αλγόριθμο Κατάταξης του Google Scholar

Με την όλο και αυξανόμενη χρήση των ακαδημαϊκών μηχανών αναζήτησης, καθίσταται σημαντικό για τους ακαδημαϊκούς συγγραφείς να είναι τα άρθρα τους σε καλή κατάταξη στις συγκεκριμένες μηχανές αναζήτησης προκειμένου να προσεγγίσουν το κοινό τους. Για τη βελτιστοποίηση των εργασιών για ακαδημαϊκές μηχανές αναζήτησης, όπως το Google Scholar ή το Scienstein.org, είναι απαραίτητη η γνώση σχετικά με τους εφαρμοσμένους αλγόριθμους κατάταξης. Για παράδειγμα, αν μια μηχανή αναζήτησης εξετάσει πόσο συχνά εμφανίζονται οι όροι αναζήτησης στο πλήρες κείμενο ενός άρθρου, οι συγγραφείς θα πρέπει να χρησιμοποιούν τις πιο σχετικές λέξεις-κλειδιά στα άρθρα τους-όποτε είναι δυνατόν-για να φέρουν τις εργασίες τους σε κορυφαία θέση.

Για τους χρήστες ακαδημαϊκών μηχανών αναζήτησης, η γνώση σχετικά με τους εφαρμοσμένους αλγόριθμους κατάταξης είναι επίσης απαραίτητη προκειμένου να αξιολογηθεί η ευρωστία των αποτελεσμάτων. Όπως επισημάνθηκε, οι ερευνητές ενδιαφέρονται να εμφανίζονται τα άρθρα τους σε κορυφαίες θέσεις από τις μηχανές αναζήτησης. Αντίστοιχα, οι χρήστες θα πρέπει να γνωρίζουν τους αλγόριθμους προκειμένου να εκτιμήσουν την ευρωστία και συνεπώς την αξιοπιστία των αποτελεσμάτων των ακαδημαϊκών μηχανών αναζήτησης.

Η γνώση των αλγορίθμων κατάταξης επιτρέπει επίσης στους ερευνητές να εκτιμήσουν τη χρησιμότητα των αποτελεσμάτων σε σχέση με την πρόθεσή τους για αναζήτηση. Για παράδειγμα, οι ερευνητές που σκοπεύουν να αναζητήσουν τις τελευταίες τάσεις στον τομέα τους θα πρέπει να χρησιμοποιήσουν μια μηχανή αναζήτησης που δίνει μεγάλη βαρύτητα στην ημερομηνία των δημοσιεύσεων. Οι χρήστες που αναζητούν τυπική βιβλιογραφία θα πρέπει να επιλέξουν μια μηχανή αναζήτησης που δίνει μεγάλη βαρύτητα στον αριθμό των αναφορών. Αντίθετα, εάν ένας χρήστης αναζητήσει άρθρα από συγγραφείς που προωθούν μια άποψη διαφορετική από την πλειονότητα, οι μηχανές αναζήτησης που δίνουν μεγάλη βαρύτητα στις μετρήσεις αναφορών ενδέχεται να μην δίνουν τα καλύτερα αποτελέσματα.

Λόγω των διαφορετικών αναγκών των χρηστών, πολλές ακαδημαϊκές βάσεις δεδομένων και μηχανές αναζήτησης επιτρέπουν στον χρήστη να επιλέξει έναν αλγόριθμο κατάταξης. Για παράδειγμα, το ScienceDirect επιτρέπει στους χρήστες να επιλέξουν μεταξύ ημερομηνίας και συνάφειας, το IEEE Xplore προσφέρει επιπλέον, κατάταξη ανά τίτλο και η Ψηφιακή Βιβλιοθήκη ACM επιτρέπει στους χρήστες να επιλέξουν εάν θα ταξινομήσουν τα αποτελέσματα κατά συνάφεια, ημερομηνία, αλφαβητικά κατά τίτλο ή περιοδικό, κατά

πλήθος αναφορών ή λήψεις. Ωστόσο, αυτοί οι «αλγόριθμοι» μπορούν να θεωρηθούν μικρής αξίας, καθώς οι χρήστες μπορούν να επιλέξουν μόνο ένα κριτήριο κατάταξης και όχι συνδυασμό αυτών.

Το Google Scholar αποτελεί μια από τις λίγες ακαδημαϊκές μηχανές αναζήτησης που συνδυάζει πολλές προσεγγίσεις σε έναν μόνο αλγόριθμο. Υπάρχουν αρκετές μελέτες σχετικά με το Google Scholar οι οποίες όμως περιλαμβάνουν: επικάλυψη δεδομένων με άλλες ακαδημαϊκές μηχανές αναζήτησης όπως το Scopus και το Web of Science (Bailey et al., 2007), την κάλυψη της βιβλιογραφίας από το Google Scholar γενικά αλλά και σε συγκεκριμένα ερευνητικά πεδία (Meier & Conkling, 2008), την καταλληλότητα χρήσης των παραπομπών του Google Scholar για τον υπολογισμό δεικτών (όπως ο h (Bar-Ilan, 2008)) και την αξιοπιστία του Google Scholar ως σοβαρής πηγής πληροφοριών γενικά (Jacsó, 2005). Ωστόσο, παρόλο που ο αλγόριθμος κατάταξης του Google Scholar επηρεάζει σημαντικά τα ακαδημαϊκά άρθρα που διαβάζονται από την επιστημονική κοινότητα, δεν κατέστη δυνατό να βρεθούν άρθρα σχετικά με τον αλγόριθμο κατάταξης του Google Scholar. Οι μόνες ασαφείς πληροφορίες σχετικά με τον αλγόριθμο προέρχονται από την ίδια την Google: Το Google Scholar ταξινομεί τα «άρθρα με τον τρόπο που κάνουν οι ερευνητές, σταθμίζοντας το πλήρες κείμενο κάθε άρθρου, τον συγγραφέα, τη δημοσίευση στην οποία εμφανίζεται το άρθρο και πόσο συχνά αναφέρεται το άρθρο σε άλλη επιστημονική βιβλιογραφία»³ Οποιοσδήποτε άλλες λεπτομέρειες ή περαιτέρω εξηγήσεις δεν είναι διαθέσιμες.

Ο στόχος της έρευνας των (Beell & Gipp, 2009) ήταν να παρουσιαστεί μια πρώτη εντύπωση για το εάν άλλοι παράγοντες εκτός από τους τέσσερις που προαναφέρθηκαν (πλήρες κείμενο, συγγραφέας, δημοσίευση και παραπομπές) λαμβάνονται υπόψη από το Google Scholar και αν ναι, πόσο βάρος δίνεται σε κάθε παράγοντα. Σκοπός δεν ήταν να δοθεί μία τελεσίδικη απάντηση σε όλα τα ερωτήματα διεξοδικά, αλλά να φανεί ποια κατεύθυνση θα έπρεπε να ακολουθήσει η περαιτέρω έρευνα. Ερευνήθηκαν οι ακόλουθοι παράγοντες:

- (1) Αριθμός αναφορών του άρθρου
- (2) Ηλικία του άρθρου
- (3) Εμφάνιση του όρου αναζήτησης στο πλήρες κείμενο ενός άρθρου
- (4) Συχνότητα του όρου αναζήτησης στο πλήρες κείμενο ενός άρθρου
- (5) Εμφάνιση του όρου αναζήτησης στον τίτλο ενός άρθρου
- (6) Εμφάνιση του όρου αναζήτησης στο όνομα του συγγραφέα ή της δημοσίευσης

Με άλλα λόγια, πραγματοποιήθηκε σύγκριση για το εάν παλιά/πρόσφατα άρθρα με υψηλό/χαμηλό αριθμό αναφορών και/ή όρο αναζήτησης ο οποίος εμφανίζεται (συχνά) στον τίτλο, το πλήρες κείμενο ή/και το όνομα συγγραφέα/δημοσίευσης που εμφανίζεται στο ερώτημα αναζήτησης είναι πιο πιθανό να εμφανίζονται σε κορυφαία θέση από το Google Scholar.

