



**ΤΜΗΜΑ ΑΡΧΕΙΟΝΟΜΙΑΣ, ΒΙΒΛΙΟΘΗΚΟΝΟΜΙΑΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΗΣΗΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΤΙΚΩΝ, ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**DEPARTMENT OF ARCHIVAL, LIBRARY AND INFORMATION STUDIES
SCHOOL OF MANAGEMENT, ECONOMICS AND SOCIAL SCIENCES**

Πτυχιακή Εργασία

Κατάταξη Αποτελεσμάτων στις Μηχανές Αναζήτησης.

**Μια Συγκριτική Μεθοδολογική Προσέγγιση μεταξύ της Google και
της Bing**

Βασίλειος - Παναγιώτης Νταραράς (ΑΜ: 59917032)

Γεώργιος Ντίμο (ΑΜ: 59917100)

Επιβλέπων: Δημήτριος Κουής

Αθήνα, Ιούλιος 2023

Επιτροπή Εξέτασης

1. Δημήτριος Κουής

2. Σαράντος Καπιδάκης

3. Ιωάννης Τριανταφύλλου

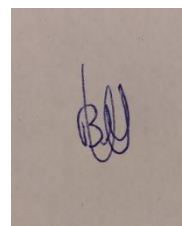
ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Νταραράς Βασίλειος - Παναγιώτης , με αριθμό μητρώου 59917032 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Διοικητικών, Οικονομικών και Κοινωνικών Επιστημών του Τμήματος Αρχαιονομίας, Βιβλιοθηκονομίας και Συστημάτων Πληροφόρησης, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της πτυχιακής/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 πτυχίου μου».

Ο Δηλών



Βασίλειος - Παναγιώτης Νταραράς

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Ντίμο Γεώργιος, με αριθμό μητρώου 59917100 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Διοικητικών, Οικονομικών και Κοινωνικών Επιστημών του Τμήματος Αρχαιονομίας, Βιβλιοθηκονομίας και Συστημάτων Πληροφόρησης, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της πτυχιακής/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 πτυχίου μου».

Ο Δηλών



Γεώργιος Ντίμο

Ευχαριστίες – Αφιερώσεις

Η παρούσα πτυχιακή εργασία με θέμα “Κατάταξη αποτελεσμάτων στις μηχανές αναζήτησης: Μια συγκριτική μεθοδολογική προσέγγιση μεταξύ της Google και της Bing” αποτελεί την πρώτη επιστημονική μας προσπάθεια. Θα θέλαμε να ευχαριστήσουμε τους καθηγητές του τμήματος για την διαρκή στήριξή τους όλα αυτά τα χρόνια και ιδιαίτερα τους επιβλέποντες καθηγητές μας Κουή Δημήτριο και Δρίβα Ιωάννη, οι οποίοι μας παρείχαν συμβουλές και παρατηρήσεις καθ’ όλη τη διάρκεια της εκπόνηση της εργασίας μας.

Θέλουμε, επίσης, να ευχαριστήσουμε τους γονείς και τους φίλους μας, οι οποίοι μας προσέφεραν κάθε μορφής στήριξης κατά τη διάρκεια των σπουδών μας στο Τμήμα Αρχαιονομίας, Βιβλιοθηκονομίας και Συστημάτων Πληροφόρησης,

Ημερομηνία 7/2023

Συγγραφέας Νταραράς Βασίλειος- Παναγιώτης, Ντίμο Γεώργιος

Περίληψη στα ελληνικά

Σκοπός και Πλαίσιο Μελέτης: Η σύγχρονη εποχή, η οποία εύστοχα χαρακτηρίζεται ως η εποχή της “πληροφορίας” λόγω του τεράστιου όγκου δεδομένων που παράγεται διαρκώς στον παγκόσμιο ιστό. Για την διαχείριση της τεράστιας ποσότητας δεδομένων και πληροφοριών επιστρατεύονται οι μηχανές αναζήτησης ως μέσα ανάκτησης, διάδοσης και οπτικοποίησης της πληροφορίας καλύπτοντας τις ανάγκες των χρηστών σε ποσότητα και ποιότητα. Η ποσότητα των αποτελεσμάτων εκφράζεται με τον αριθμό που ανακτάται ανά ερώτημα που υποβάλλει ο εκάστοτε χρήστης, ενώ η ποιότητα παρουσιάζεται μέσω της κατάταξης των αποτελεσμάτων σε μία μηχανή αναζήτησης. Η παρούσα μελέτη έχει σκοπό να συγκρίνει τα αποτελέσματα δύο μηχανών αναζήτησης της Google και της Bing τόσο ως προς την πιθανότητα αλληλοκάλυψης και το βαθμό ομοιότητας τους, όσο και ως προς την απόδοση αυτών των αποτελεσμάτων ως ιστοσελίδες. Η συγκεκριμένη απόδοση λαμβάνεται υπόψη μέσω τριών αξόνων αξιολόγησης, το βαθμό επιμέλειας των μεταδεδομένων των ιστοσελίδων, την ταχύτητα και την ασφάλεια.

Σχεδιασμός Έρευνας/Μεθοδολογία: Για την επίτευξη του σκοπού μελέτης αναπτύχθηκε σχετική μεθοδολογία, η οποία υποστήριξε συνολικά την ερευνητική μας προσπάθεια. Πιο συγκεκριμένα, αναζητούμε τους 10 δημοφιλέστερους όρους του 2021 μέσω της εφαρμογής Google Trends. Στη συνέχεια, αναζητούμε τους 10 όρους στις μηχανές αναζήτησης Google και Bing και συγκεντρώνουμε τα 10 πρώτα αποτελέσματα για κάθε έναν από τους 10 όρους. Ύστερα, διαλέγουμε τα 5 πρώτα αποτελέσματα-URLs του κάθε όρου και από τις δύο μηχανές αναζήτησης. Στην επόμενη φάση, συγκεντρώνουμε τα εναπομείναντα αποτελέσματα, τα οποία αναλύουμε μέσω του εργαλείου Checkbot. Με βάση τα αποτελέσματα του Checkbot, δημιουργούμε δύο φύλλα Excel για τα αποτελέσματα καθεμίας μηχανής αναζήτησης. Τα φύλλα αυτά, έχουν 3 στήλες (Search Term, URL, Website Score), στις οποίες καταχωρούμε τους όρους, τα 5 URLs του κάθε όρου και η βαθμολογία που συγκεντρώνει στο Checkbot. Συνολικά χρησιμοποιήσαμε περισσότερες από 34 μετρικές απόδοσης. Ένα βήμα παρακάτω αναπτύχθηκαν μοντέλα πρόβλεψης μέσω γραμμικών παλινδρομήσεων κατά ζεύγη για την προτεραιοποίηση της βελτιστοποίησης ανά μετρική.

Ενδεικτικά Αποτελέσματα: Η ερευνητική μας προσπάθεια παρουσίασε μία σειρά από αποτελέσματα-μετρήσεις τα οποία αποτυπώνουν τη δυναμική των δύο μηχανών αναζήτησης Google και Bing. Ειδικότερα, ενδιαφέρον στατιστικό δεδομένο αποτελεί το 5% ποσοστό αλληλοκάλυψης (Overlapping). Πιο συγκεκριμένα, μόλις 5% των συνολικών αποτελεσμάτων των Google και Bing ανήλθαν σε κοινή θέση στην κατάταξη αποτελεσμάτων που όρισαν οι

μηχανές. Επιπλέον, συγκεκριμένες μετρικές εμφανίζουν την ανάλογη μεγαλύτερη απήχηση βελτιστοποίησης συγκριτικά με άλλες και για τους τρεις άξονες αξιολόγησης.

Πρωτοτυπία/ Συνεισφορά: Η εν λόγω έρευνα αποτελεί μια από τις πιο πρόσφατες προσπάθειες σύγκρισης μεταξύ Google και Bing. Για την επίτευξη των στόχων της έρευνας, όπως αναφέρεται παραπάνω συγκροτήθηκε μεθοδολογική προσέγγιση, η οποία δημιουργήθηκε εξ ολοκλήρου χωρίς την ύπαρξη προγενέστερων παρόμοιων ολοκληρωμένων ερευνητικών προσεγγίσεων. Η μεθοδολογία μας οδήγησε στην δημιουργία βημάτων ανάλυσης και ερμηνείας της απόδοσης των ιστοσελίδων. Με άλλα λόγια, μια απλή και συνεχόμενη διαδικασία βελτίωσης των μετρικών των ιστοσελίδων, ώστε να επιτύχουν καλύτερη θέση κατάταξης και παράλληλα καλύτερη εμπειρία χρήστη στο περιβάλλον της ιστοσελίδας και στις δύο μηχανές αναζήτησης.

Λέξεις Κλειδιά: (μηχανές αναζήτησης, google, bing, βελτιστοποίηση μηχανών αναζήτησης, seo, κατάταξη ιστοσελίδων, google trends, σύγκριση, μετρικές ιστού, βελτιστοποίηση ιστοσελίδων)

Περίληψη στα αγγλικά

Ranking Results in Search Engines. A Comparative Methodological Perspective between Google & Bing

Purpose and Context of the Study: The modern age, is aptly described as the "information age" due to the huge amount of data that is constantly being generated on the web. To manage the huge amount of data and information, search engines are being enlisted as a means of retrieving, disseminating and visualizing information to cover the needs of users in terms of quantity and quality. The quantity of results is expressed in terms of the number of retrieved documents per query submitted by each user, while the quality is represented by the ranking of results in a search engine. This study aims to compare the results of two search engines Google and Bing both in terms of the probability of overlap and their degree of similarity, and in terms of the performance of these results as web pages. This performance is taken into account through three evaluation aspects, the degree of metadata curation of the web pages, speed and security.

Research Design/Methodology: To achieve the study purpose, a relevant methodology was developed, that supported our overall research effort. More specifically, we searched for the 10 most popular terms of 2021 through the Google Trends web application. We then search for the 10 terms on Google and Bing search engines and compile the top 10 results for each of the 10 terms. Then, we pick the top 5 results-URLs of each term, that is the first five websites from both search engines to evaluate the performance of these selected websites. The selected websites have been analyzed through the CheckBot crawler web-based tool. According to the Checkbot results, we create two Excel sheets for the results of each search engine. These sheets, have 3 columns (Search Term, URL, Website Score), in which we enter the terms, the 5 URLs of each term and the score that it accumulates in Checkbot. Totally, we used more than 34 performance metrics of metadata, speed and security effectiveness. One step further, predictive models were developed using stepwise linear regressions to prioritize optimization by metric.

Indicative Results: Our research effort presented a series of results-measurements that encapsulate the dynamics of the two search engines Google and Bing. In particular, an interesting statistic is the 5% overlap rate. It is also worth noting that only 5% of the total results of Google and Bing were ranked together in the results ranking defined by the engines.

Furthermore, specific metrics showed a proportionately higher optimization impact compared to others for all three evaluation axes.

Originality/Contribution: This research is one of the most recent attempts to compare Google and Bing. In order to achieve the research objectives, as mentioned above, a methodological approach was put together, which was created entirely without the existence of previous similar integrated research approaches. The methodology led us to create steps for analyzing and interpreting the performance of the websites. In other words, a simple and continuous process of improving the metrics of the websites to achieve better ranking position, and at the same time, greater user experience in the website's interface in both search engines.

Keywords: (search engines, google, bing, search engine optimization, search engine optimization, seo, web ranking, google trends, comparison, web metrics, web optimization)

Πίνακας περιεχομένων

ΕΠΙΤΡΟΠΗ ΕΞΕΤΑΣΗΣ	II
ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ	III
ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ	IV
ΕΥΧΑΡΙΣΤΙΕΣ – ΑΦΙΕΡΩΣΕΙΣ	V
ΠΕΡΙΛΗΨΗ ΣΤΑ ΕΛΛΗΝΙΚΑ	VI
ΠΕΡΙΛΗΨΗ ΣΤΑ ΑΓΓΛΙΚΑ	VIII
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	X
ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ	XII
ΠΙΝΑΚΑΣ ΠΙΝΑΚΩΝ	XIII
ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ	1
1.1 ΠΛΑΙΣΙΟ ΤΗΣ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ	1
1.2 ΣΚΟΠΟΣ- ΣΤΟΧΟΙ	2
1.3 ΜΕΘΟΔΟΛΟΓΙΑ	2
1.4 ΟΡΓΑΝΩΣΗ ΚΕΦΑΛΑΙΩΝ / ΔΙΑΡΘΡΩΣΗ ΤΗΣ ΕΡΓΑΣΙΑΣ	3
1.5 ΟΡΙΣΜΟΙ	4
ΚΕΦΑΛΑΙΟ 2. ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ – ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΡΕΥΝΑ – ΣΧΕΤΙΚΕΣ ΠΡΟΣΠΑΘΕΙΕΣ	5
2.1 ΜΗΧΑΝΕΣ ΑΝΑΖΗΤΗΣΗΣ ΚΑΙ ΣΚΟΠΟΣ	5
2.2 ΣΧΕΤΙΚΕΣ ΠΡΟΣΠΑΘΕΙΕΣ – ΕΡΕΥΝΕΣ	5
2.2.1 <i>Προσπάθειες στην μηχανή αναζήτησης Google</i>	5
2.2.2 <i>Συγκριτικές Μελέτες</i>	9
ΚΕΦΑΛΑΙΟ 3. ΜΕΘΟΔΟΛΟΓΙΑ Η/ΚΑΙ ΣΧΕΔΙΟ ΕΡΓΑΣΙΩΝ– ΥΛΟΠΟΙΗΣΗ – ΕΦΑΡΜΟΓΗ	13
3.1 ΣΥΛΛΟΓΗ ΚΑΙ ΟΡΓΑΝΩΣΗ ΔΕΔΟΜΕΝΩΝ	14
3.1.1 <i>Επεξήγηση μετρικών απόδοσης ιστοσελίδων</i>	23
3.2 ΜΕΘΟΔΟΙ & ΕΡΓΑΛΕΙΑ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ	30
3.2.1 <i>Περιγραφική στατιστική ανάλυση</i>	31
3.2.2 <i>Ανάλυση Παλινδρόμησης</i>	32
ΚΕΦΑΛΑΙΟ 4. ΑΠΟΤΕΛΕΣΜΑΤΑ – ΕΥΡΗΜΑΤΑ / ΕΠΙΤΕΥΓΜΑΤΑ	34
4.1 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΕΡΙΓΡΑΦΙΚΗΣ ΣΤΑΤΙΣΤΙΚΗΣ ΑΝΑΛΥΣΗΣ	34
4.1.1 <i>Διαγράμματα Μηχανής Αναζήτησης Google</i>	39