Δεδομένου ότι το Google Scholar προσφέρει δύο βασικούς τρόπους αναζήτησης, μια αναζήτηση στο πλήρες κείμενο και μια αναζήτηση στον τίτλο, αναλύθηκε επίσης (7) το εάν χρησιμοποιείται ο ίδιος αλγόριθμος κατάταξης και για τους δύο τρόπους αναζήτησης. Ο επόμενος στόχος (8) ήταν να αναλυθεί το πώς διαφέρουν οι ταξινομήσεις των άρθρων που ανακτώνται μέσω των λειτουργιών «αναφέρονται από» και «σχετικά άρθρα» από αυτές που ανακτώνται μέσω της κανονικής αναζήτησης λέξεων-κλειδιών. Ο ερευνητικός στόχος (9) ήταν να αναλυθεί εάν τελικά το Google Scholar ευρετηριάζει κείμενο από σχήματα και πίνακες που είναι ενσωματωμένοι ως εικόνες στα άρθρα.

Το Google Scholar εμφανίζει για κάθε άρθρο τον αριθμό των αναφορών του στη λίστα αποτελεσμάτων. Για να αποκτηθεί ένα πλήθος αναφορών, αναπτύχθηκε ένα πρόγραμμα για την ανάλυση του ιστότοπου του Google Scholar. Αυτό το πρόγραμμα στέλνει ερωτήματα αναζήτησης στο Google Scholar και αποθηκεύει τις μετρήσεις και τις θέσεις των αναφορών όλων των επιστρεφόμενων αποτελεσμάτων σε ένα αρχείο .csv. Λόγω των περιορισμών του Google Scholar, ήταν δυνατή η ανάκτηση μόνο 1.000 αποτελεσμάτων ανά ερώτημα αναζήτησης. Η ανάλυση πραγματοποιήθηκε δύο φορές, κάθε φορά με 10 ερωτήματα αναζήτησης. Στην πρώτη εκτέλεση, έγινε αναζήτηση του πλήρους κειμένου των άρθρων. Στη δεύτερη εκτέλεση έγινε αναζήτηση μόνο του τίτλου του άρθρου. Αυτό είχε ως αποτέλεσμα 20 ερωτήματα αναζήτησης, να επιστρέψουν συνολικά 19.612 άρθρα. Οι μετρήσεις των αναφορών και οι ταξινομήσεις των άρθρων αποθηκεύτηκαν και αναλύθηκαν. Για την επαλήθευση της σωστής εκτέλεσης του αναλυτή του Google Scholar, πραγματοποιήθηκαν επιτόπιοι έλεγχοι.

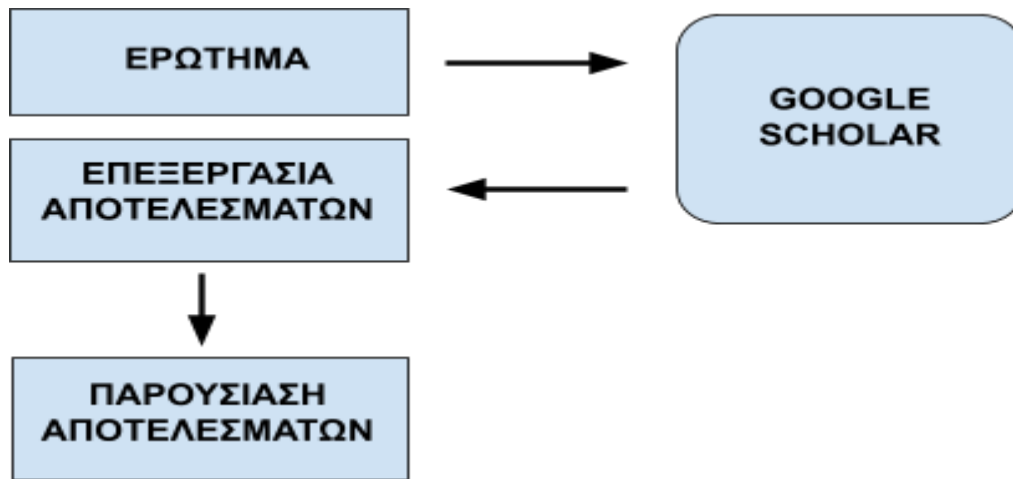
Κεφάλαιο 4. Υλοποίηση και άντληση δεδομένων από την βάση του Google Scholar

Στο πλαίσιο της παρούσας διπλωματικής εργασίας αναπτύχθηκε εφαρμογή η οποία αντλεί δεδομένα από τη βάση του Google Scholar ώστε να μπορεί να εξάγει τα δεδομένα αυτά σε αρχείο txt ή xls με στόχο να είναι άμεσα επεξεργάσιμα ώστε να είναι δυνατή η επεξεργασία του ερευνητικού αντίκτυπου μεμονωμένου ερευνητή ή ομάδας ερευνητών/συγγραφέων. Τόσο τα απαραίτητα εργαλεία για την ανάπτυξη όσο και ο κώδικας της ίδιας της εφαρμογής αλλά και ενδεικτικά αποτελέσματα παρουσιάζονται σε αυτό το κεφάλαιο. Στο παρακάτω διάγραμμα παρουσιάζεται η ροή της πληροφορίας από τη διαμόρφωση του ερωτήματος μέχρι την ολοκλήρωση της απάντησης από το Google Scholar.



Τα εργαλεία που απαιτούνται για την δημιουργία και εκτέλεση της εφαρμογής είναι τα:

- Εφαρμογή ANACONDA
- Γλώσσα προγραμματισμού PYTHON
- Υπηρεσίας SERPAPI
- Google API
- Framework και χρήση Web Server μέσω του flask
- JSON
- HTML,CSS



Σχήμα 1. Η διαδικασία που ακολουθείται κατά την διατύπωση ενός ερωτήματος στο Google Scholar με στόχο την επεξεργασία και την παρουσίαση του τελικού αποτελέσματος.

Με βάση τα παραπάνω πρέπει να επισημανθεί ότι τόσο το ερώτημα όσο και η απάντηση του διαμορφώνονται σε τοπικό επίπεδο ενώ το Google Scholar χρησιμοποιείται μόνο για να επιστρέψει την απάντηση στο ερώτημα που τίθεται. Τόσο το ερώτημα όσο και η επεξεργασία, διαμόρφωση και οπτικοποίηση των δεδομένων πραγματοποιείται σε τοπικό επίπεδο όπως παρουσιάζεται και στο διάγραμμα που ακολουθεί.