4.1.2	Σύγκριση διαγραμμάτων Μηχανών Google και Bing	42
4.2	ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΝΑΛΥΣΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ	43
4.2.1	Αποτελέσματα Παλινδρόμησης για τον Άξονα του SEO	43
4.2.2	Αποτελέσματα Παλινδρόμησης για τον Άξονα της Ταχύτητας	47
4.2.3	Security Αποτελέσματα Παλινδρόμησης για τον Άξονα της Ασφαλείας	50
4.2.4	Διαγράμματα Παλινδρόμησης και Εκτιμώμενες Τιμές Μεταβολής	52
ΚΕΦΑΛΑΙΟ 5. ΣΥΖΗΤΗΣΗ – ΣΥΜΠΕΡΑΣΜΑΤΑ – ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ.....		55
5.1	ΣΥΖΗΤΗΣΗ ΕΠΙ ΤΗΣ ΜΕΘΟΔΟΛΟΓΙΑΣ	55
5.2	ΒΑΣΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ	61
5.3	ΠΡΟΒΛΗΜΑΤΑ ΕΡΕΥΝΑΣ	61
5.4	ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ / ΠΡΑΚΤΙΚΕΣ ΠΡΟΕΚΤΑΣΕΙΣ ΤΗΣ ΈΡΕΥΝΑΣ	62
ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ.....		63
ΠΑΡΑΡΤΗΜΑ.....		67
5.5	ΔΕΔΟΜΕΝΑ ΕΡΕΥΝΑΣ	67

Πίνακας Σχημάτων

Εικόνα 1. Βήματα Μεθοδολογίας	13
Εικόνα 2. Περιβάλλον Google Trends	15
Εικόνα 3. Περιβάλλον Checkbot.....	18
Εικόνα 4. Περιβάλλον ανάλυσης URL Checkbot	19
Εικόνα 5. Περιβάλλον αποτελεσμάτων ανάλυσης Checkbot	20
Εικόνα 6. Περιβάλλον αναλυτικών αποτελεσμάτων SEO Checkbot	21
Εικόνα 7. Περιβάλλον αναλυτικών αποτελεσμάτων ταχύτητας (Speed) Checkbot	22
Εικόνα 8. Περιβάλλον αναλυτικών αποτελεσμάτων ασφάλειας (Security) Checkbot	23
Εικόνα 9. Περιβάλλον επεξεργασίας δεδομένων στην πλατφόρμα του JASP.....	31
Εικόνα 10. Περιβάλλον ανάλυσης των ερευνητικών δεδομένων στο JASP.....	31
Εικόνα 11. Βαθμολογίες Total Website Score Bing.....	36
Εικόνα 12. Βαθμολογίες SEO Bing.....	37
Εικόνα 13. Βαθμολογίες Speed Bing	37
Εικόνα 14. Βαθμολογίες Security Bing	38
Εικόνα 15. Βαθμολογίες Total Website Score Google	39
Εικόνα 16. Βαθμολογίες SEO Google	40
Εικόνα 17. Βαθμολογίες Speed Google.....	40
Εικόνα 18. Βαθμολογίες Security Google.....	41
Εικόνα 19. Διάγραμμα σύγκρισης των βασικών μετρικών Google και Bing.....	42
Εικόνα 20. Πίτα ποσοστού αλληλοκάλυψης των αποτελεσμάτων Google και Bing	42
Εικόνα 21. Εκτιμώμενες τιμές μεταβολής για κάθε μετρική στον άξονα SEO.....	52
Εικόνα 22 Εκτιμώμενες τιμές μεταβολής για κάθε μετρική στον άξονα Speed	53
Εικόνα 23. Εκτιμώμενες τιμές μεταβολής για κάθε μετρική στον άξονα Security	54
Εικόνα 24. Βήματα Ανάλυσης Απόδοσης και Βελτίωσης Ιστοσελίδων.....	59

Πίνακας Πινάκων

Πίνακας 1. Οι 10 δημοφιλέστεροι οροί για την μηχανή αναζήτησης της Google (2021).....	14
Πίνακας 2. Μέσος όρος της απόδοσης των ιστοσελίδων των δέκα πρώτων αποτελεσμάτων ανάκτησης στην μηχανή Bing.....	34
Πίνακας 3. Μέσος όρος της απόδοσης των ιστοσελίδων των δέκα πρώτων αποτελεσμάτων ανάκτησης στη μηχανή Google	35
Πίνακας 4. Σύγκριση. Μέσος όρος των μετρικών απόδοσης ανάμεσα στις δύο μηχανές.	35
Πίνακας 5. Αποτελέσματα παλινδρόμησης για τον άξονα του SEO	44
Πίνακας 6. Αποτελέσματα παλινδρόμησης για τον άξονα της ταχύτητας	48
Πίνακας 7. Αποτελέσματα παλινδρόμησης για τον άξονα της ασφαλείας.....	50

Κεφάλαιο 1. Εισαγωγή

1.1 Πλαίσιο της πτυχιακής εργασίας

Στην εποχή που διανύουμε, η πληροφορία έχει γίνει αναπόσπαστο κομμάτι της καθημερινότητας. Πλέον, είναι εύκολα προσβάσιμη, καθώς τα μέσα που τη δημιουργούν και τη διακινούν, ψάχνουν τρόπους να βελτιώνονται συνεχώς στο κομμάτι τη πρόσβασης αλλά και της ταχύτητας. Σήμερα, οι άνθρωποι έχοντας λύσει το πρόβλημα της πρόσβασης και της ταχύτητας στην αναζήτηση πηγών, καλούνται να διαχειριστούν τον τεράστιο όγκο δεδομένων και να τον αξιολογήσουν, προκειμένου να έχουν την καλύτερη δυνατή εμπειρία χρήστη στο διαδίκτυο. Στο σημείο αυτό, αναλαμβάνουν δράση οι μηχανές αναζήτησης, οι οποίες φροντίζουν για την εύρεση και προβολή της πληροφορίας, καθώς και την ασφάλεια που τη συνοδεύει.

Με απλά λόγια, μια μηχανή αναζήτησης αποτελεί το εργαλείο ανάκτησης των πληροφοριών στον παγκόσμιο ιστό. Δηλαδή η πληροφορία υπάρχει, εντούτοις αυτό που χρειάζεται είναι ο τρόπος ανάκτησης της. Και ανάκτηση της πληροφορίας πραγματοποιείται από τις μηχανές αναζήτησης, ενώ βασικός τους στόχος είναι η ανάκτηση της πληροφορίας στο βέλτιστο δυνατό χρόνο, ποσότητα, και ποιότητα (Kisiel, 2010; Ren, Yang and Diao, 2010).

Για την επίτευξη του συγκεκριμένου στόχου, οι μηχανές αναζήτησης επιστρατεύουν την κατάταξη των ιστοσελίδων, η οποία αξιολογεί και τοποθετεί σε συγκεκριμένη σειρά τα αποτελέσματα της εκάστοτε αναζήτησης, χρησιμοποιώντας κριτήρια που ορίζει η καθεμία. Η αξιολόγηση χρησιμοποιεί ως κριτήρια τη λειτουργικότητα, την ασφάλεια, το περιεχόμενο αλλά και το πλήθος επισκέψεων των χρηστών ώστε να προβάλλει και στους υπόλοιπους χρήστες το καλύτερο δυνατό περιεχόμενο που σχετίζεται με τις λέξεις κλειδιά του θέματος που αναζητούν (Mavridis and Symeonidis, 2015).

Ωστόσο, αυτό που παραμένει ακόμα θολό στην ερευνητική κοινότητα είναι ο βαθμός σχέσης μεταξύ των κριτηρίων που χρησιμοποιούν οι εκάστοτε μηχανές αναζήτησης, καθώς κάθε μηχανή ενδέχεται να χρησιμοποιεί διαφορετικά κριτήρια κατάταξης. Με βάση τα παραπάνω, γίνεται αντιληπτή η ανάγκη σύγκρισης των μηχανών αναζήτησης με σκοπό την κατανόηση των τεχνικών χαρακτηριστικών που εφόσον καταγραφούν μπορούν να δώσουν ένα καλά πληροφορημένο οδηγό βελτίωσης των ιστοσελίδων στους διαχειριστές για υψηλότερες κατατάξεις ανεξάρτητα τις μηχανές αναζήτησης (Google, Bing, Yahoo! κ.α).

1.2 Σκοπός- Στόχοι

Η έρευνα μας, με τον τρόπο που δομήθηκε το περιεχόμενο της έχει, ως στόχο την σύγκριση των μηχανών αναζήτησης Google και Bing μέσω της χρήσης από πλήθος μετρικών που καταγράφουν την απόδοση των ιστοσελίδων ως απολέσματα. Οι μετρήσεις, που πραγματοποιήθηκαν, αντικατοπτρίζουν τις διαφορές που έχουν οι δύο μηχανές αλλά και την βαθμολογική “απόσταση” που καταγράφεται αναμεσά τους. Επιπλέον, οι μετρήσεις των σημαντικότερων υπομετρικών κάθε μετρικής απευθύνονται στους δημιουργούς και διαχειριστές των ιστοσελίδων παρουσιάζοντας τους μία μεθοδολογία, που στοχεύει στη βελτίωση των ιστοσελίδων που επιμελούνται. Η βελτίωση των ιστοσελίδων κρίνουμε, ότι γίνεται αποδοτικότερη με τη χρήση του μεθοδολογικού μοτίβου που προτείνεται, από την ανάκτηση, την προ-επεξεργασία την ανάλυση, έως και την δημιουργία μοντέλων προτεραιοποίησης των απαιτήσεων για βελτιστοποίηση. Πιο συγκεκριμένα η παρούσα πτυχιακή εργασία στοχεύει στα εξής:

- Αρχικά να καταγράψει το βαθμό απόδοσης συγκεκριμένων ιστοσελίδων που ήρθαν πρώτες στα αποτελέσματα της Google και της Bing πάνω σε τρεις άξονες. Την απόδοση Μεταδεδομένων (SEO Metadata Crawling Performance), Την απόδοση Ταχύτητας (Speed Performance) και την απόδοση Ασφάλειας (Security Performance).
- Σε δεύτερο στάδιο στοχεύει στην καταγραφή των πιθανών διαφοροποιήσεων ανάμεσα στις δύο μηχανές αναζήτησης ως προς το βαθμό απόδοσης για τις ιστοσελίδες που εμφανίστηκαν ως αποτελέσματα σε κάθε μία από αυτές
- Να παρουσιάσει μοντέλα προτεραιοποίησης των απαιτήσεων αναδεικνύοντας μετρικές απόδοσης που ενέχουν αμεσότερης βελτίωσης έναντι άλλων βοηθώντας πρακτικά τους διαχειριστές

1.3 Μεθοδολογία

Για την διεκπεραίωση της ερευνητικής μας προσπάθειας και την κάλυψη των στόχων αναπτύχθηκαν πολλαπλά επιμέρους μεθοδολογικά βήματα. Αρχικά, διαλέξαμε τους δημοφιλέστερους όρους που αναζητήθηκαν το 2021 στην μηχανή αναζήτησης της Google. Οι όροι επιλέχθηκαν από το Google Trends, το οποίο είναι εφαρμογή της Google, όπως προδίδει το όνομα της. Μετά την επιλογή των όρων, ακολούθησε η ανάκτηση από τη Google και την Bing, των δέκα πρώτων αποτελεσμάτων- URLs για κάθε έναν όρο. Στη συνέχεια, από τα δέκα αποτελέσματα για κάθε όρο, διατηρήσαμε τα 5 πρώτα, προσδίδοντας με αυτόν τον τρόπο,

μεγαλύτερη ακρίβεια, όσον αφορά το περιεχόμενο του όρου και την μέτρηση της απόδοσής τους. Με τη σειρά τους, τα 5 πρώτα αποτελέσματα κάθε όρου, αναλύθηκαν μέσω του Checkbot, ενός εργαλείου που καταγράφει περισσότερες από 60 μετρικές απόδοσης των ιστοσελίδων. Πιο συγκεκριμένα, η εν λόγω εφαρμογή ανέλυσε τα URLs στις βασικές μετρικές Total Website Score, SEO, Speed, Security και των υπομετρικών τους, δίνοντας στο κάθε URL μια βαθμολογία. Τα ευρήματα του Checkbot, περιελάμβαναν τα δεδομένα που συγκέντρωσε η κάθε μετρική αλλά και οι υπομετρικές τους, οι οποίες ευθύνονται για την τελική βαθμολογία της ιστοσελίδας που βρίσκεται υπό ανάλυση. Τα δεδομένα αυτά, εκφράζονται μέσω διαγραμμάτων στο κεφάλαιο 4.

Το επόμενο στάδιο της έρευνας μας, αφορά την ανάλυση των δεδομένων των υπομετρικών του Total Website Score που επιλέξαμε κρίνοντας, ότι επηρεάζουν την πρόβλεψη της μελλοντικής βελτίωσης των τριών βασικών μετρικών και της συνολικής απόδοσης μιας ιστοσελίδας (Total Website SEO Score). Για την επίτευξη της παραπάνω πρότασης χρησιμοποιήθηκε το περιβάλλον της JASP (Jeffreys's Amazing Statistics Program). Το συγκεκριμένο πρόγραμμα στατιστικής ανάλυσης υποστηρίζει τη στατιστική ικανότητα πρόβλεψης (prediction capability) της κάθε υπομετρικής. Τα δεδομένα του JASP διατίθενται στο κεφάλαιο 4 μέσω διαγραμμάτων.

1.4 Οργάνωση Κεφαλαίων / Διάρθρωση της Εργασίας

Η ερευνητική μας προσπάθεια διαρθρώνεται σε μικρότερα μέρη με σκοπό την καλύτερη οργάνωσή αλλά και κατανόησή της από τους δυνητικούς αναγνώστες. Στο **κεφάλαιο 1**, αναλύεται το εισαγωγικό κείμενο, το οποίο περιγράφει το πλαίσιο τους στόχους και τους σκοπούς της εργασίας. Στο **κεφάλαιο 2**, αναφέρονται προγενέστερες ερευνητικές προσπάθειες του επιστημονικού αντικειμένου που ερευνούμε. Στο **κεφάλαιο 3**, αναλύεται λεπτομερώς η μεθοδολογία που χρησιμοποιήθηκε για την εκπόνηση της έρευνας κάνοντας αναφορά στα εργαλεία που χρησιμοποιήθηκαν για να τελεσφορήσει η ερευνητική μας προσπάθεια. Στο **κεφάλαιο 4**, ακολουθεί η προβολή των ευρημάτων-αποτελεσμάτων που προήλθαν από τα εργαλεία που χρησιμοποιήθηκαν. Τέλος, στο **κεφάλαιο 5**, η έρευνα κλείνει με τη συζήτηση επί των ευρημάτων, το συμπέρασμα αλλά και τις μελλοντικές ερευνητικές προσπάθειες.