Η διαδικασία που ακολουθείται για την εγκατάσταση της εφαρμογής είναι η εξής:

- Για την λήψη των δεδομένων από το Google Scholar χρησιμοποιείται η υπηρεσία της Serpapi ώστε να γίνει η λήψη ενός κλειδιού API που θα χρησιμοποιείται για την λήψη δεδομένων.
- Στην συνέχεια αναπτύχθηκε online εφαρμογή με framework flask για την δημιουργία σελίδων στην Python.

Εκτέλεση εφαρμογής

1. Δημιουργία φακέλου `...` και δημιουργία ενός αρχείου `app.py`. Σε αυτό το αρχείο θα οριστούν και τα `routes` για το ποια συνάρτηση θα εκτελεστεί για κάθε URL.
2. Το αρχείο `app.py` εκτελείται μέσω του `anaconda prompt` με δικαιώματα διαχειριστή δίνοντας την εντολή `flask run`.
3. Εκτός από το αρχείο `app.py` υπάρχει ο φάκελος `static` για τα στατιστικά αρχεία όπως `css` και ο φάκελος `template` που βρίσκεται αποθηκευμένος ο `html` κώδικας.
4. Στον φάκελο `template` θα δημιουργηθούν κατά την εκτέλεση 3 διαφορετικά αρχεία `html` (`home`, `base`, `result`).
5. Στο αρχείο `home` θα δημιουργηθεί η φόρμα στην οποία ο χρήστης θα δίνει το ερώτημα προς αναζήτηση (προαιρετικά θα έχει επιλογή για χρονολογία).
6. Οι υπόλοιπες σελίδες επεκτείνουν το αρχείο `base` ορίζοντας το `block content`.
7. Το αρχείο `result` θα καλεί το API μόλις πατηθεί η επιλογή `submit` όπου θα λαμβάνει τα δεδομένα και θα δίνει το αποτέλεσμα.

4.1. Χρήση API

Για τη λήψη δεδομένων από το Google Scholar γίνεται χρήση της υπηρεσίας serpari.com. Πριν την εκτέλεση πραγματοποιείται η εγγραφή στην ιστοσελίδα Google Scholar ώστε να δημιουργηθεί ένα API key, το οποίο θα χρησιμοποιηθεί όταν χρειαστεί να γίνει λήψη δεδομένων.

Αρχικά δημιουργήθηκε μια φόρμα σε HTML και με Javascript ώστε να γίνεται η χρήση του API ώστε να είναι δυνατόν να παρουσιάζονται τα αποτελέσματα. Σημειώνεται ότι οι browsers για λόγους ασφαλείας δεν επιτρέπουν να δείχνουν «ξένο» περιεχόμενο (πρόβλημα σχετικά με το Cross-Origin Resource Sharing - CORS). Αυτή η ρύθμιση φαίνεται ότι είναι server-side, οπότε πιθανόν το serpari να μην επιτρέπει τη δημιουργία σελίδων με το δικό του περιεχόμενο. Για να αντιμετωπιστεί το πρόβλημα αυτό χρησιμοποιήθηκε η γλώσσα Python για την ανάπτυξη σχετικού λογισμικού.

Ο κώδικας (δες Παράρτημα 1) δέχεται ένα ερώτημα από το χρήστη (ό,τι θα πληκτρολογήσει στο πεδίο search του Google Scholar), καλεί το API και εμφανίζει τα αποτελέσματα. Ο κώδικας έχει εκτενή σχόλια και κατάλληλα ονόματα μεταβλητών. Σημειώνεται ότι επειδή το κείμενο προς αναζήτηση μπορεί να περιέχει κενά ή άλλα σημεία στίξης, πρέπει να κωδικοποιηθεί σε μορφή αποδεκτή για URL. Στη συνέχεια καλείται το API δίνοντας το κείμενο και το κλειδί. Τα αποτελέσματα λαμβάνονται σε μορφή JSON. Οπότε, γίνεται η λήψη των αποτελεσμάτων και εμφανίζονται τα πιο βασικά πεδία της απάντησης (εικόνα 1). {Παραπομπή στο Παράρτημα 1}

```
Result 1:
Title: Machine learning
Publication info: ZH Zhou - 2021 - books.google.com
Link: https://books.google.com/books?hl=en&lr=&id=ctM-EAAAQBAJ&oi=fnd&pg=PR6&dq=machine
+learning&ots=oZPi_7XrYu&sig=vaxdh6tssaa-A1qc9pcZgY7qaCs
Cited by: 1198

Result 2:
Title: Machine learning
Publication info: TM Mitchell, TM Mitchell - 1997 - profs.info.uaic.ro
Link: https://profs.info.uaic.ro/~ciortuz/SLIDES/2017s/ml0.pdf
Cited by: 34730

Result 3:
Title: Machine learning
Publication info: TG Dietterich - Annual review of computer science, 1990 -
annualreviews.org
Link: https://www.annualreviews.org/doi/pdf/10.1146/annurev.cs.04.060190.001351
Cited by: 211
```

Εικόνα 3. Αποτελέσματα κλήσης για το ερώτημα «Machine Learning».

4.2. Ανάπτυξη ιστοσελίδας με Flask

Στη συνέχεια, αναπτύχθηκε εφαρμογή WEB σε Flask που αποτελεί το framework για την δημιουργία σελίδων σε Python, ώστε να υπάρχει μια γραφική διεπαφή με το χρήστη.

4.2.1. Προαπαιτούμενα

Γίνεται η εγκατάσταση της γλώσσας προγραμματισμού Python 3.9 μέσω του Anaconda. Στη συνέχεια, από το anaconda prompt εγκαθίσταται το flask σύμφωνα με τα παρακάτω:

```
pip install flask  
pip install Flask-excel  
pip install Flask-Session
```

Στη συνέχεια στο φάκελο που θα εγκατασταθεί η εφαρμογή δημιουργείται ένα αρχείο app.py με περιεχόμενο που θα γίνει αργότερα η αναφορά του. Σε αυτό το αρχείο γίνεται ο καθορισμός των συναρτήσεων που θα εκτελούνται για κάθε URL. Αυτό το αρχείο εκτελείται με χρήση anaconda prompt με δικαιώματα διαχειριστή, δίνοντας την εντολή:

```
flask --app app run
```

Επειδή έχει χρησιμοποιηθεί το προκαθορισμένο όνομα app.py δεν είναι απαραίτητο να δοθεί το "--app app". Ο server εκτελείται και είναι προσβάσιμος στην τοπική διεύθυνση (local address) του υπολογιστή :

```
http://127.0.0.1:5000
```