1.5 Ορισμοί

SEO – Search Engine Optimization: Βελτιστοποίηση Μηχανών Αναζήτησης. Διαδικασία βελτιστοποίησης ενός ιστότοπου και του περιεχομένου του για τη βελτίωση της ορατότητας και της κατάταξής του στις σελίδες αποτελεσμάτων των μηχανών αναζήτησης.

Search Engine: Μια μηχανή αναζήτησης είναι ένα πρόγραμμα λογισμικού που αναζητά σε μια βάση δεδομένων πληροφορίες σύμφωνα με το ερώτημα του χρήστη. Η μηχανή παρέχει έναν κατάλογο αποτελεσμάτων που ταιριάζουν καλύτερα σε αυτό που προσπαθεί να βρει ο χρήστης.

Κατάταξη: Η κατάταξη στις μηχανές αναζήτησης αναφέρεται στη θέση μιας ιστοσελίδας στα αποτελέσματα αναζήτησης. Υψηλότερη κατάταξη σημαίνει καλύτερη ορατότητα και περισσότερη οργανική επισκεψιμότητα.

Αποτελέσματα μηχανής αναζήτησης: Τα αποτελέσματα στις μηχανές αναζήτησης είναι οι ιστοσελίδες που εμφανίζονται στους χρήστες μετά την εισαγωγή ενός ερωτήματος αναζήτησης.

Google Trends: Το Google Trends είναι ένα εργαλείο που παρέχει πληροφορίες σχετικά με τη δημοτικότητα και τον όγκο αναζήτησης συγκεκριμένων λέξεων-κλειδιών ή θεμάτων με την πάροδο του χρόνου.

JASP: Το JASP (Jeffreys’s Amazing Statistics Program) είναι ένα ελεύθερο λογισμικό ανοικτού κώδικα για στατιστική ανάλυση.

Αλληλοκάλυψη (Overlapping): Η αλληλοκάλυψη στις κατατάξεις αναφέρεται στο σενάριο, όπου πολλοί ιστότοποι διεκδικούν τις ίδιες λέξεις-κλειδιά, οδηγώντας σε κοινές θέσεις στα αποτελέσματα των μηχανών αναζήτησης.

Checkbot: Μια επέκταση προγράμματος περιήγησης (browser extension) που βοηθά τους χρήστες να βελτιστοποιήσουν τους ιστότοπούς τους, ελέγχοντας και παρακολουθώντας το SEO, την ταχύτητα και την ασφάλειά τους.

Κεφάλαιο 2. Θεωρητικό μέρος – Βιβλιογραφική έρευνα – Σχετικές προσπάθειες

2.1 Μηχανές αναζήτησης και σκοπός

Είναι κοινά παραδεκτό, ότι με την εξέλιξη της τεχνολογίας και την συνακόλουθη διάδοση του διαδικτύου οι μηχανές αναζήτησης αποτελούν μεγάλο μέρος της καθημερινότητας του μέσου ανθρώπου. Μέσω των μηχανών αναζήτησης, ικανοποιούνται οι πληροφοριακές και οι καταναλωτικές ανάγκες των χρηστών, λόγω των δυνατοτήτων που προσφέρουν. Η κάλυψη των πληροφοριακών αναγκών επιτυγχάνεται μέσω της αλληλεπίδρασης των χρηστών με το εκάστοτε πληροφοριακό σύστημα. Με βάση, τα στατιστικά στοιχεία της Firms99 “ο μέσος άνθρωπος πραγματοποιεί 3-4 αναζητήσεις ημερησίως” αναδεικνύοντας ακόμα περισσότερο τις δυνατότητες που προσφέρονται στους τελικούς δέκτες της πληροφορίας. Η ανάγκη αυτή, έχει συμβάλει στην δημιουργία μιας πληθώρας μηχανών αναζήτησης με κυρίαρχο την Google και ακολουθούν οι υπόλοιπες.

2.2 Σχετικές προσπάθειες – έρευνες

2.2.1 Προσπάθειες στην μηχανή αναζήτησης Google

Οι Cahill και Chalut (2009) προσπάθησαν να ενισχύσουν την γνώση της βελτιστοποίησης μηχανών αναζήτησης (SEO) και πώς μπορεί να συμβάλει στις βιβλιοθήκες και τις παρεχόμενες ηλεκτρονικές υπηρεσίες τους. Οι συγγραφείς επικεντρώθηκαν εκτενώς στην εξήγηση της διαφοροποίησης μεταξύ των τεχνικών “λευκού και μαύρου καπέλου”. Δύο πτυχές που στις μέρες μας μάλλον μοιάζουν ξεπερασμένες, ενώ είναι σαφές σε όλους τους προγραμματιστές ιστοσελίδων, ότι η μηχανή Google ποινικοποιεί τις αθέμιτες τακτικές (black-hat tactics). Σε κάθε περίπτωση, παρόλο που η εργασία δημοσιεύθηκε το 2009, οι Cahill και Chalut ρίχνουν φως στο ρόλο των βιβλιοθηκονόμων όσον αφορά την πτυχή του SEO. Καταλήγουν σε δύο σημαντικές συστάσεις. Πρώτον, ότι οι βιβλιοθηκονόμοι θα πρέπει να διδάξουν στους χρήστες να γνωρίζουν τον όρο SEO και πώς αυτό επηρεάζει τις σελίδες αποτελεσμάτων των μηχανών αναζήτησης από την άποψη της πληροφοριακής παιδείας. Και δεύτερον, σχετικά με το πώς οι βιβλιοθηκονόμοι θα πρέπει να εισάγουν τους χρήστες όχι μόνο στην ανάκτηση πληροφοριών που προέρχονται μόνο από τις μηχανές αναζήτησης, αλλά

και να τους εισάγουν σε εξειδικευμένο θεματικό περιεχόμενο και βάσεις δεδομένων άρθρων πλήρους κειμένου που προέρχονται από πηγές του Deep Web (Βαθύς ιστός).

Με βάση τον αριθμό των αναφορών, η εργασία των Arlitsch και O'Brian (2012) που δημοσιεύθηκε στο *Library Hi Tech* το 2012 αποτελεί την πιο σημαντική ερευνητική προσπάθεια γύρω από το θέμα του SEO και την εφαρμογή του στο πλαίσιο των Βιβλιοθηκών, Αρχείων και Μουσείων (εφεξής BAM). Αυτό που καθιστά αυτή την έρευνα ιδιαίτερα δημοφιλή συγκριτικά με τις υπόλοιπες, είναι ότι κατά κύριο λόγο στηρίζεται σε τρεις άξονες που είναι σε θέση να αξιοποιηθούν πρακτικά στο πλαίσιο της στρατηγικής ανάπτυξης ενός ιδρυματικού αποθετηρίου. Πρώτον, εξετάζει με συγκριτικό τρόπο ποιο λογισμικό αποθετηρίου ορίζει τα υψηλότερα ποσοστά ευρετηρίασης στο Google Scholar. Γεγονός ικανό να υποστηρίξει τη λήψη αποφάσεων από τους διαχειριστές βιβλιοθηκών όσον αφορά το είδος του λογισμικού που θα υιοθετήσουν τελικά στο ιδρυματικό τους αποθετήριο. Δεύτερον, εξετάζεται ότι ο μετασχηματισμός των προεπιλεγμένων σχημάτων μεταδεδομένων των ιδρυματικών αποθετηρίων σε προτιμώμενα σχήματα του Google Scholar είχε ως αποτέλεσμα τη σημαντική αύξηση των ποσοστών ευρετηρίασης. Και τρίτον, παρείχαν πρακτικές και βήμα προς βήμα προτάσεις για αυτόν τον μετασχηματισμό των προεπιλεγμένων μεταδεδομένων των αποθετηρίων σε προτιμώμενα σχήματα μεταδεδομένων Google Scholar, όπως τα Highwire Prints, Eprints ή BePress και PRISM (Publishing Requirements for Industry Standard Metadata). Ένα χρόνο αργότερα, οι ερευνητές δημοσιεύουν παρόμοια μελέτη ωστόσο αυτή τη φορά δίνουν έμφαση στην δημιουργία και οριοθέτηση SEO πολιτικής στο πλαίσιο της ανάπτυξης της δημοφιλίας ενός οργανισμού στον Παγκόσμιο Ιστό (Arlitsch et al. 2013)

Οι Onaifo και Rassmussen (2013) προχώρησαν σε μια διαδικασία αξιολόγησης 52 ιστότοπων δημόσιων βιβλιοθηκών στο Οντάριο. Οι συγγραφείς συνέλεξαν δεδομένα για τέσσερις κατηγορίες: την παγκόσμια κατάταξη επισκεψιμότητας, τη φήμη, τις συνολικές ημερήσιες επισκέψεις σελίδων και τις συνολικές σελίδες που έχουν αναπροσαρμοστεί στο Google. Η παγκόσμια κατάταξη επισκεψιμότητας είναι μια κατάταξη που απονέμεται από την Alexa σε κάθε ιστότοπο στον Παγκόσμιο Ιστό. Η φήμη εκτιμήθηκε με βάση τους εξωτερικούς συνδέσμους που έρχονται στον εξεταζόμενο ιστότοπο οι οποίοι ανακτήθηκαν επίσης μέσω της Alexa. Οι αριθμοί συνολικών ημερήσιων επισκέψεων σελίδων ανακτήθηκαν μέσω του Websiteoutlook.com. Τέλος, ο αριθμός των ευρετηριασμένων σελίδων της Google συλλέχθηκε μέσω του RankCheck.org. Εκτός από την παρουσίαση περιγραφικών στατιστικών

στοιχείων που δείχνουν την τάση των εξεταζόμενων βιβλιοθηκών με βάση αυτές τις μετρήσεις, οι συγγραφείς διαπίστωσαν αρκετές συσχετίσεις μεταξύ των μετρήσεων. Η παγκόσμια κατάταξη της ιστοσελίδας και η φήμη της ιστοσελίδας έδειξαν συντελεστή συσχέτισης Pearson (-,368), ενώ η παγκόσμια επισκεψιμότητα της ιστοσελίδας και η φήμη της ιστοσελίδας συσχετίστηκαν με την τιμή 0,-759 ρ. Τέλος, η συνολική ημερήσια επισκεψιμότητα ιστοσελίδας διατύπωσε σημαντική συσχέτιση με το σύνολο των αναπροσαρμοσμένων από την Google σελίδων της ιστοσελίδας (.962). Χωρίς αμφιβολία, η τελευταία συσχέτιση που διαπίστωσαν οι συγγραφείς, ενισχύει την ανάγκη θέσπισης τεχνικών SEO που είναι ικανές να αυξήσουν τον αριθμό των ευρετηριασμένων σελίδων που ορίζουν οι ιστότοποι βιβλιοθηκών στις μηχανές αναζήτησης. Με άλλα λόγια, όσο μεγαλύτερος είναι ο αριθμός των ευρετηριασμένων ιστοσελίδων, τόσο μεγαλύτερος είναι ο αριθμός των επισκέψεων των χρηστών σε μια ιστοσελίδα και, ως εκ τούτου, μεγαλύτερες πιθανότητες αλληλεπίδρασης τους με το παρεχόμενο.

Σε σύγκριση με προηγούμενες μεθόδους ανάλυσης ιστού, οι Chakravarty και Wasan (2015) χρησιμοποίησαν μια διαφορετική υπολογιστική προσέγγιση. Υιοθέτησαν τον Web Impact Factor (WIF) και τον Revised Web Impact Factor (R-WIF) για να αξιολογήσουν 10 ιστότοπους βιβλιοθηκών ανώτατων εκπαιδευτικών ιδρυμάτων στην Ινδία. Η μεθοδολογική τους προσέγγιση ήταν απλή και σαφής και κατανεμήθηκε σε τρία βήματα. Αρχικά, για να βρεθεί ο αριθμός των σελίδων self-link, in-link και εξωτερικές σελίδες συνδέσμων ιστοτόπων, δεύτερον για τον υπολογισμό του WIF και του R-WIF και τέλος, για την εκτίμηση της πιθανής συσχέτισης μεταξύ WIF και R-WIF μέσω του συντελεστή συσχέτισης Spearman. Διαπίστωσαν υψηλή τιμή του συντελεστή συσχέτισης (.702) μεταξύ των δύο παραγόντων, η οποία συνεπάγεται μεγάλη εγγύτητα μεταξύ τους. Πιθανόν να ήταν εκτός του πεδίου εφαρμογής της μελέτης των συγκεκριμένων ερευνητών, ωστόσο, δεν υπάρχουν αποτελέσματα που θα ήταν ικανά να βελτιστοποιήσουν τους εξεταζόμενους ιστότοπους πανεπιστημιακών βιβλιοθηκών. Με άλλα λόγια, πώς αυτά τα αποτελέσματα θα μπορούσαν να χρησιμοποιηθούν από τους προγραμματιστές των μηχανών αναζήτησης για τη βελτίωση του πεδίου εφαρμογής τους, δηλαδή της καλύτερης δυνατής ανάκτησης πληροφοριών σε χρόνο, ποσότητα και ποιότητα με βάση τους όρους αναζήτησης που πληκτρολογούν οι χρήστες.

Μια λεπτομερής ερευνητική εξέταση πραγματοποιήθηκε από τους Krstić και Masliković (2018). Οι συγγραφείς πρότειναν μια μεθοδολογία δύο επιπέδων με σκοπό να κατανοήσουν το επίπεδο ορατότητας αναζήτησης των ιστότοπων βιβλιοθηκών, αρχείων και μουσείων στην

περιοχή της Σερβίας. Στο πρώτο επίπεδο αναπτύχθηκε μια έρευνα για την κατανόηση της εξοικείωσης του προσωπικού με τις τεχνικές SEO. Στο δεύτερο επίπεδο οι συγγραφείς εξέτασαν τους ιστότοπους μέσω διαδικτυακών εργαλείων SEO με βάση τις προτεινόμενες πτυχές SEO των Enge et al. (2015), δηλαδή On-Site, Off-Site και Technical SEO. Το WooRank χρησιμοποιήθηκε για το On-Site, το SEOReviwTools χρησιμοποιήθηκε για το Off-Site και ένας συνδυασμός μεταβλητών του WooRank, του Google PageSpeedInsights και του Google MobileFriendlyTest εφαρμόστηκε για την εκτίμηση της απόδοσης του Τεχνικού SEO παράγοντα. Με βάση τα ευρήματά τους, οι ιστοσελίδες αρχαικών οργανισμών υπέδειξαν τη μεγαλύτερη τεχνική απόδοση SEO σε σύγκριση με τις βιβλιοθήκες και τα μουσεία. Οι βιβλιοθήκες υπέδειξαν το υψηλότερο επίπεδο επιδόσεων στην πτυχή On-Site, ενώ τα μουσεία είχαν ως αποτέλεσμα την υψηλότερη απόδοση στην πτυχή Off-Site. Εκτός από τα αποτελέσματα των διαδικτυακών εργαλείων SEO, οι συγγραφείς υπογράμμισαν το μέτριο ή χαμηλό επίπεδο του προσωπικού ως προς τις τεχνικές SEO και τις αναλυτικές δεξιότητες που γνωρίζουν.