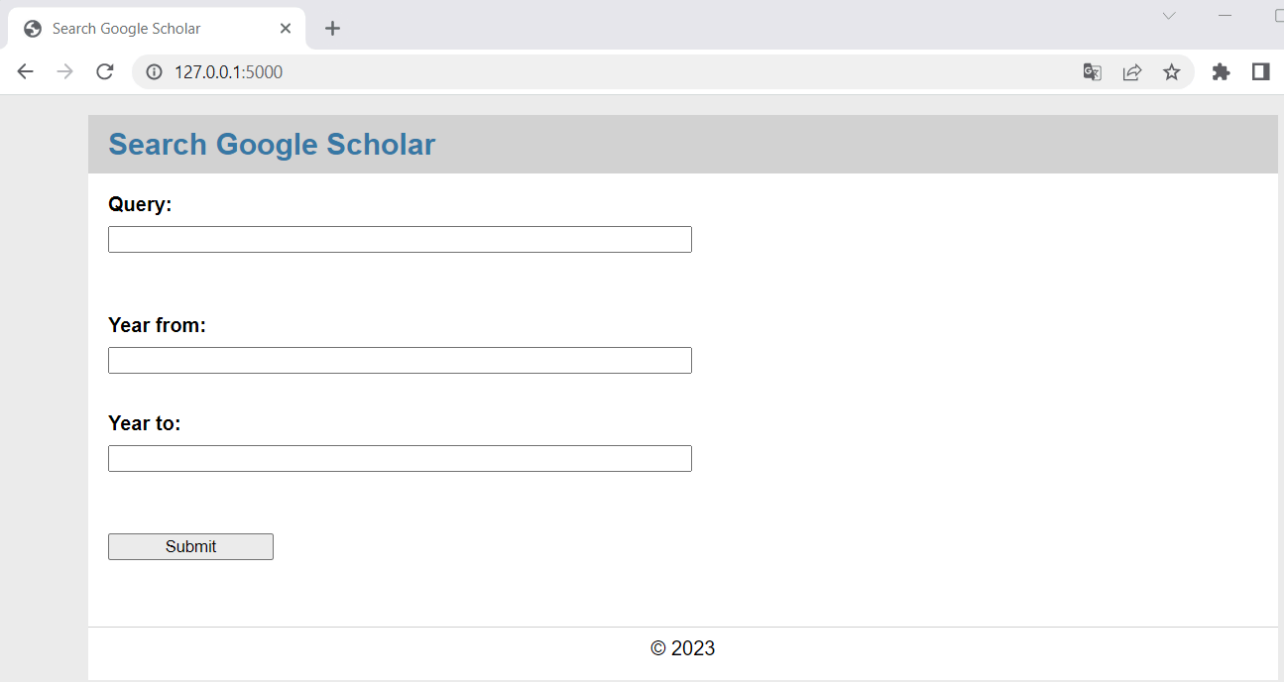
4.2.2. Εφαρμογή flask

Εκτός από το αρχείο `py`, υπάρχει ένας φάκελος `static` για τα στατικά αρχεία, όπως `css`. Επίσης, υπάρχει ένας φάκελος `templates` με τον HTML κώδικα για κάθε σελίδα.

Για την αρχική σελίδα (`/`) ορίζονται στο αρχείο `.py` όσα παρουσιάζονται στο Παράρτημα 2.

Για τη διεπαφή, αρχικά δημιουργείται στον φάκελο `templates` ένα αρχείο `base.html`, το οποίο θα έχει τα γενικά στοιχεία, όπως την επικεφαλίδα. {δες Παραπομπή στο Παράρτημα 3}.

Στην αρχική σελίδα (αρχείο `templates/home.html`) δημιουργείται η φόρμα που ο χρήστης δίνει το ερώτημα προς αναζήτηση και προαιρετικά επιλέγει έτη (από-μέχρι). Το αποτέλεσμα φαίνεται στην εικόνα 4. Επειδή τα έτη είναι ακέραιοι αριθμοί, χρησιμοποιείται το `input` με `type="number"` ώστε να περιοριστεί η είσοδο σε αποδεκτές τιμές. Σημειώνεται ότι τα έτη είναι προαιρετικά, οπότε δεν λαμβάνεται υπόψη η κενή επιλογή. Επίσης, επιτρέπεται να οριστεί μόνο η μια από τις δύο επιλογές.



The image shows a web browser window displaying the 'Search Google Scholar' form. The browser's address bar shows the URL '127.0.0.1:5000'. The form itself is titled 'Search Google Scholar' and contains three input fields: 'Query', 'Year from', and 'Year to'. Below these fields is a 'Submit' button. The footer of the page shows '© 2023'.

Εικόνα 4. Η αρχική σελίδα της φόρμας που ο χρήστης δίνει το ερώτημα προς αναζήτηση και προαιρετικά επιλέγει έτη (από-μέχρι).

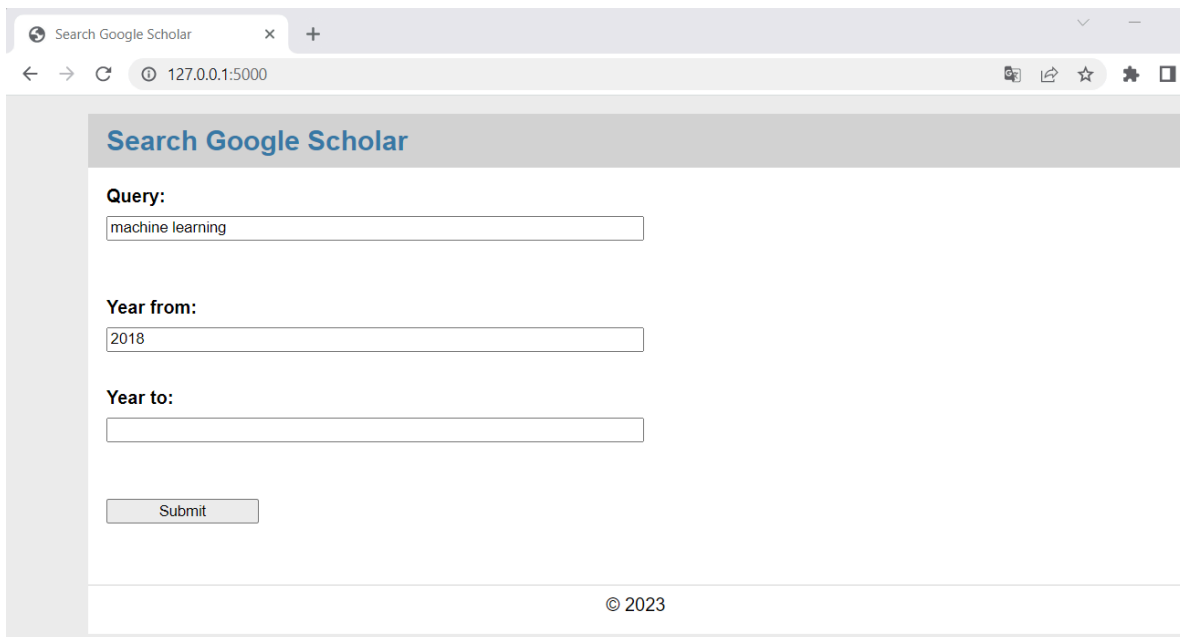
Οι υπόλοιπες σελίδες επεκτείνουν το αρχείο `base.html` ορίζοντας το `block content`. Πιο συγκεκριμένα, ο κώδικας του αρχείου `home.html` βρίσκεται στο Παράρτημα 4.

Τέλος, με την επιλογή του `submit`, τα δεδομένα της φόρμας στέλνονται στη σελίδα `results`, η οποία καλεί το API, λαμβάνει τα δεδομένα και εμφανίζει το αποτέλεσμα.