Οι Tavosi και Naghshineh (2021) διερεύνησαν 42 δικτυακούς τόπους πανεπιστημιακών βιβλιοθηκών και τις επιδόσεις τους στην περιοχή του Ιράν. Χρησιμοποίησαν 3 διαφορετικά εργαλεία, το SimilarWeb, το Ahrefs και τα διαθέσιμα εργαλεία του W3C για τον εντοπισμό ευθυγράμμισης SSL και σφαλμάτων JavaScript ή CSS. Η αξιολόγησή τους στηρίχθηκε σε επτά παράγοντες, δηλαδή τον αριθμό των backlinks, τη μέση διάρκεια επίσκεψης, το επίπεδο του ποσοστού αναπήδησης σε άλλη σελίδα (bounce rate), τον αριθμό των κλικ, το πιστοποιητικό SSL, τον αριθμό των σφαλμάτων σε Javascript και CSS και τέλος τον αριθμό των σελίδων που έχουν ευρετηριαστεί στη Google. Οι συγγραφείς παρουσίασαν μέσω γραφημάτων τις επιδόσεις των εξεταζόμενων ιστοσελίδων πανεπιστημιακών βιβλιοθηκών με βάση τους επτά παράγοντες. Επίσης, εξίσωσαν τα ευρήματά τους με άλλες προηγούμενες προσεγγίσεις που κατέληξαν σε παρόμοια συμπεράσματα.

Μια άλλη μελέτη στην περιοχή του Ιράν προέρχεται από τους Fahimnia και Eltemasi (2021). Οι συγγραφείς χρησιμοποίησαν δύο διαφορετικά διαδικτυακά εργαλεία, το SEOSiteCheckUp και το SEOptimer, για να εξάγουν κοινά στοιχεία που θα μπορούσαν να χρησιμοποιηθούν για την αξιολόγηση ιστοτόπων. Παρατηρήθηκε επαρκής συντελεστής συμφωνίας 0,77 μεταξύ των δύο εργαλείων. Τέλος, διερεύνησαν 22 ακαδημαϊκές ιατρικές βιβλιοθήκες χρησιμοποιώντας 30 μετρικές SEO. Οι συγγραφείς χρησιμοποίησαν επίσης το εργαλείο Ahrefs για να εκτιμήσουν και να συγκρίνουν το συνολικό ποσό των backlinks μεταξύ των 22

περιπτώσεων. Τα αποτελέσματά τους έδειξαν τις χαμηλές επιδόσεις των ιστοτόπων όσον αφορά αυτά τα συστατικά, ενώ μόνο 5 από τους 22 εξεταζόμενους ιστοτόπους έλαβαν από 50 έως 75 βαθμούς στην τελική εκατοσταβάθμια βαθμολογία SEO. Κατέληξαν στο συμπέρασμα ότι η έρευνά τους θα πρέπει να ληφθεί υπόψη από το προσωπικό των ιατρικών βιβλιοθηκών του Ιράν με στόχο να βελτιωθεί η δυνατότητα ανακάλυψη και εύρεση τέτοιων ιστοτόπων στον Παγκόσμιο Ιστό.

2.2.2 Συγκριτικές Μελέτες

Μέσω μιας λεπτομερούς μεθοδολογίας οι French και Fagan (2019) εξέτασαν σε ποιο βαθμό τα αρχεία αρχών, τα αναγνωριστικά των ερευνητών και τα κοινωνικά προφίλ ακαδημαϊκής δικτύωσης, επηρεάζουν την ορατότητα αναζήτησης των ερευνητών στις μηχανές αναζήτησης Google και Bing. Αυτή η ερευνητική προσπάθεια θα μπορούσε να προσφέρει πρακτικές συμβουλές στις ακαδημαϊκές βιβλιοθήκες και στον τρόπο βελτίωσης της ευρεσιμότητας των ερευνητικών αντικειμένων. Πιο συγκεκριμένα, οι συγγραφείς εξήγαγαν τις εγγραφές αυθεντίας από το Library of Congress Name Authority File (LCNAF), το Virtual International Authority File (VIAF) και τα WorldCat Identities. Τα αναγνωριστικά των ερευνητών εξήχθησαν από το ORCID και το International Standard Name Identifier (ISNI). Προφίλ ακαδημαϊκής κοινωνικής δικτύωσης (ASN) που αντλήθηκαν από τα Academia, Research Gate, bpress και Google Scholar. Το δείγμα αποτελούνταν από 24 μέλη Διδακτικού Ερευνητικού Προσωπικού και 35 δημοσιεύσεις που σχετίζονται με τα αρχεία αρχής τους. Οι συγγραφείς κατέληξαν στο συμπέρασμα ότι τα μέλη ΔΕΠ με περισσότερα προφίλ ASN είναι πιο ορατά στις μηχανές αναζήτησης και επίσης τα μέλη ΔΕΠ με αρχεία κύρους είναι πιθανότερο να κατατάσσονται υψηλότερα στη μηχανή αναζήτησης Google. Τα Research Gate, bpress και Google Scholar αυξάνουν δραστικότερα την ορατότητα των μελών ΔΕΠ στα αποτελέσματα των μηχανών αναζήτησης σε σύγκριση με τα υπόλοιπα προφίλ ASN. Σημειώνεται ότι οι French και Fagan απέδειξαν έμμεσα τη δύναμη που έχει η δημιουργία backlink ως τεχνική Off-Site SEO για την αύξηση της ορατότητας αναζήτησης των δημοσιεύσεων των ερευνητών. Ανοίγουν επίσης νέους ερευνητικούς διαλόγους σχετικά με το πώς οι ακαδημαϊκές βιβλιοθήκες είναι σε θέση να αναπτύξουν τις υπηρεσίες SEO που παρέχονται στα μέλη ΔΕΠ για την αύξηση της ορατότητας των ερευνητικών εργασιών τους στον Παγκόσμιο Ιστό.

Ο Lewandowski, κ.ά (2006) στην μέλετη τους χρησιμοποίησαν 3 μηχανές αναζήτησης, οι οποίες ήταν οι Google, Yahoo και MSN. Η μελέτη αυτή, μετρά τη συχνότητα με την οποία, οι

μηχανές αναζήτησης ενημερώνουν τους δείκτες τους. Τα συμπεράσματα είναι, ότι η ποιότητα των διαφόρων ευρετηρίων των μηχανών αναζήτησης ποικίλλει και ότι θα πρέπει να χρησιμοποιούνται περισσότερες από μία μηχανές κατά την αναζήτηση τρέχοντος περιεχομένου.

Ο Bar-Ilan, J. (2004) από το Εβραϊκό Πανεπιστήμιο της Ιερουσαλήμ, εξέτασε την απόδοση των μηχανών αναζήτησης στον Παγκόσμιο Ιστό με την πάροδο του χρόνου. Το κεφάλαιο συζητά επίσης εν συντομία την αρχιτεκτονική των μηχανών αναζήτησης. Αποδεικνύοντας, ότι οι μηχανές αναζήτησης δεν αντιμετωπίζουν πάντα ικανοποιητικά τις δυναμικές αλλαγές.

Οι Höchstätter και Lewandowski (2009) διερευνούν την σύνθεση των αποτελεσμάτων των μηχανών αναζήτησης, ορίζοντας ποια στοιχεία χρησιμοποιούν οι πιο δημοφιλείς μηχανές αναζήτησης στο διαδίκτυο στις σελίδες αποτελεσμάτων τους και σε ποιο βαθμό χρησιμοποιούνται για δημοφιλή έναντι σπάνιων ερωτημάτων. στέλνοντας 500 ερωτήματα στις Google, Yahoo, Live.com και Ask. Συνολικά, η συγκεκριμένη μελέτη βασίστηκε σε 42.758 στοιχεία.

Σύμφωνα με τους Shahzad (2018) κ.ά η βελτιστοποίηση μηχανών αναζήτησης (SEO) παίζει πολύ σημαντικό ρόλο στην ανάπτυξη επαγγελματικών ιστότοπων. Υπάρχουν ορισμένες μηχανές αναζήτησης διαθέσιμες στο διαδίκτυο, όπως οι Yahoo, Ask.com, AOL.com, Baidu και Bing. Μεταξύ των οποίων, η Google είναι η πιο ευρέως χρησιμοποιούμενη μηχανή αναζήτησης. Κάθε μηχανή αναζήτησης χρησιμοποιεί διαφορετική τεχνική και αλγόριθμο SEO, ο οποίος όχι μόνο αποτελεί τη βάση του SEO αλλά επηρεάζει και τη θέση ενός ιστότοπου στα οργανικά αποτελέσματα αναζήτησης. Καθώς, η Google τροποποιεί τον αλγόριθμό της περίπου 500 ή περισσότερες φορές το χρόνο, ο σχεδιασμός ιστοσελίδων και το διαδίκτυο εξελίσσονται επίσης δυναμικά λόγω των αλλαγών στις τεχνικές και τους αλγόριθμους SEO. Ωστόσο, το πόσο καλά οι ιστότοποι των πανεπιστημίων της Μαλαισίας είναι βελτιστοποιημένοι για άλλες μηχανές αναζήτησης είναι αμφίβολο, ιδίως λόγω των κρίσιμων διαφορών μεταξύ των τεχνικών και των αλγορίθμων κατάταξης των μηχανών αναζήτησης. Η παρούσα ερευνητική εργασία τείνει να απαντήσει σε αυτά τα ζωτικά ερωτήματα προτείνοντας μια συγκριτική ανάλυση της Bing και της Google σε ορισμένες ιστοσελίδες πανεπιστημίων της Μαλαισίας, αναλύοντας τις παραμέτρους βελτιστοποίησης των μηχανών αναζήτησης και τα αποτελέσματα της χρήσης της Microsoft Bing σε σύγκριση με τον κύριο ανταγωνιστή της, τη Google.

Ο Edosomwan (2010) συνέκρινε τις επιδόσεις της ανάκτησης πληροφοριών ορισμένων δημοφιλών μηχανών αναζήτησης (συγκεκριμένα, Google, Yahoo, AlltheWeb, Gigablast, Zworks και AltaVista και Bing/MSN) σε απάντηση σε έναν κατάλογο δέκα ερωτημάτων, διαφορετικής πολυπλοκότητας. Τα ερωτήματα αυτά εκτελέστηκαν σε κάθε μηχανή αναζήτησης και καταγράφηκαν η ακρίβεια και ο χρόνος απόκρισης των ανακτηθέντων αποτελεσμάτων. Τα πρώτα δέκα έγγραφα σε κάθε αποτέλεσμα ανάκτησης αξιολογήθηκαν ως "σχετικά" ή "μη σχετικά" για την αξιολόγηση της ακρίβειας της μηχανής αναζήτησης. Για την αξιολόγηση του χρόνου απόκρισης, υπολογίστηκαν κανονικοποιημένοι λόγοι ανάκλησης σε διάφορα σημεία αποκοπής για κάθε ερώτημα και μηχανή αναζήτησης. Η μελέτη αυτή δείχνει ότι η Google φαίνεται να είναι η καλύτερη μηχανή αναζήτησης τόσο από άποψη μέσης ακρίβειας (70%) όσο και από άποψη μέσου χρόνου απόκρισης (2 s). Η Gigablast και η AlltheWeb είχαν τις χειρότερες επιδόσεις συνολικά σε αυτή τη μελέτη.

Σύμφωνα με την έρευνα του Almukhtar (2021) κ.ά. οι μηχανές αναζήτησης παίζουν καθοριστικό ρόλο στην παρουσίαση των σωστών σελίδων στον χρήστη λόγω της διαθεσιμότητας ενός τεράστιου αριθμού ιστότοπων. Οι μηχανές αναζήτησης, όπως η Google, χρησιμοποιούν τον αλγόριθμο κατάταξης σελίδων για να βαθμολογήσουν τις ιστοσελίδες ανάλογα με τη φύση του περιεχομένου τους και την ύπαρξή τους στον παγκόσμιο ιστό. Το SEO μπορεί να χαρακτηριστεί ως η μεθοδολογία που χρησιμοποιείται για την ανύψωση του ιστότοπου έχοντας κατά νου τον τελικό στόχο να έχει υψηλή κατάταξη, δηλαδή, κορυφαίο αποτέλεσμα. Σε αυτό το έγγραφο οι συγγραφείς παρουσιάζουν τις περισσότερες μηχανές βελτιστοποίησης αναζήτησης όπως (Google, Bing, MSN, Yahoo, κ.λπ.), και συγκρίνουν με την απόδοση της βελτιστοποίησης μηχανών αναζήτησης. Οι συγγραφείς παρουσιάζουν επίσης τα οφέλη, τους περιορισμούς, τις προκλήσεις και την εφαρμογή της βελτιστοποίησης μηχανών αναζήτησης στις επιχειρήσεις.

Κατά τον Britvic (2014), κ.ά. η βελτιστοποίηση μηχανών αναζήτησης είναι ένα σημαντικό έργο στον επαγγελματικό σχεδιασμό ιστοσελίδων. Η Google κυριαρχεί στην αγορά των μηχανών αναζήτησης, αλλά υπάρχουν και άλλες, δηλαδή η Bing, η Yahoo κ.λπ. Πόσο καλά βελτιστοποιούνται οι ιστότοποι για άλλες μηχανές αναζήτησης και αξίζει τον κόπο; Ποιες είναι οι κύριες διαφορές μεταξύ των αλγορίθμων των μηχανών αναζήτησης; Αυτή η ερευνητική εργασία τείνει να απαντήσει σε αυτά τα ερωτήματα συγκρίνοντας την Google και την Bing. Η έρευνα περιλαμβάνει μελέτη παραδείγματος κροατικών ειδησεογραφικών

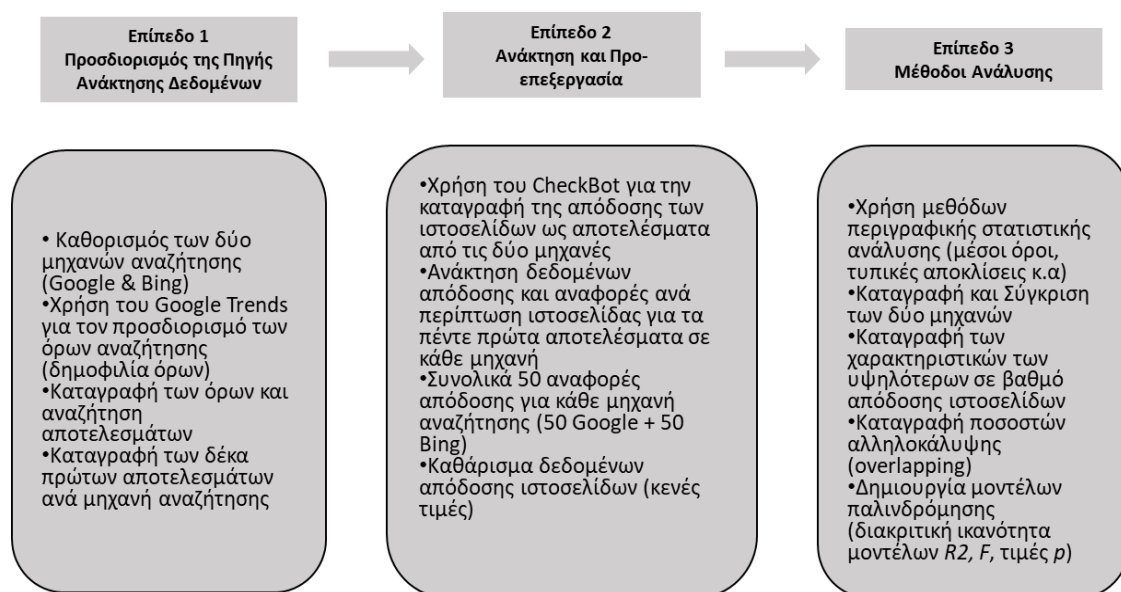
πυλών, αναλύοντας τις παραμέτρους SEO τους και τα αποτελέσματα που παίρνουν σε μια Google και τον κύριο ανταγωνιστή της, το Microsoft Bing.

Όπως αναφέρεται στην έρευνα του Barbar και Ismail (2019) με την ανάπτυξη του διαδικτύου τα τελευταία χρόνια, οι μηχανές αναζήτησης όπως η Google, η Bing και η Yahoo γίνονται όλο και πιο σημαντικές και αξιόπιστες. Ο ρόλος των μηχανών αναζήτησης είναι να ευρετηριάσουν δισεκατομμύρια ιστοσελίδες και να εμφανίζουν μόνο τα πιο σχετικά αποτελέσματα για ένα δεδομένο ερώτημα αναζήτησης. Κατά τη δημιουργία της ιστοσελίδας, πολλοί διαχειριστές ιστοσελίδων ξεχνούν να λάβουν υπόψη τους έναν ουσιαστικό παράγοντα, ο οποίος είναι η γνωστοποίηση της ιστοσελίδας τους στον κόσμο. Τις περισσότερες φορές το κύριο βάρος δίνεται στο να γίνει ο δικτυακός τόπος όσο το δυνατόν πιο φιλικός προς τον χρήστη, σταθερός, γρήγορος και ασφαλής. Ωστόσο, όλες αυτές οι τεχνικές μπορεί να είναι άχρηστες αν ο ιστότοπος δεν έχει επισκέπτες ή αν οι άνθρωποι απλώς δεν μπορούν να τον βρουν. Για την επίλυση αυτού του προβλήματος και για την ενίσχυση της δομής της ιστοσελίδας ώστε να γίνει πιο φιλική προς τις μηχανές αναζήτησης, αναπτύσσεται μια διαδικτυακή εφαρμογή που αναλύει οποιαδήποτε δεδομένη ιστοσελίδα και παρέχει πληροφορίες για το πώς μπορεί να ενισχυθεί και να βελτιωθεί η δομή της ώστε να γίνει πιο φιλική προς τις μηχανές αναζήτησης και να βελτιωθεί η κατάταξή της στις μηχανές αναζήτησης. Η διαδικασία αυτή είναι επίσης γνωστή ως βελτιστοποίηση μηχανών αναζήτησης (SEO).

Κεφάλαιο 3. Μεθοδολογία ή/και Σχέδιο εργασιών– Υλοποίηση – Εφαρμογή

Η παρούσα μελέτη έχει ως σκοπό να συγκρίνει τις μηχανές αναζήτησης της Google και της Bing. Θα εστιάσει στην αξιολόγηση του πόσο καλά λειτουργούν αυτές οι μηχανές όταν παρέχουν αποτελέσματα για ιστοσελίδες. Θα χρησιμοποιήσει διάφορα κριτήρια για να μετρήσει πόσο καλά επιδόθηκαν οι ιστοσελίδες σε τρεις πτυχές: τις λέξεις-κλειδιά που περιέχουν (SEO), την ταχύτητα φόρτωσής τους και την ασφάλειά τους. Επίσης, θα εξετάσει τις διαφορές μεταξύ των δύο μηχανών αναζήτησης στο πώς αποδίδουν τις ιστοσελίδες. Με βάση αυτές τις αξιολογήσεις, θα παρουσιάσει μοντέλα για να βοηθήσει τους διαχειριστές ιστοσελίδων να βελτιώσουν την απόδοσή τους.

Για την επίτευξη αυτών των στόχων, η παρακάτω εικόνα παρουσιάζει εν τάχει τη μεθοδολογία της ερευνητικής μας προσπάθειας. Με βάση το σχήμα, η μεθοδολογία χωρίζεται σε τρία μέρη.



Εικόνα 1. Βήματα Μεθοδολογίας

Το Επίπεδο 1, με τίτλο “Προσδιορισμός της Πηγής Ανάκτησης Δεδομένων” παρουσιάζει την επιλογή των μηχανών αναζήτησης που πήραν μέρος στη σύγκριση αλλά και τον τρόπο επιλογής των 10 όρων και των αποτελεσμάτων που συγκέντρωσε η κάθε μηχανή, στα οποία στηρίχθηκε η έρευνα. Στο επίπεδο 2, με τίτλο “Ανάκτηση και Προεπεξεργασία”

πραγματοποιήθηκε η συλλογή των δεδομένων, όπως ανακτήθηκαν από το Checkbot. Τέλος, στο επίπεδο 3 με τίτλο “Μέθοδοι Ανάλυσης” πραγματοποιήθηκε εκτενής ανάλυση των δεδομένων, η οποία εκφράστηκε με μία σειρά από διαγράμματα.

3.1 Συλλογή και Οργάνωση Δεδομένων

Στο πλαίσιο της εργασίας μας, αναζητήσαμε μέσω του Google Trends τους 10 δημοφιλέστερους όρους, που επισκέφτηκαν οι χρήστες παγκοσμίως το έτος 2021, ώστε η επιλογή των όρων να βασίζεται σε γενικά και αντικειμενικά κριτήρια και όχι σε προσωπικές προτιμήσεις. Στον παρακάτω πίνακα, παρατίθενται, οι εν λόγω όροι:

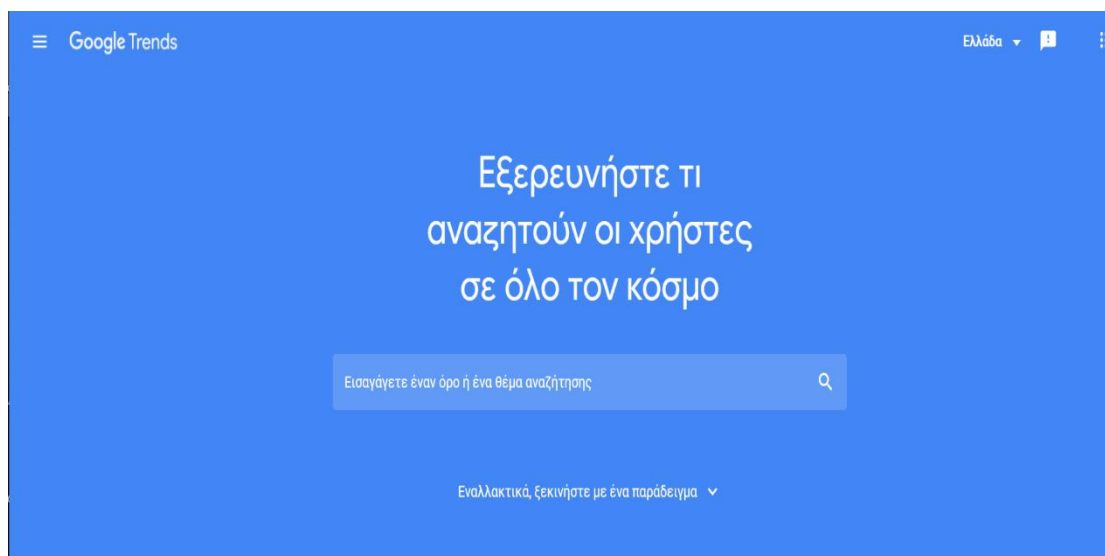
Top 10 trending searches in 2021
1) Australia vs India
2) India vs England
3) IPL
4) NBA
5) Euro 2021
6) Copa América
7) India vs New Zealand
8) T20 World Cup
9) Squid Game
10) DMX

Πίνακας 1. Οι 10 δημοφιλέστεροι όροι για την μηχανή αναζήτησης της Google (2021)

Το Google Trends είναι μία ιστοσελίδα της μηχανής αναζήτησης Google, η οποία συγκεντρώνει και αναλύει τους δημοφιλέστερους όρους που αναζητήθηκαν από τους χρήστες. Το εν λόγω εργαλείο, μπορεί να ενημερώσει τους χρήστες για την θέση του εκάστοτε όρου στις τάσεις αναζητήσεις, αν δηλαδή βρίσκεται σε ακμή- υψηλή θέση ή παρακμή- χαμηλή. Για να επιτευχθεί η παραπάνω πρόταση, το Google Trends χρησιμοποιεί τον όγκο των αναζητήσεων, δηλαδή την δραστηριότητα του κάθε όρου σε βάθος χρόνου, αλλά και σχετικές με τον όρο, που αναζητείται, λέξεις. Συμπληρωματικά, δίνει στους χρήστες τη δυνατότητα να κάνουν φιλτράρισμα στην αναζήτηση κάθε όρου, επιλέγοντας συγκεκριμένη ημερομηνία αλλά και γεωγραφική περιοχή. Το χρηστικό περιβάλλον και η αξιοπιστία του, το καθιστούν σημαντικό εργαλείο των ειδικών του SEO. Σημειώνεται ότι οι όροι που

αναφέρονται στον πίνακα 1, εμφανίζονται στο σύνδεσμο του Trends που ακολουθεί:
<https://trends.google.com/trends/yis/2021/GLOBAL/> [Ημερομηνία Πρόσβασης: 26/06/2023]

Παρακάτω ακολουθεί η παρουσίαση του περιβάλλοντος Google Trends.



Εικόνα 2. Περιβάλλον Google Trends

Ακολουθούν οι θεματικές λίστες του Google Trends. Η πρώτη από αυτές, με τίτλο Searches, είναι η λίστα που χρησιμοποιήθηκε στην συγκεκριμένη έρευνα.

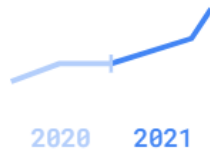
Searches	News	Actors
1 Australia vs India	1 Afghanistan	1 Alec Baldwin
2 India vs England	2 AMC Stock	2 Pete Davidson
3 IPL	3 COVID Vaccine	3 Aryan Khan
4 NBA	4 Dogecoin	4 Gina Carano
5 Euro 2021	5 GME Stock	5 Armie Hammer
6 Copa América	6 Stimulus Check	6 Carmen Salinas
7 India vs New Zealand	7 Georgia Senate Race	7 Shehnaaz Gill
8 T20 World Cup	8 Hurricane Ida	8 Elliot Page
9 Squid Game	9 COVID	9 Dave Chappelle

Athletes	Foods	Games
1 Christian Eriksen	1 Birria tacos	1 PopCat
2 Tiger Woods	2 Nasi goreng	2 FIFA 22
3 Simone Biles	3 Feta pasta	3 Battlefield 2042
4 Emma Raducanu	4 Charcuterie board	4 モンハンライズ (Monster Hunter Rise)
5 Henry Ruggs III	5 生姜焼き (Shōgayaki)	5 Resident Evil Village
6 Neeraj Chopra	6 Potato Soup	6 Genshin Impact
7 大谷翔平 (Shohei Ohtani)	7 ぶりの照り焼き (Teriyaki Amberjack)	7 Call of Duty: Vanguard
8 Odell Beckham Jr.	8 豚汁 (Tonjiru)	8 Far Cry 6
9 Rafael Nadal	9 Baked oats	9 Madden NFL 22
10 Tyson Fury	10 Overnight oats	10 Metroid Dread

Songs	Sports Teams	TV Shows
1 drivers license, Olivia Rodrigo	1 Real Madrid CF	1 Squid Game
2 MONTERO (Call Me by Your Name), Lil Nas X	2 Chelsea F.C.	2 Bridgerton
3 INDUSTRY BABY, Lil Nas X feat. Jack Harlow	3 Paris Saint-Germain F.C.	3 WandaVision
4 Fancy Like, Walker Hayes	4 FC Barcelona	4 Cobra Kai
5 MAPA, SB19	5 Sociedade Esportiva Palmeiras	5 Loki
6 good 4 u, Olivia Rodrigo	6 Los Angeles Lakers	6 Sweet Tooth
7 Butter, BTS	7 Sport Club Corinthians Paulista	7 Lupin
	8 Boston Red Sox	8 Ginny and Georgia
	9 São Paulo FC	9 True Beauty
	10 Manchester City F.C.	10 BBB21

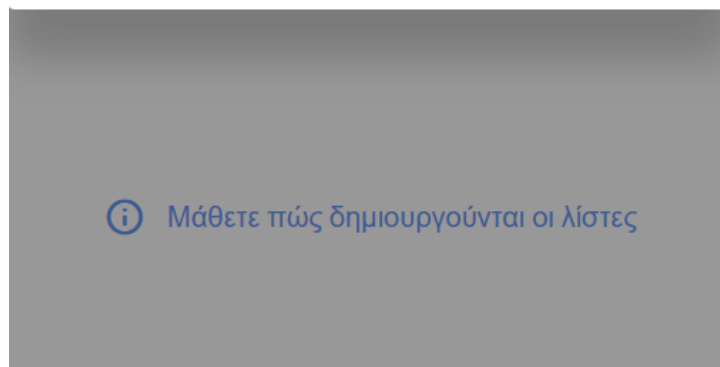
Movies	Passings	People
1 Eternals	1 DMX	1 Alec Baldwin
2 Black Widow	2 Gabby Petito	2 Kyle Rittenhouse
3 Dune	3 Marília Mendonça	3 Christian Eriksen
4 Shang-Chi and the Legend of the Ten Rings	4 Prince Philip	4 Tiger Woods
5 Red Notice	5 Brian Laundrie	5 Simone Biles
6 Mortal Kombat	6 MC Kevin	6 Joe Biden
7 Cruella	7 Lázaro Barbosa de Sousa	7 Emma Raducanu
8 Halloween Kills	8 Peter R. de Vries	8 Derek Chauvin
9 Godzilla vs. Kong	9 Paulo Gustavo	9 Henry Ruggs III
10 Army of the Dead	10 Helen McCrory	10 Pete Davidson

Ο τρόπος επιλογής των όρων και δημιουργίας των λιστών στο Google Trends απεικονίζεται στη παρακάτω εικόνα. Η αιτιολόγηση παρέχεται στον χρήστη δικαιολογώντας τις επιλογές της ιστοσελίδας.