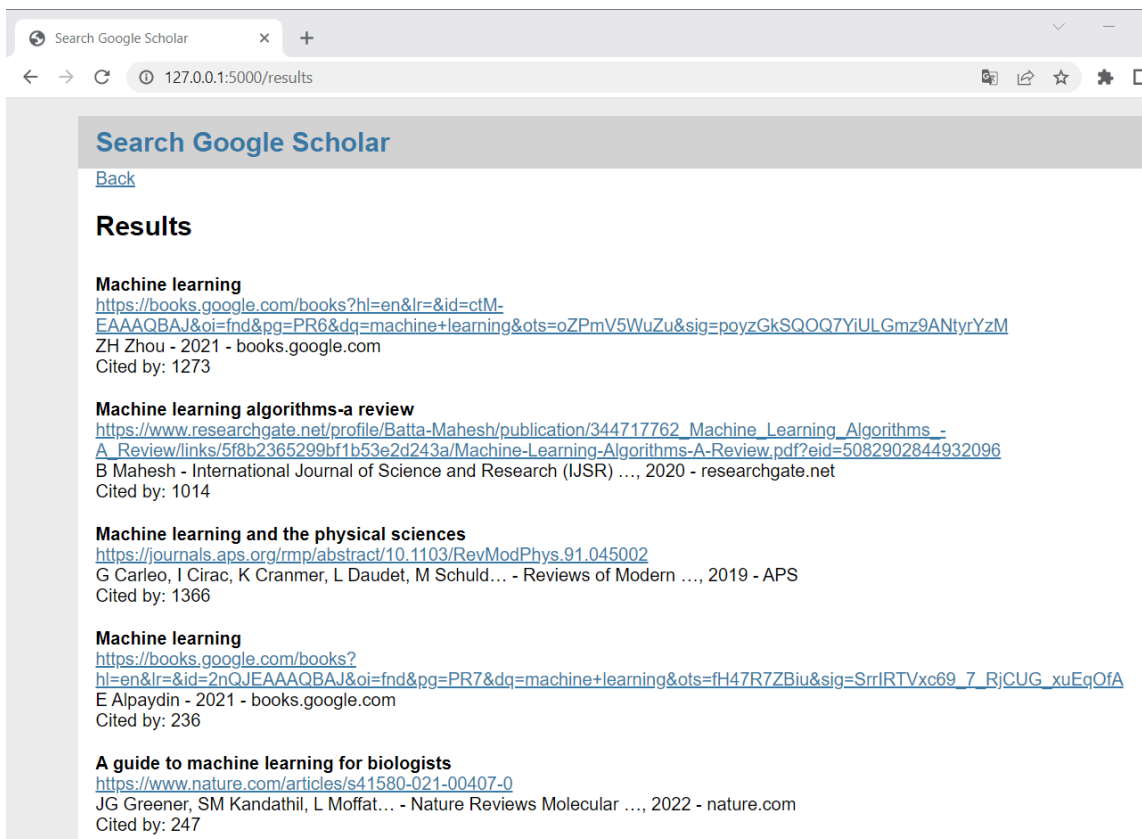
Στο επόμενο βήμα περιγράφεται ο κώδικας της συνάρτησης `results` στο αρχείο `app.py` και δημιουργείται το αντίστοιχο αρχείο `view` στο φάκελο `templates`. Για τη συλλογή των αποτελεσμάτων ορίζεται η συνάρτηση `results`. Αρχικά, λαμβάνονται τα δεδομένα από τη φόρμα της αρχικής σελίδας. Στη συνέχεια σχηματίζεται το URL, συμπεριλαμβάνοντας το ερώτημα και τα έτη, εφόσον τα έδωσε ο χρήστης. Καλείται το API και λαμβάνεται το αποτέλεσμα. Το αποτέλεσμα τοποθετείται στη μεταβλητή `results` τύπου λεξικού, ώστε να την προωθήσει στο αρχείο `view`. Ο κώδικας εκτελείται όπως παρουσιάστηκε παραπάνω μέσω της γραμμής εντολών {δες παραπομπή στο Παράρτημα 5}.

Ο κώδικας της συνάρτησης `view` για τα αποτελέσματα (`results.html`) παρουσιάζεται στο Παράρτημα 6 παρακάτω. Μπορεί να γίνει εκτέλεση του κώδικα Python εντός μπλοκ `{% ... %}` και να εμφανιστούν μεταβλητές εντός μπλοκ `{{ ... }}`.

Ένα παράδειγμα εκτέλεσης φαίνεται στην Εικόνα 5, όπου δώσαμε τον όρο `Machine Learning` και έτος το 2018. Τα αποτελέσματα φαίνονται στην Εικόνα 6.

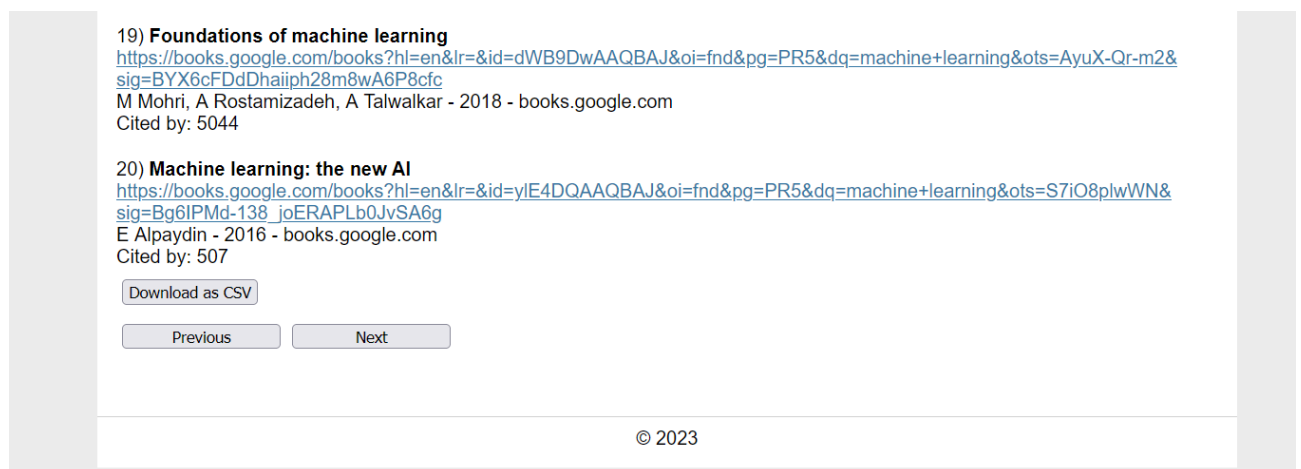


Εικόνα 5. Παράδειγμα χρήσης της φόρμας.



Εικόνα 6. Το αποτέλεσμα μετά την εισαγωγή για αναζήτηση του λεκτικού “Machine Learning” για έτη μετά το 2018.

Σημειώνεται ότι το API επιστρέφει 10 αποτελέσματα και μετά χρειάζεται κλήση εκ νέου για να συνεχίσει με τα υπόλοιπα αποτελέσματα. Οπότε, προστέθηκαν επιλογές Previous και Next για μετάβαση σε προηγούμενη/επόμενη σελίδα αποτελεσμάτων (Εικόνα 6). Όπως φαίνεται στον κώδικα παραπάνω, η κάθε επιλογή καλεί την results μία φορά, αλλάζοντας την παράμετρο start, ώστε να δείχνει τα κατάλληλα αποτελέσματα. Επίσης, υπάρχει επιλογή για λήψη των δεδομένων ως CSV ώστε να είναι δυνατή η απευθείας επεξεργασία τους σε excel. Η επιλογή αυτή καλεί την downloads, η οποία μετατρέπει τα δεδομένα (τα οποία έχουν αποθηκευτεί στο session όταν κλήθηκε η results) σε αρχείο CSV.



Εικόνα 7. Το αποτέλεσμα μετά την εισαγωγή για αναζήτηση του λεκτικού “Machine Learning” για έτη μετά το 2018.