Οι λίστες βασίζονται σε όρους αναζήτησης που παρουσίασαν τη μεγαλύτερη κορύφωση φέτος σε σχέση με το προηγούμενο έτος.

ΤΟ ΚΑΤΑΛΑΒΑ

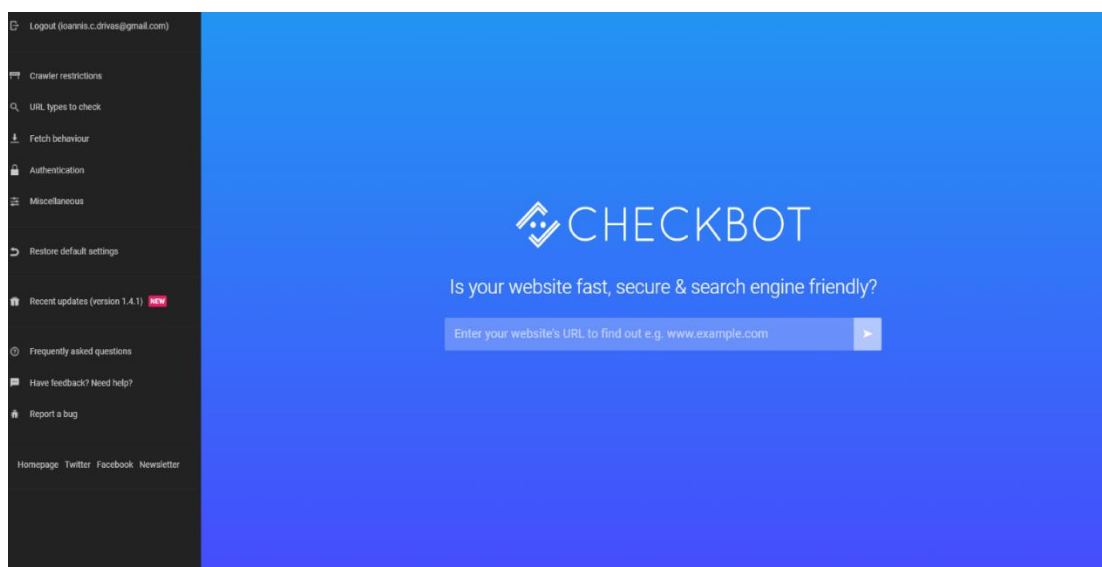


<https://trends.google.com/trends/vis/2021/GLOBAL/>

Στη συνέχεια, αναζητήσαμε τους παραπάνω όρους στις μηχανές αναζήτησης, Google και Bing και αποθηκεύσαμε για κάθε όρο, τα 10 πρώτα αποτελέσματα- URLs σε 2 λογιστικά φύλλα (1 για κάθε μηχανή αναζήτηση που χρησιμοποιήθηκε), τα οποία αρχικά είχαν δύο στήλες, Search Term & Domain. Η πρώτη περιείχε τους 10 όρους ενώ η δεύτερη είχε τα URLs. Συγκεντρώσαμε 100 αποτελέσματα από τη Google και 100 από τη Bing. Έπειτα, από τα 200 συνολικά αποτελέσματα αποφασίσαμε, να αξιοποιήσουμε τα 5 πρώτα URL κάθε όρου, ώστε να προχωρήσουμε την έρευνα μας, με τα πιο σχετικά αποτελέσματα. Στα εναπομείναντα 100 αποτελέσματα έγινε περαιτέρω ανάλυση μέσω του εργαλείου Checkbot. Η εξέλιξη αυτή, οδήγησε στη δημιουργία άλλων δύο υπολογιστικών φύλλων στο excel. Τα συγκεκριμένα φύλλα έχουν 3 στήλες (Search Term, URL, Website Score), στις οποίες καταχωρήθηκαν οι όροι, τα 5 URLs του κάθε όρου και η βαθμολογία που συγκέντρωσε στο Checkbot.

Το Checkbot είναι ένα χρήσιμο εργαλείο που χρησιμοποιείται για την βελτιστοποίηση της ταχύτητας και της ασφάλειας του εκάστοτε website. Ελέγχει μια πληθώρα παραμέτρων και αξιολογεί τον ιστότοπο και τον βαθμολογεί. Κάποιες από αυτές είναι η ασφάλεια, οι H1 επικεφαλίδες, η ταχύτητα φόρτωσης ιστοσελίδας, οι ανενεργοί σύνδεσμοι, το μήκος του URL,

η αντιγραφή περιεχομένου (Duplicate content), η βελτιστοποίηση για κινητές συσκευές, ο αριθμός ιστοσελίδων που συνδέονται κ.α. Το συγκεκριμένο εργαλείο υπάρχει, ως επέκταση στο Chrome Web Store.



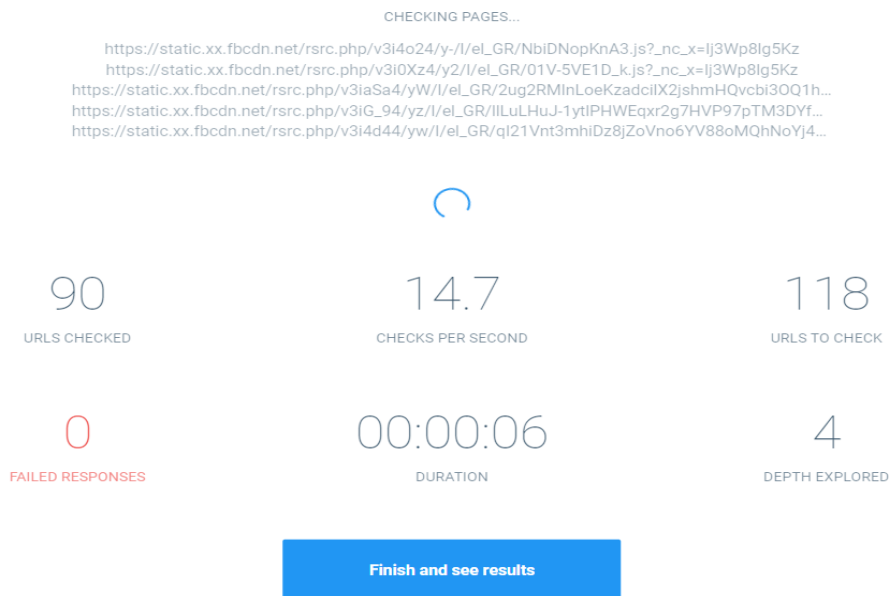
Εικόνα 3. Περιβάλλον Checkbot

Τα 100 URLs αναλύθηκαν από την εν λόγω επέκταση. Τα αποτελέσματα κάθε συνδέσμου αποθηκεύτηκαν σε μορφότυπο PDF. Ύστερα, αποθηκεύτηκαν εκ νέου σε 2 λογιστικά φύλλα, ώστε να είναι πιο εύκολος ο διαχωρισμός και μετέπειτα η σύγκριση Google & Bing. Εν συνεχεία, διαλέξαμε τις 4 πιο σημαντικές μετρικές του Checkbot. Κάθε μία μετρική απαρτίζεται από άλλες επιμέρους υπό-μετρικές οι οποίες συσσωρευτικά καταγράφουν σε εκατοσταβάθμια κλίμακα την τελική βαθμολογία για κάθε μία από τις 4 γενικές μετρικές:

- **SEO:** Καταγράφει την συνολική απόδοση των ιστοσελίδων που εξετάστηκαν ως προς τα τεχνικά χαρακτηριστικά της ιστοσελίδας που σχετίζονται με το βαθμό προσπελασιμότητας της από τις μηχανές αναζήτησης
- **Speed:** Καταγράφει την συνολική απόδοση των ιστοσελίδων που εξετάστηκαν ως προς τα τεχνικά χαρακτηριστικά που επηρεάζουν την ταχύτητα φόρτωσης του περιεχόμενου
- **Security:** Καταγράφει την συνολική απόδοση των ιστοσελίδων που εξετάστηκαν ως προς τα τεχνικά χαρακτηριστικά που επηρεάζουν την ασφάλεια πλοήγησης των χρηστών και την αποφυγή κακόβουλου λογισμικού (malware)
- **Total Website Score** λαμβάνοντας υπόψη την έρευνα των Drivas et al. (2021) η παρούσα μετρική αναπτύσσεται μέσω του συγκερασμού των τριών διαφορετικών μετρικών (SEO, Speed, Security).

και σημειώσαμε τους μέσους όρους των 5 πρώτων συνδέσμων που συγκέντρωσαν οι 10 όροι και δημιουργήσαμε δύο πίνακες.

Τα παρακάτω στιγμιότυπα παρουσιάζουν το περιβάλλον του Checkbot κατά τη χρήση του.



Εικόνα 4. Περιβάλλον ανάλυσης URL Checkbot

Στην παραπάνω εικόνα παρατηρούμε την ανάλυση του κάθε συνδέσμου από το checkbot ελέγχοντας μία σειρά από παραμέτρους.

Website audit summary for www.facebook.com

Website score










Category scores

SEO	80%
Speed	97%
Security	84%

URLs checked

Website checked on Tue, 25 Oct 2022 16:42:06 GMT

 Crawled URLs	14	<div style="width: 100%;"><div style="width: 100%;"></div></div>
 Pages	2	<div style="width: 20%;"><div style="width: 20%;"></div></div>
 Internal URLs	4	<div style="width: 40%;"><div style="width: 40%;"></div></div>
 External URLs	10	<div style="width: 75%;"><div style="width: 75%;"></div></div>
 Images	0	<div style="width: 0%;"><div style="width: 0%;"></div></div>
 CSS	1	<div style="width: 10%;"><div style="width: 10%;"></div></div>
 JavaScript	9	<div style="width: 75%;"><div style="width: 75%;"></div></div>

Εικόνα 5. Περιβάλλον αποτελεσμάτων ανάλυσης Checkbot

Στο τέλος της εξέτασης κάθε συνδέσμου το Checkbot εξάγει σε μορφή PDF τα αποτελέσματα των 4 βασικών μετρικών (Website score, SEO, Speed, Security) καθώς και των υπό-μετρικών που περιλαμβάνει η κάθε μετρική.

SEO results

80%

Page titles

✓ Set page titles	2 out of 2	<div style="width: 100%;"><div style="width: 100%;"></div></div>	100%
✗ Use optimal length titles	0 out of 2	<div style="width: 0%;"><div style="width: 0%;"></div></div>	0%
✓ Use unique titles	1 out of 1	<div style="width: 100%;"><div style="width: 100%;"></div></div>	100%

Page headings

✗ Set H1 headings	0 out of 2	<div style="width: 0%;"><div style="width: 0%;"></div></div>	0%
✗ Use one H1 heading per page	0 out of 2	<div style="width: 0%;"><div style="width: 0%;"></div></div>	0%
✗ Use optimal length H1 headings	0 out of 2	<div style="width: 0%;"><div style="width: 0%;"></div></div>	0%
✓ Use unique H1 headings	1 out of 1	<div style="width: 100%;"><div style="width: 100%;"></div></div>	100%

Page descriptions

✓ Set page descriptions	2 out of 2	<div style="width: 100%;"><div style="width: 100%;"></div></div>	100%
✗ Use optimal length descriptions	0 out of 2	<div style="width: 0%;"><div style="width: 0%;"></div></div>	0%
✓ Use unique descriptions	1 out of 1	<div style="width: 100%;"><div style="width: 100%;"></div></div>	100%

Duplicate content





✓ Set canonical URLs	2 out of 2	<div style="width: 100%;"><div style="width: 100%;"></div></div>	100%
✓ Avoid duplicate page content	1 out of 1	<div style="width: 100%;"><div style="width: 100%;"></div></div>	100%

Εικόνα 6. Περιβάλλον αναλυτικών αποτελεσμάτων SEO Checkbot

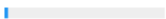

Speed results

97%

Page size

✓ Use compression	13 out of 13		100%
✓ Avoid recompressing data			100%
✓ Use minification	10 out of 10		100%
✓ Avoid inline source maps	10 out of 10		100%




Javascript

✗ Avoid excessive inline JavaScript	0 out of 2		0%
✓ Avoid render-blocking JavaScript	2 out of 2		100%

CSS

✓ Avoid excessive inline CSS	2 out of 2		100%
✓ Avoid CSS @import	1 out of 1		100%

Redirects

✓ Avoid internal link redirects	2 out of 2		100%
✓ Avoid resource redirects	2 out of 2		100%
✓ Avoid redirect chains			100%

Caching

✓ Use caching	10 out of 10		100%
✓ Use long caching times	10 out of 10		100%
✓ Avoid duplicate resources	10 out of 10		100%

Εικόνα 7. Περιβάλλον αναλυτικών αποτελεσμάτων ταχύτητας (Speed) Checkbot

Security results

84%

HTTPS

✓ Use HTTPS	2 out of 2	<div style="width: 100%;"></div>	100%
✓ Avoid mixed content	2 out of 2	<div style="width: 100%;"></div>	100%
✓ Use secure password forms		<div style="width: 100%;"></div>	100%

HSTS

✓ Use HSTS	2 out of 2	<div style="width: 100%;"></div>	100%
✗ Use HSTS preload	0 out of 2	<div style="width: 0%;"></div>	0%

Content sniffing

✓ Set MIME types	4 out of 4	<div style="width: 100%;"></div>	100%
✓ Use content sniffing protection	4 out of 4	<div style="width: 100%;"></div>	100%

Miscellaneous

✓ Use clickjack protection	2 out of 2	<div style="width: 100%;"></div>	100%
✗ Use XSS protection	0 out of 2	<div style="width: 0%;"></div>	0%
✓ Hide server version data	4 out of 4	<div style="width: 100%;"></div>	100%

Εικόνα 8. Περιβάλλον αναλυτικών αποτελεσμάτων ασφάλειας (Security) Checkbot

3.1.1 Επεξήγηση μετρικών απόδοσης ιστοσελίδων

Παρακάτω ακολουθούν οι πίνακες με τις υπομετρικές της SEO, Speed, Security, οι οποίες επιλέχθηκαν από τις προτάσεις της [Google](#) και [Bing](#), ως οι σημαντικότερες για την βελτίωση της ποιότητας και κατά συνέπεια της θέσης κατάταξης της εκάστοτε ιστοσελίδας.