Ο Κώδικας της συνάρτησης download περιέχεται στο Παράρτημα 7.

Κεφάλαιο 5. Συμπεράσματα

Στόχος της εργασίας ήταν η ανάπτυξη ολοκληρωμένης εφαρμογής ανάκτησης και ανάλυσης δεδομένων της πηγής επιστημονικής πληροφόρησης Google Scholar. Δεδομένης της αμφισβήτησης που δέχεται η παραπάνω πηγή έγινε μία προσπάθεια διεξοδικής περιγραφής τόσο του τρόπου λειτουργίας της όσο και των πλεονεκτημάτων και μειονεκτημάτων που αυτή παρουσιάζει. Γίνεται αναλυτική βιβλιογραφική προσέγγιση στη λειτουργία της μέσα από επιστημονικές δημοσιεύσεις που έχουν προσπαθήσει να προσεγγίσουν την αξιοπιστία των δεδομένων του scholar αλλά και να επιχειρήσουν μια κριτική προσέγγιση στη λειτουργία του. Επιπλέον έγινε πλήρης ανάπτυξη τόσο του περιβάλλοντος λειτουργίας της εφαρμογής (web server και database) όσο και της πλήρους εφαρμογής από το στάδιο της διαμόρφωσης του html ερωτήματος όσο και της λήψης της απάντησης, της οπτικοποίησης της και της αποθήκευσης των αντίστοιχων δεδομένων σε μορφή Comma Separated Value (CSV) για να επεξεργαστούν σε excel. Αποτελεί επόμενο βήμα η άντληση και περισσότερων στοιχείων πέρα της λίστας δημοσιεύσεων ή του αποκλειστικού αριθμού αναφορών ανά δημοσίευση καθώς και η λήψη δεδομένων από άλλες βάσεις επιστημονικής πληροφόρησης όπως το Scopus.

ΠΑΡΑΡΤΗΜΑ 1

```
# -*- coding: utf-8 -*-

import requests
import urllib.parse

#input string
#query = 'machine learning'
query = input('Query: ')

#encode spaces and special characters
query_encoded = urllib.parse.quote(query)

#Serapi key
api_key =
'483f3bfc7292b39ecef791f815abc4928a17daddc9fefe8bd6b27d66c334de22'

#construct url
url= 'https://serpapi.com/search?engine=google_scholar&api_key=' +
api_key + '&q=' + query_encoded

#get results from api
r = requests.get(url)

#get json data
data = r.json()

#get results and print them
results = data['organic_results']
```

```
for i in range(len(results)):
    item = results[i]

    print('Result %d:' % (i+1))

    print('Title: %s' % item['title'])

    print('Publication info: %s' %
item['publication_info']['summary'])

    print('Link: %s' % item['link'])

    if ('cited_by' in item['inline_links'].keys()):

        print('Cited by: %d' %
item['inline_links']['cited_by']['total'])

    else:

        print('Cited by: 0')

    print('')
```

ΠΑΡΑΡΤΗΜΑ 2

```
from flask import Flask

from flask import (Blueprint, flash, g, redirect, render_template,
request, session,url_for, make_response)

import flask_excel as excel

from flask_session import Session

import requests

import urllib.parse

app = Flask(__name__)

excel.init_excel(app)

app.secret_key = 'super secret key'

app.config['SESSION_TYPE'] = 'filesystem'

Session(app)

@app.route("/")

def home():

    return render_template('home.html')
```

ΠΑΡΑΡΤΗΜΑ 3

```
<!doctype html>

<title>Search Google Scholar</title>

<link rel="stylesheet" href="{ url_for('static',
filename='style.css') }" >

<nav>

  <h1><a href="/" style='text-decoration:none'>Search Google
  Scholar</a></h1>

</nav>

<section class="content">

{% block content %}{% endblock %}

</section>

<hr>

<section class="footer">

&copy; 2023

</section>
```


ΠΑΡΑΡΤΗΜΑ 4

```
{% extends 'base.html' %}

{% block content %}

<form name='form1' action="results" method="POST">
    <label for='query'>Query:</label>
    <input type='text' name='query'><br><br>

    <label for='alg'>Year from:</label>
    <input type='number' name='as_ylo' value=''>
    <br>

    <label for='alg'>Year to:</label>
    <input type='number' name='as_yhi' value=''>
    <br>
    <br>

    <input type='submit' value='Submit'>
</form>

{% endblock %}
```

ΠΑΡΑΡΤΗΜΑ 5

```
@app.route('/results', methods=("GET", "POST"))

def results():

    txt = ""

    if request.method == "POST":

        #get form data

        d = request.form

        #for key in d:

        # txt += "var %s = %s<br>" % (key, d[key])

        query = d["query"]

        as_ylo = d["as_ylo"]

        as_yhi = d["as_yhi"]

        start = d["start"]

        start_int = int(start)

        #encode spaces and special characters

        query_encoded = urllib.parse.quote(query)

        #Serapi key

        api_key =

'483f3bfc7292b39ecef791f815abc4928a17daddc9fefe8bd6b27d66c334de22'

        #construct url

        url=

'https://serpapi.com/search?engine=google_scholar&api_key=' +

api_key + '&q=' + query_encoded

        #if optional params were given

        if as_ylo:
```

```

        url = url + '&as_ylo=' + as_ylo

    if as_yhi:
        url = url + '&as_yhi=' + as_yhi

    if start_int:
        url = url + '&start=' + start

    #print(url)

    #get results from api
    r = requests.get(url)

    #get json data
    data = r.json()

    #get results and print them
    res = data['organic_results']

    session['res'] = res

    data = {}
    data["query"] = query
    data["as_ylo"] = as_ylo
    data["as_yhi"] = as_yhi
    data["start"] = start
    data["start_int"] = start_int

    return render_template("results.html", results = res,
data=data)

```

ΠΑΡΑΡΤΗΜΑ 6

```
{% extends 'base.html' %}

{% block content %}
<a href='/'>Back</a><br>

<h2>Results</h2>

{% for item in results:%}

    <div class='res_item'>

        <div
class='res_title'><b>{{item['title']}}</b></div>

        <div class='res_link'><a
href='{{item['link']}}'>{{item['link']}}</a></div>

        <div
class='res_summary'>{{item['publication_info']['summary']}}</div>

        {% if ('cited_by' in item['inline_links'].keys()) %}

            <div class='res_citations'>Cited by:
{{item['inline_links']['cited_by']['total']}}</div>

        {% else: %}

            <div class='res_citations'>Cited by: 0</div>

        {% endif%}

    </div>

{% endfor %}

<div>
```

```

        <a href="download" style="margin: 0
5px;"><button>Download as CSV</button></a>

</div>

```

```

<div class='pagination'>
    {% if (data['start_int'] > 0) %}
    <div class='page_button'>
        <form method="POST">
            <input type="hidden" name="query"
value="{{data['query']}}">
            <input type="hidden" name="as_ylo"
value="{{data['as_ylo']}}">
            <input type="hidden" name="as_yhi"
value="{{data['as_hi']}}">
            <input type="hidden" name="start"
value="{{data['start_int'] - 10}}">
            <input type="submit" name="Previous"
value="Previous">
        </form>
    </div>
    {% endif%}

```

```

<div class='page_button'>
    <form method="POST">
        <input type="hidden" name="query"
value="{{data['query']}}">
        <input type="hidden" name="as_ylo"
value="{{data['as_ylo']}}">
        <input type="hidden" name="as_yhi"
value="{{data['as_hi']}}">
        <input type="hidden" name="start"
value="{{data['start_int'] + 10}}">
    </form>

```

```
                                <input type="submit" name="Next"
value="Next">
                                </form>
                                </div>
                                </form>
                                </div>

{% endblock %}
```

ΠΑΡΑΡΤΗΜΑ 7

```
@app.route('/download', methods=['GET'])
def download():
    header = ["Title", "Link", "Cited_by", "Summary"]
    data = [header]

    for item in session['res']:
        cit = 0
        if 'cited_by' in item['inline_links'].keys():
            cit = item['inline_links']['cited_by']['total']

        data.append([item['title'], item['link'], cit,
item['publication_info']['summary']])

    output = excel.make_response_from_array(data, 'csv')

    return output
```

BIBΛΙΟΓΡΑΦΙΑ

- Aguillo, I. F. (2012). Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics*, 91(2), 343-351.
- Arlitsch, K., & O'Brien, P. S. (2012). Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar. *Library Hi Tech*, 30(1), 60–81
- Bailey, J., Zhang, C., Budgen, D., Turner, M., & Charters, S. (2007, May). Search Engine Overlaps: Do they agree or disagree?. In *Second International Workshop on Realising Evidence-Based Software Engineering (REBSE'07)* (pp. 2-2). IEEE.
- Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *BMC Biomedical Digital Libraries*, 3, 7.
- Bar-Ilan, J. (2008). Which h-index?—A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271.
- Beel, J., & Gipp, B. (2009, July). Google Scholar's ranking algorithm: an introductory overview. In *Proceedings of the 12th international conference on scientometrics and informetrics (ISSI'09)* (Vol. 1, pp. 230-241).
- Beel, J., & Gipp, B. (2010). On the robustness of Google Scholar against spam. In *Proceedings of the 21 st ACM conference on hypertext and hypermedia* (pp. 297–298). Retrieved from. <http://dl.acm.org/citation.cfm?id=1810683>
- Bornmann, L., & Mutz, R. (2014). From P100 to P100': A new citation-rank approach. *Journal of the Association for Information Science and Technology*, 65(9), 1939-1943.
- Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A., & Daniel, H. D. (2009). Convergent validity of bibliometric Google Scholar data in the field of chemistry—Citation counts for papers that were accepted by *Angewandte Chemie International Edition* or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts. *Journal of informetrics*, 3(1), 27-35.
- Bramer, W. M., Giustini, D., Kramer, B. M., & Anderson, P. F. (2013). The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: A review of searches used in systematic reviews. *Systematic Reviews*, 2(1), 1
- Cheek, J., Garnham, B., & Quan, J. (2006). What's in a number? Issues in providing evidence of impact and quality of research (ers). *Qualitative health research*, 16(3), 423-435.

- Chitu, A. (2009). Google's Market Share in Your Country. *Website, March*.
- Couzin, G., & Grappone, J. (2006). Search Engine Optimization: an hour a day.
- Croft, W. B., Metzler, D., & Strohman, T. (2015). Search Engines. In *Information retrieval in practice*. Pearson Education, Inc..
- Cusker, J. (2013). Elsevier Compendex and Google Scholar: a quantitative comparison of two resources for engineering research and an update to prior comparisons. *The Journal of Academic Librarianship*, 39(3), 241–243.
- DeGraff, J. V., DeGraff, N., & Romesburg, H. C. (2013). Literature searches with Google Scholar: Knowing what you are and are not getting. *GSA Today*, 23(10), 44–45
- Delgado López-Cózar, E., & Robinson-García, N. (2012). Repositories in Google Scholar Metrics or what is this document type doing in a place as such? Retrieved from. <http://digibug.ugr.es/handle/10481/22019>
- Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science*. Elsevier Science Publishers.
- Evans, M. P. (2007). Analysing Google rankings through search engine optimization data. *Internet research*.
- García-Pérez, M. A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in Psychology. *Journal of the American Society for Information Science and Technology*, 61(10), 2070–2085.
- García-Pérez, M. A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in Psychology. *Journal of the American society for information science and technology*, 61(10), 2070-2085.
- Geng, G. G., Wang, C. H., & Li, Q. D. (2008, January). Improving Spamdexing Detection Via a Two-Stage Classification Strategy. In *Asia Information Retrieval Symposium* (pp. 356-364). Springer, Berlin, Heidelberg.
- Georgas, H. (2014). Google vs. the library (part II): student search patterns and behaviors when using Google and a federated search tool. *portal: Libraries and the Academy*, 14(4), 503-532.

- Giles, L., & Khabisa, M. (2014). The Number of scholarly documents on the Web. *PLoS One*, 9(5), e93949.
- Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., Haak, L. L., & Kington, R. (2011). Race, ethnicity, and NIH research awards. *Science*, 333(6045), 1015-1019.
- Giustini, D., & Boulos, M. N. K. (2013). Google Scholar is not enough to be used alone for systematic reviews. *Online Journal of Public Health Informatics*, Retrieved from <http://journals.uic.edu/ojs/index.php/ojphi/article/view/4623>
- Goodman, D. (2005). Web of Science (2004 version) and Scopus. *The Charleston Advisor*, 6(3), 5-5.
- Google Scholar Metrics. . Google Scholar. [Internet]. Available from <http://scholar.google.com/intl/en/scholar/metrics.html>
- Google Scholar Team. 2012 May 21. Google Scholar Sources. [Personal e-mail]. Accessed 2012 Jun 25.
- Grigas, V., Juzeniene, S., & Velickaite, J. (2017). " Just Google It"--The Scope of Freely Available Information Sources for Doctoral Thesis Writing. *Information Research: An International Electronic Journal*, 22(1), n1.
- Haines, L. L., Light, J., O'Malley, D., & Delwiche, F. A. (2010). Information-seeking behavior of basic science researchers: implications for library services. *Journal of the Medical Library Association: JMLA*, 98(1), 73.
- Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature. *Journal of informetrics*, 11(3), 823-834.
- Harzing, A. W. (2010). *The publish or perish book*. Melbourne, Australia: Tarma Software Research Pty Limited.
- Harzing, A. W., & Alakangas, S. (2017). Microsoft Academic: is the phoenix getting wings?. *Scientometrics*, 110(1), 371-383.
- Harzing, A.-W., & Van Der Wal, R. (2009). A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. *Journal of the American Society for Information Science and Technology*, 60(1), 41–46