SEO

Metric	Περιγραφή
--------	-----------

Use unique titles	Κάθε τίτλος θα πρέπει να είναι μοναδικός και να μην χρησιμοποιείται πουθενά αλλού σε άλλες ιστοσελίδες, αποτρέποντας τις δυσκολίες των χρηστών να καταλάβουν ποια σελίδα να επιλέξουν μια ιστοσελίδα από τα SERPs και των crawlers να ευρετηριάσουν σωστά την ιστοσελίδα.
Set H1 headings	Κάθε σελίδα θα πρέπει να έχει μια περιγραφική επικεφαλίδα H1 για να βοηθήσει τις μηχανές αναζήτησης και τους χρήστες να καταλάβουν τι περιέχει η συγκεκριμένη σελίδα. Η πιο σημαντική και υψηλότερης κατάταξης επικεφαλίδα στην HTML ονομάζεται επικεφαλίδα H1. Επιπλέον, υπάρχει μια σειρά επιπέδων σπουδαιότητας από τις επικεφαλίδες H2, H3, H4, H5 και H6.
Use one H1 per page	Πολλές επικεφαλίδες H1 μπορεί να προκαλέσουν σύγχυση τόσο στους χρήστες κατά την ανάγνωση μιας ιστοσελίδας όσο και στη διαδικασία ανίχνευσης από τις μηχανές αναζήτησης. Έτσι, συνιστάται από τις μηχανές αναζήτησης να χρησιμοποιείται μόνο ένας τίτλος H1 ανά ιστοσελίδα.
Use optimal length H1 headings	Κάθε σελίδα θα πρέπει να έχει μια επικεφαλίδα H1 που δεν θα είναι μεγάλη, ώστε οι χρήστες και οι μηχανές αναζήτησης να κατανοούν εύκολα το θέμα της σελίδας. Η Google συνιστά να αποφεύγονται οι υπερβολικά μεγάλες επικεφαλίδες, ώστε οι αναγνώστες να μπορούν να "σαρώσουν" γρήγορα. Η σύσταση είναι να μην υπερβαίνει τους 70 χαρακτήρες.
Use unique H1 headings	Κάθε ιστοσελίδα θα πρέπει να έχει έναν τίτλο H1 που να είναι μοναδικός από όλες τις άλλες σελίδες του ιστότοπου, ώστε να αποφεύγονται προβλήματα διπλού περιεχομένου.
Set page descriptions	Για κάθε ιστοσελίδα, θα πρέπει να υπάρχει μια μετα-περιγραφή για να βοηθήσει τόσο τους ανιχνευτές να ευρετηριάσουν το περιεχόμενο της σελίδας όσο και τους αναγνώστες να καταλάβουν αν το περιεχόμενο της σελίδας ταιριάζει με τους όρους αναζήτησής τους όταν η ιστοσελίδα εμφανίζεται στα SERPs..

Use optimal length descriptions	Τόσο η Google όσο και η Bing συνιστούν ένα εύρος από 100 έως 320 χαρακτήρες. Διαφορετικά, οι εκτεταμένοι χαρακτήρες δεν θα εμφανίζονται στα αποτελέσματα αναζήτησης, γεγονός που πιθανώς αποτρέπει τους χρήστες από το να κάνουν κλικ στην ιστοσελίδα.
Set canonical URLs	Όταν μια συλλογή διευθύνσεων URL στον ιστότοπο επιστρέφει διπλό ή σχεδόν διπλό περιεχόμενο, οι μηχανές αναζήτησης θα επιλέξουν μια μοναδική οριστική διεύθυνση URL για αυτό το περιεχόμενο, γνωστή ως κανονική διεύθυνση URL. Αυτή η διεύθυνση URL θα ανιχνεύεται συχνότερα και θα δίνεται προτεραιότητα στα αποτελέσματα αναζήτησης έναντι των διευθύνσεων URL με διπλό περιεχόμενο, ενώ οι backlinks προς τις διευθύνσεις URL με διπλό περιεχόμενο θα ερμηνεύονται ως σύνδεσμοι προς την κανονική διεύθυνση URL. Ως εκ τούτου, είναι σημαντικό για τους διαχειριστές ιστοτόπων να καθορίζουν έγκυρες κανονικές διευθύνσεις URL.
Use short URLs	Οι σύντομες διευθύνσεις URL βοηθούν τις μηχανές αναζήτησης να ανιχνεύσουν ταχύτερα τον ιστότοπο, ενώ οι χρήστες μπορούν να διαβάσουν ταχύτερα μια σύντομη διαδρομή URL στη γραμμή του προγράμματος περιήγησης παρά μια μακρά. Το εργαλείο CheckBot συνιστά μια διεύθυνση URL έως 100 χαρακτήρες.
Avoid duplicate page content	Το περιεχόμενο [ή το κείμενο] πρέπει να είναι μοναδικό σε κάθε ιστοσελίδα. Αυτό συμβαίνει επειδή οι μηχανές αναζήτησης δεν ευρετηριάζουν το ίδιο περιεχόμενο ακόμη και αν έχει εμφανιστεί σε διαφορετικές διαδρομές URL.
Set image alt texts	Κάθε εικόνα που περιλαμβάνεται σε μια σελίδα με ετικέτες θα πρέπει να έχει ακριβή περιγραφή με τη χρήση των χαρακτηριστικών alt. Αυτό βελτιώνει καταρχάς την προσβασιμότητα, καθώς οι υποστηρικτικές τεχνολογίες, όπως οι βοηθητικοί αναγνώστες, βασίζονται στα alt attributes. Επίσης, οι ετικέτες εικόνας βοηθούν τους ανιχνευτές να

	ευρετηριάζουν τα στοιχεία μιας σελίδας που δεν είναι σε μορφή κειμένου.
Set mobile scaling	Αυτή η μέτρηση εκτιμά πόσο φιλική είναι μια ιστοσελίδα για να εμφανιστεί σε προγράμματα περιήγησης για κινητά τηλέφωνα. Αντιφατικά, όσο περισσότερο χρειάζεται ο χρήστης να κάνει zoom-in και zoom-out από την οθόνη του κινητού του τηλεφώνου, τόσο λιγότερο συμβατός είναι ο ιστότοπος με τις ιδιότητες κλιμάκωσης για κινητά τηλέφωνα.
Avoid deeply nested URLs	Οι απλές δομές στις διευθύνσεις URL ευρετηριάζονται γρηγορότερα από τους ανιχνευτές, ελαχιστοποιώντας με αυτόν τον τρόπο τον αριθμό των υποφακέλων που πρέπει να αναλύσουν. Ένας γενικός κανόνας που προέρχεται από την τεκμηρίωση του CheckBot είναι όχι περισσότεροι από πέντε υποφάκελοι.

Speed

Metric	Description
Avoid resources redirects	Θα πρέπει να αποφεύγεται η φόρτωση πόρων ιστοσελίδας μέσω διευθύνσεων URL που εκτελούν ανακατευθύνσεις, επειδή οι ανακατευθύνσεις επιβραδύνουν το χρόνο φόρτωσης της ιστοσελίδας. Πιο συγκεκριμένα, εάν ένα αρχείο CSS φορτωνόταν χρησιμοποιώντας τη διεύθυνση URL <code>example.com/styles.css</code> , η οποία στη συνέχεια ανακατευθύνεται στη διεύθυνση <code>www.example.com/styles.css</code> , η ανακατεύθυνση θα είχε αντίκτυπο στο αρχείο που πρέπει να ανακτηθεί και ως εκ τούτου θα χρειαζόταν περισσότερο χρόνο. Αυτό θα μπορούσε να διορθωθεί με απευθείας σύνδεση στον προορισμό ανακατεύθυνσης.
Avoid internal links redirects	Είναι ζωτικής σημασίας για την αύξηση της ταχύτητας η αποφυγή υπερσυνδέσμων προς διευθύνσεις URL που εκτελούν ανακατευθύνσεις για την αύξηση της ταχύτητας πλοήγησης μεταξύ των ιστοσελίδων ενός ιστότοπου. Για παράδειγμα, αν υποθέσουμε ότι είχαμε έναν υπερσύνδεσμο που έδειχνε στο

	<p>/news που τώρα ανακατευθύνεται στο /our-news λόγω της μετεγκατάστασης της ιστοσελίδας. Καθώς ακολουθείται η ανακατεύθυνση, ένας χρήστης που επισκέπτεται τον υπερσύνδεσμο θα βιώσει μια σημαντική καθυστέρηση φόρτωσης της ιστοσελίδας.</p>
Use compression	<p>Για να μειωθούν οι χρόνοι μεταφοράς, απαιτείται διαμόρφωση του διακομιστή για την αποστολή δεδομένων σε συμπιεσμένη μορφή. Η συμπίεση μπορεί να μειώσει τον όγκο των δεδομένων που πρέπει να αποσταλούν κατά περίπου 70% για συμπιεσμένα αρχεία με ελάχιστη προσπάθεια διαμόρφωσης. Η HTML, το CSS, η JavaScript, το απλό κείμενο, η XML, το JSON και το SVG είναι όλα μορφές δεδομένων που βασίζονται σε κείμενο και μπορούν να σταλούν με ενεργοποιημένη συμπίεση.</p>
Use minification	<p>Αυτό βοηθά στην ελαχιστοποίηση τόσο των αρχείων CSS όσο και της JavaScript για τη μείωση του βάρους της ιστοσελίδας. Με απλότητα, η διαδικασία ελαχιστοποίησης διατηρεί τη μορφολογία και τη συμπεριφορά του κώδικα- ωστόσο, μειώνει το μέγεθός του αφαιρώντας τα σχόλια των προγραμματιστών και το κενό διάστημα τόσο στα αρχεία JavaScript όσο και στα αρχεία CSS.</p>
Avoid render-blocking JavaScript	<p>Η εξωτερική JavaScript θα πρέπει να χρησιμοποιείται στις ιστοσελίδες με τρόπο που να μην παρεμβαίνει στην απόδοση της σελίδας. Μια ετικέτα <code>script src="..."</code> θα αποτρέψει την απόδοση της HTML μέχρι να ανακτηθεί το καθορισμένο αρχείο JavaScript καθώς και μέχρι να ολοκληρωθεί η εκτέλεση του περιεχομένου του αρχείου.</p>
Use valid HTML	<p>Τα σφάλματα εγκυρότητας HTML θα πρέπει να αποφεύγονται στις ιστοσελίδες. Η μη έγκυρη HTML μπορεί να προκαλέσει προβλήματα εμφάνισης στους χρήστες, επειδή οι ιστοσελίδες μπορεί να μην εμφανίζονται όπως σκοπεύουν να εμφανίσουν οι διαχειριστές των ιστοσελίδων. Επιπλέον, επειδή οι φυλλομετρητές αντιμετωπίζουν διαφορετικά τον άκυρο κώδικα και ο χρόνος φόρτωσης διαφέρει για κάθε επικύρωση, οι</p>

	<p>διαχειριστές θα πρέπει να χρησιμοποιούν πάντα έγκυρη HTML για να αποφεύγονται τα προβλήματα που σχετίζονται με τους φυλλομετρητές. Ομοίως, οι μηχανές αναζήτησης που προσπαθούν να ερμηνεύσουν την άκυρη HTML ενδέχεται να παρερμηνεύσουν το περιεχόμενο μιας ιστοσελίδας.</p>
Avoid excessive inline JavaScript	<p>Οι διαχειριστές των ιστοτόπων πρέπει να αποφεύγουν την παγίδα της inline JavaScript. Ο κώδικας JavaScript μπορεί να ενσωματωθεί απευθείας στις σελίδες χρησιμοποιώντας την ετικέτα <script>. Φυσικά, αυτό είναι τουλάχιστον βολικό κατά τη διάρκεια της ανάπτυξης ιστοτόπων, αλλά έχει δύο μειονεκτήματα. Πρώτον, μπλοκάρει την απόδοση της HTML μέχρι να αναλυθεί και να εκτελεστεί ο κώδικας JavaScript, και δεύτερον, δεν αποθηκεύεται στην προσωρινή μνήμη εάν ο κώδικας μοιράζεται μεταξύ σελίδων.</p>
Avoid excessive inline CSS	<p>Ομοίως με την inline JavaScript, η αποφυγή μεγάλου όγκου inline CSS είναι προτιμότερη για τη βελτίωση της αποτελεσματικότητας της προσωρινής αποθήκευσης.</p>

Security

Metric	Description
Use content sniffing protection	<p>Τα προγράμματα περιήγησης μερικές φορές αγνοούν τον τύπο πολυμέσων (ή τύπο MIME) στην επικεφαλίδα Content-Type κατά την επεξεργασία μιας απόκρισης και προσπαθούν να μαντέψουν τον τύπο με βάση το περιεχόμενο της απόκρισης. Αυτό είναι γνωστό ως content sniffing και χρησιμοποιείται για τη βελτίωση της εμπειρίας του χρήστη όταν οι επικεφαλίδες Content-Type είναι λανθασμένες ή λείπουν. Αυτή η συμπεριφορά, ωστόσο, μπορεί να γίνει αντικείμενο εκμετάλλευσης. Για παράδειγμα, εάν ο ιστοτόπος επιτρέπει στους χρήστες να μοιράζονται εικόνες, κακόβουλο λογισμικό θα μπορούσε να ανεβάσει ένα ειδικά διαμορφωμένο αρχείο εικόνας που περιέχει κώδικα JavaScript. Το κακόβουλο αρχείο θα μπορούσε στη συνέχεια να εκτελεστεί από ένα πρόγραμμα περιήγησης που εκτελεί sniffing περιεχομένου.</p>

Use clickjack protection	<p>Το κακόβουλο λογισμικό χρησιμοποιεί clickjacking για να εμφανίσει ιστοσελίδες μέσα σε ένα iframe σε έναν ιστότοπο που ελέγχει. Ένα είδος κακόβουλου λογισμικού, για παράδειγμα, θα μπορούσε να τοποθετήσει ένα στρώμα UI πάνω από ένα iframe για να εξαπατήσει τους επισκέπτες να εκτελέσουν ενέργειες στην ιστοσελίδα και να τοποθετήσει πεδία φόρμας πάνω από αυτά για να κλέψει προσωπικά δεδομένα. Επιπλέον, το clickjacking έχει χρησιμοποιηθεί για να εξαπατήσει τους χρήστες ώστε να μοιραστούν συνδέσμους στα μέσα κοινωνικής δικτύωσης, να κάνουν κλικ σε διαφημίσεις και να κλέψουν κωδικούς πρόσβασης.</p>
Use HTTPS	<p>Συνιστάται ανεπιφύλακτα από τα περισσότερα προγράμματα περιήγησης να χρησιμοποιείτε HTTPS σε ιστότοπους. Το HTTPS κρυπτογραφεί τα δεδομένα που αποστέλλονται μεταξύ του ιστότοπου και των προγραμμάτων περιήγησης, καθιστώντας αδύνατη την ανάγνωση ή την αλλοίωση του περιεχομένου των δεδομένων από κακόβουλο λογισμικό.</p>
Hide server version data	<p>Συνιστάται η αφαίρεση των επικεφαλίδων απόκρισης που υποδεικνύουν ποιο λογισμικό εκτελείται στον διακομιστή ιστού για να αποφύγετε τη διαρροή πληροφοριών σε κακόβουλο λογισμικό. Το όνομα και η έκδοση του λογισμικού που χρησιμοποιείται για την απάντηση σε ένα αίτημα παρέχονται στις επικεφαλίδες απάντησης πολλών διακομιστών ιστού. Αυτό μπορεί να δώσει πληροφορίες στο κακόβουλο λογισμικό σχετικά με το πού να αναζητήσει ευπάθειες στην πλευρά του διακομιστή.</p>
Use HSTS	<p>Η ρύθμιση των κεφαλίδων απόκρισης HSTS (HTTP Strict Transport Security) που αναγκάζουν τα προγράμματα περιήγησης να χρησιμοποιούν πάντα HTTPS σε έναν ιστότοπο αντί για HTTP είναι ένα σημαντικό χαρακτηριστικό ασφαλείας στο πυροβολικό ασφαλείας του ιστότοπου. Όταν ένα πρόγραμμα περιήγησης επισκέπτεται την ιστοσελίδα και βλέπει την κεφαλίδα απόκρισης Strict-Transport-Security, το πρόγραμμα περιήγησης λαμβάνει εντολή να μετατρέψει όλα τα μελλοντικά αιτήματα URL http:// προς την ιστοσελίδα σε αιτήματα https://.</p>