- Hauser, A. (2012) May 21. Google Scholar Sources. [Personal e-mail]. Accessed 2012 Jun 25.
- Hendrix, C. (2015). Google Scholar Metrics and Scholarly Productivity in International Relations. *Duck of Minerva*.
- Holden, G., Rosenberg, G., & Barker, K. (2005). Tracing thought through time and space: A selective review of bibliometrics in social work. *Social Work in Health Care*, 41(3-4), 1-34.
- Jacobs, J. A. (2011). Journal rankings in sociology: Using the H index with google scholar. *The American Sociologist*, 1–33.
- Jacobs, J. A. (2016). Journal rankings in sociology: Using the H index with google scholar. *American Sociologist*, 47(2–3), 192–224. <http://dx.doi.org/10.1007/s12108-015-9292-7>
- Jacsó, P. (2005). Google Scholar: the pros and the cons. *Online information review*.
- Jacsó, P. (2008). Google scholar revisited. *Online information review*, 32(1), 102-114.
- Kalyvas, S. N. (1999). Wanton and senseless? The logic of massacres in Algeria. *Rationality and Society*, 11(3), 243-285.
- Kalyvas, S. N. (2006). *The logic of violence in civil war*. Cambridge University Press.
- Khabisa, M., & Giles, C. L. (2014). The number of scholarly documents on the public web. *PLoS One*, 9(5), e93949. <http://dx.doi.org/10.1371/journal.pone.0093949>
- Khan, S., Liu, X., Shakil, K. A., & Alam, M. (2017). A survey on scholarly data: From big data perspective. *Information Processing & Management*, 53(4), 923-944.
- Kousha, K., Thelwall, M., & Rezaie, S. (2011). Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus. *Journal of the American Society for Information Science and Technology*, 62(11), 2147–2164.
- Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S. S. (2005). The effect of use and access on citations. *Information processing & management*, 41(6), 1395-1402.
- Lascar, C., & Mendelsohn, L. D. (2011a). The evolution–intelligent design debate: a meaningful context for teaching the nature of science in information literacy. Part 1:

- historical background and philosophical considerations. *Science & Technology Libraries*, 30(4), 354-371.
- Lascar, C., & Mendelsohn, L. D. (2011β). The evolution–intelligent design debate: a meaningful context for teaching the nature of science in information literacy. Part 1: historical background and philosophical considerations. *Science & Technology Libraries*, 30(4), 354-371.
- Levay, P., Ainsworth, N., Kettle, R., & Morgan, A. (2016). Identifying evidence for public health guidance: A comparison of citation searching with Web of Science and Google Scholar. *Research Synthesis Methods*, 7(1), 34–45. <http://dx.doi.org/10.1002/jrsm.1158>
- Levine-Clark, M., & Gil, E. L. (2008). A comparative citation analysis of web of science, scopus, and google scholar. *Journal of Business & Finance Librarianship*, 14(1), 32–46
- Lewandowski, D. (2010). Google Scholar as a tool for discovering journal articles in library and information science. *Online Information Review*, 34(2), 250–262.
- Lewandowski, D., & Mayr, P. (2006). Exploring the academic invisible web. *Library hi tech*.
- Lopez-Cozar, E. D., Robinson-Garcia, N., & Torres-Salinas, D. (2012). Manipulating google scholar citations and google scholar metrics: Simple, easy and tempting. arXiv Preprint arXiv:1212.0638. Retrieved from. <http://arxiv.org/abs/1212.0638>
- Martín-Martín, A., Orduna-Malea, E., Ayllón, J. M., & López-Cózar, E. D. (2014). Does Google Scholar contain all highly cited documents (1950–2013)? arXiv:1410.8464 [Cs] [Retrieved from]. <http://arxiv.org/abs/1410.8464>
- Mastrangelo, G., Fadda, E., Rossi, C. R., Zamprogno, E., Buja, A., & Cegolon, L. (2010). Literature search on risk factors for sarcoma: PubMed and Google Scholar may be complementary sources. *BMC Research Notes*, 3(1), 131.
- Mayr, P., & Walter, A. K. (2007). An exploratory study of Google Scholar. *Online information review*.
- Mayr, P., & Walter, A. K. (2008). Studying journal coverage in Google Scholar. *Journal of Library Administration*, 47(1-2), 81-99.

- Mayr, P., & Walter, A.-K. (2008). Studying journal coverage in google scholar. *Journal of Library Administration*, 47(1–2), 81–99.
- Meier, J. J., & Conkling, T. W. (2008). Google Scholar's coverage of the engineering literature: an empirical study. *The Journal of Academic Librarianship*, 34(3), 196-201.
- Meier, J. J., & Conkling, T. W. (2008). Google Scholar's coverage of the engineering literature: an empirical study. *The Journal of Academic Librarianship*, 34(3), 196-201.
- Meier, J. J., & Conkling, T. W. (2008). Google Scholar's coverage of the engineering literature: An empirical study. *The Journal of Academic Librarianship*, 34(3), 196–201.
- Mikki, S. (2010). Comparing google scholar and ISI web of science for earth sciences. *Scientometrics*, 82(2), 321–331.
- Minasny, B., Hartemink, A. E., McBratney, A., & Jang, H.-J. (2013). Citations and the h index of soil researchers and journals in the Web of Science, Scopus, and Google Scholar. *PeerJ*, 1, e183.
- Moed, H. F., Bar-Ilan, J., & Halevi, G. (2016). A new methodology for comparing Google Scholar and Scopus. *Journal of Informetrics*, 10(2), 533–551.
<http://dx.doi.org/10.1016/j.joi.2016.04.017>
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the national academy of sciences*, 109(41), 16474-16479.
- Mullen, L. B., & Hartman, K. A. (2006). Google Scholar and the library web site: The early response by ARL libraries. *College & Research Libraries*, 67(2), 106-122.
- Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. *Journal of Informetrics*, 1(2), 161–169.
- Noruzi, A. (2005). Google Scholar: The new generation of citation indexes.
- Orduna-Malea, ~ E., Ayllón, J. M., Martín-Martín, A., & López-Cózar, E. D. (2014). About the size of Google Scholar: Playing the numbers. arXiv Preprint arXiv:1407.6239.
 Retrieved from. <http://arxiv.org/abs/1407.6239>

Orduña-Malea, E., Martín-Martín, A., Ayllon, J. M., & López-Cózar, E. D. (2014). The silent fading of an academic search engine: the case of Microsoft Academic Search. *Online information review*.

Prins, A. A., Costas, R., van Leeuwen, T. N., & Wouters, P. F. (2016). Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data. *Research Evaluation*, 25(3), 264-270.

Reiter, D. (2016). Citation Count Data and Faculty Promotion. *Duck of Minerva*.

Scholar Help. Google Scholar. [Internet]. Available from:

<http://scholar.google.com/intl/en/scholar/help.html>

Shariff, S. Z., Bejaimal, S. A., Sontrop, J. M., Iansavichus, A. V., Haynes, R. B., Weir, M. A., & Garg, A. X. (2013). Retrieving clinical evidence: A comparison of PubMed and Google Scholar for quick clinical searches. *Journal of Medical Internet Research*, 15(8), e164.

Shultz, M. (2007). Comparing test searches in PubMed and Google Scholar. *Journal of the Medical Library Association: JMLA*, 95(4), 442.

Stephan, P., Veugelers, R., & Wang, J. (2017). Reviewers are blinkered by bibliometrics. *Nature*, 544(7651), 411-412.

Tober, M. (2011). PubMed, ScienceDirect, Scopus or Google Scholar—Which is the best search engine for an effective literature research in laser medicine? *Medical Laser Application*, 26(3), 139–144.

Top Publications. Google Scholar. [Internet]. Available from: http://scholar.google.com/citations?view_op=top_venues&hl=en

van Aalst, J. (2010). Using Google Scholar to estimate the impact of journal articles in education. *Educational Researcher*, 39(5), 387–400.

Walters, W. H. (2007). Google Scholar coverage of a multidisciplinary field. *Information Processing & Management*, 43(4), 1121–1132

Walters, W. H. (2011). Comparative recall and precision of simple and expert searches in Google Scholar and eight other databases. *portal: Libraries and the Academy*, 11(4), 971-1006.

Wang, D., Song, C., & Barabási, A. L. (2013). Quantifying long-term scientific impact. *Science*, 342(6154), 127-132.