Use protection	XSS	<p>Στις επιθέσεις XSS, ένα είδος κακόβουλου λογισμικού εισάγει κώδικα σε μια σελίδα που αποστέλλεται στους χρήστες. Ορισμένα προγράμματα περιήγησης περιλαμβάνουν προστασία XSS. Αυτή η ρύθμιση θα σταματήσει την απόδοση της ιστοσελίδας εάν το πρόγραμμα περιήγησης εντοπίσει επίθεση XSS.</p>
Avoid Content	Mixed	<p>Στο σύνολο της κάθε ιστοσελίδας και οι συνδέσμοι που περιλαμβάνει οφείλουν να κατέχουν το πιστοποιητικό HTTPS. Σε συγκεκριμένες περιπτώσεις μπορούν να φορτωθούν css ή javascript αρχεία τα οποία μπορεί να βελτιώνουν την ευχρηστία της ιστοσελίδας, ωστόσο αυτά δεν κατέχουν το εν λόγω πιστοποιητικό, σημείο που μπορεί να βλάψει την ασφάλεια της ιστοσελίδας αφήνοντας κακόβουλο λογισμικό να στείλει πολλά αιτήματα (requests) για την φόρτωση των συγκεκριμένων αρχείων υπερφορτώνοντας έτσι τον διακομιστή (server).</p>

3.2 Μέθοδοι & Εργαλεία Ανάλυσης Δεδομένων

Το JASP (Jeffreys's Amazing Statistics Program) είναι ένα δωρεάν πρόγραμμα στατιστικής ανάλυσης ανοικτού κώδικα που υποστηρίζεται από το Πανεπιστήμιο του Άμστερνταμ. Έχει σχεδιαστεί για να είναι εύκολο στη χρήση και οικείο στους χρήστες του SPSS ή άλλων παρόμοιων προγραμμάτων ανάλυσης δεδομένων. Προσφέρει τυποποιημένες διαδικασίες ανάλυσης τόσο στην κλασική όσο και στην Μπεϋζιανή μορφή τους. Το JASP παράγει γενικά πίνακες αποτελεσμάτων και διαγράμματα τύπου APA για να διευκολύνει τη δημοσίευση. Προωθεί την ανοικτή επιστήμη με την ενσωμάτωση των ρυθμίσεων ανάλυσης στα αποτελέσματα. Η ανάπτυξη του JASP υποστηρίζεται οικονομικά από διάφορα πανεπιστήμια και ερευνητικά ταμεία. Παρακάτω ακολουθούν δύο εικόνες από το περιβάλλον εργασίας του JASP. Στην πρώτη περίπτωση εμφανίζεται το σύνολο των δεδομένων προς επεξεργασία, ενώ στη δεύτερη η ανάλυση τους. (Wikipedia contributors, 2023)

Query	Website Search Result	Total Website SEO Score	Use_unique_titles	Set_H1_headings
1	https://www.cricbuzz.com/cricket-series/4464/australia-tour-of-india-2022/matches	89	82	97
2	https://www.espn.cricinfo.com/series/australia-in-india-2022-1327497/india-vs-australia-1st-t20i-1327503/live-cricknet-score	69	93	54
3	https://www.icc-cricknet.com/match/101160	67	74	89
4	https://www.foxsports.com.au/cricket/australia-australian-cricknet-2022-3rd-t20i-australia-vs-india-in-hyderabad-how-to-watch-sco	58	81	76
5	https://en.wikipedia.org/wiki/Australian_cricknet_team_in_India_in_2022%E2%80%9C23	75	100	100
6	https://www.espn.cricinfo.com/series/india-in-england-2022-1276893	69	93	53
7	https://www.cricbuzz.com/cricket-series/3636/india-tour-of-england-2022/matches	89	82	97
8	https://www.business-standard.com/topic/india-vs-england	72	91	96
9	https://www.icc-cricknet.com/match/22456#scorecard	67	74	89
10	https://timesofindia.indiatimes.com/sports/cricket/india-in-england	68	96	92
11	https://www.skoutz.gr/c605/apoththikes-mihanes.html?keyword=ipl	77	77	99
12	http://newyorklaser.gr/laser_vs_id.html	51	0	100
13	https://www.beautyhomeclinic.gr/papitosoxi-laser-ipl/photolysis-ipl/	67	100	89
14	https://antiagers.gr/ipl/	66	96	42
15	https://www.bestprice.gr/call/6773/foxiysi-morini-apothikos.html?qi=ipl	82	74	97
16	https://www.nba.com/	75	86	56
17	https://www.sport24.gr/nba	78	99	99
18	https://www.espn.com/nba/	65	97	100
19	https://www.gazzetta.gr/basketball/nba	72	99	96
20	https://www.facebook.com/nba/	86	100	0
21	https://www.foxbet.gr/euro-2020	80	99	100
22	https://www.iefimerida.gr/euroeuro-2021-panorama-apoliteamata-bathmologies-programma	81	99	99
23	https://en.wikipedia.org/wiki/UEFA_Euro_2020	81	100	100
24	https://sportbet.gr/euro-2020-2021/	63	87	57
25	https://www.kathimerini.gr/work/561418156/euro-2021-epithesi-verothoy-kata-ueda-ga-to-schedon-gemato-goyempler	81	93	99
26	https://en.wikipedia.org/wiki/Copa_Am%C3%A9rica	75	100	100

Εικόνα 9. Περιβάλλον επεξεργασίας δεδομένων στην πλατφόρμα του JASP

Regression Use Unique Titles

Regression H1 Headings

Regression Use H1 heading per page

Results

Regression Use Unique Titles

Model Summary - Total Website SEO Score

Model	R	R ²	Adjusted R ²	RMSE
H ₀	0.000	0.000	0.000	8.533
H _a	0.433	0.187	0.170	7.773

ANOVA

Model	Sum of Squares	df	Mean Square	F	p	
H ₀	Regression	667.357	1	667.357	11.045	0.002
	Residual	2900.163	48	60.420		
	Total	3567.520	49			

Note: The intercept model is omitted, as no meaningful information can be shown.

Coefficients

Model	Unstandardized	Standard Error	Standardized	t	p	
H ₀	(Intercept)	73.640	1.207	61.026	<.001	
H _a	(Intercept)	57.366	5.019	11.431	<.001	
	Use unique titles	0.186	0.056	0.433	3.323	0.002

Regression H1 Headings

Model Summary - Total Website SEO Score

Model	R	R ²	Adjusted R ²	RMSE
H ₀	0.000	0.000	0.000	8.533
H _a	0.148	0.022	0.002	8.326

ANOVA

Εικόνα 10. Περιβάλλον ανάλυσης των ερευνητικών δεδομένων στο JASP.

3.2.1 Περιγραφική στατιστική ανάλυση

Αρχικά, ένας από τους πρώτους στατιστικούς ελέγχους που εφαρμόστηκαν ήταν η υπόθεση ότι τα ανακτηθέντα δεδομένα ακολουθούν κανονική κατανομή. Για να γίνει αυτό, ακολουθήσαμε τον κανόνα των Dziuban και Shirkey (1974) για τη μέτρηση της κανονικότητας των δεδομένων στο πρώτο 25% του ανακτηθέντος δείγματος. Οι πληροφορίες αυτές θα βοηθήσουν τα διαδοχικά βήματα ανάπτυξης αξιόπιστων και έγκυρων συστημάτων μέτρησης

επιδόσεων, συμπεριλαμβανομένων των συγκεκριμένων μετρήσεων και της ανάπτυξης μοντέλων παλινδρόμησης. Ακολουθήσαμε αυτή την πρόταση για τον έλεγχο της κανονικότητας των δεδομένων στο πρώτο 25% της ανάκτησης για κάθε πλατφόρμα όπου συλλέχθηκαν μετρήσεις ιστού.

Για την αξιολόγηση της κανονικότητας των δεδομένων, επιλέχθηκε ο έλεγχος Shapiro-Wilk ως ένας από τους πιο ισχυρούς δείκτες ελέγχου κανονικότητας για την ανίχνευση της μη κανονικότητας σε σύγκριση με άλλους ελέγχους, όπως οι Kolmogorov-Smirnov, Anderson-Darling και Lilliefors (Mishra et al., 2019- Razali and Wah, 2011). Για κάθε μετρική που εμπλέκεται στο τμήμα αποτελεσμάτων, η τιμή Shapiro-Wilk λαμβάνει χώρα στους πίνακες περιγραφικής στατιστικής. Όσο πιο κοντά είναι η τιμή στο 1, τόσο βέλτιστη είναι η κανονικότητα των δεδομένων σε κάθε μετρική και αντίστροφα- δηλαδή, όσο πιο κοντά στο 0, τόσο λιγότερο κανονική είναι η κατανομή ανά μετρική. Για τον προσδιορισμό της κανονικότητας των δεδομένων χρησιμοποιήθηκαν οι τιμές p-values Shapiro-Wilk. Σε όλες τις μετρικές ορίστηκε ένα επίπεδο άλφα p-value > 0,05 - όσο μεγαλύτερη είναι η τιμή p-value, τόσο ισχυρότερη είναι η υπόθεση ότι κάθε μετρική δεν παρουσίαζε ενδείξεις μη κανονικότητας.

Χρησιμοποιήσαμε τον μέσο όρο και την τυπική απόκλιση ανά μετρική για να μετρήσουμε την κεντρική τάση μεταξύ των μετρικών που συλλέχθηκαν. Η λοξότητα συμμετέχει επίσης στην κατανόηση της τάσης κάθε μετρικής. Η τελευταία λαμβάνει θετικές και αρνητικές τιμές. Για παράδειγμα, το εργαλείο CheckBot εξάγει τις τιμές των μετρικών σε εκατοντάβαθμη κλίμακα (von Hippel, 2005). Οι αρνητικές τιμές για συγκεκριμένες μετρικές υποδηλώνουν ότι το δείγμα στην εξεταζόμενη μετρική τείνει να είναι 100. Στην αντίθετη κατεύθυνση, η θετική λοξότητα επισημαίνει ότι οι τιμές τους τείνουν να είναι 0. Για όλες τις μετρικές, έχει αρθρωθεί μια τιμή λοξότητας για να γίνει κατανοητό αν οι εμπλεκόμενες μετρικές ανά εξέταση τείνουν προς τις ελάχιστες ή τις μέγιστες τιμές.

3.2.2 Ανάλυση Παλινδρόμησης

Όσον αφορά την ανάπτυξη μοντέλων γραμμικής παλινδρόμησης, η κανονικότητα ή η μη κανονικότητα των δεδομένων δεν επηρεάζει, σε μεγάλο βαθμό, την παραβίαση του μοντέλου εξόδου της παλινδρόμησης (Schmidt and Finan, 2018- Li et al., 2012). Σε αυτό το σημείο, προηγούμενες επιπτώσεις ανέφεραν ότι η κανονικότητα των δεδομένων πρέπει να υπάρχει κατά την ανάπτυξη μοντέλων γραμμικής παλινδρόμησης- διαφορετικά, οι έξοδοι δεν θα έχουν τις κατάλληλες εκ των υστέρων προσδοκίες (Spanos, 1995- Goel and DeGroot, 1980). Στη συγκεκριμένη μελέτη, οι ανακτημένες διαδικτυακές μετρήσεις εξέφραζαν

κανονικότητα στη φύση τους. Το γεγονός αυτό βοήθησε στην ανάπτυξη προγνωστικών μοντέλων γραμμικής παλινδρόμησης για την αμερόληπτη εκτίμηση των τυπικών σφαλμάτων και, συνεπώς, των διαστημάτων εμπιστοσύνης και των τιμών p (Schmidt and Finan, 2018).

Για τα αποτελέσματα της εξίσωσης παλινδρόμησης, πραγματοποιήθηκαν πέντε διαφορετικοί δείκτες. Αυτό συνέβη καθώς αποτελεί κανόνα να αποφεύγονται οι ερμηνείες γραμμικής παλινδρόμησης που βασίζονται μόνο στην τιμή R^2 (Schmidt and Finan, 2018- Colton and Bower, 2002). Ως εκ τούτου, αναπτύχθηκαν η σταθερή τιμή, το R^2 , το προσαρμοσμένο R^2 , το F και η τιμή p . Η σταθερή τιμή υποδηλώνει την τιμή της εξαρτημένης μεταβλητής, ενώ οι επεξηγηματικές μεταβλητές έχουν μηδενικές τιμές (Cai and Hall, 2006). Η τιμή R^2 (ή ο συντελεστής προσδιορισμού) είναι η μέτρηση που εκτιμά το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής που θα μπορούσε να εξηγηθεί από την ανεξάρτητη μεταβλητή (Helland, 1987). Με άλλα λόγια, αναφέρεται στο πόσο καλά τα εμπλεκόμενα δεδομένα ταιριάζουν στο μοντέλο παλινδρόμησης. Οι πολύ υψηλές τιμές R^2 θα πρέπει να αποφεύγονται, καθώς υποδηλώνουν υπερβολική προσαρμογή- συνεπώς, δεν παρέχουν επαρκή προσαρμογή για μελλοντικές παρατηρήσεις που θα μπορούσαν να συμπεριληφθούν στο μοντέλο (Colton and Bower, 2002). Το προσαρμοσμένο R^2 τείνει να μετρά με αισιόδοξο τρόπο την προσαρμογή του προτεινόμενου μοντέλου γραμμικής παλινδρόμησης, ενώ προσδιορίζει το ποσοστό της διακύμανσης του πεδίου-στόχου που εξηγείται από την είσοδο ή τις εισόδους (Leach and Henson, 2007). Η τιμή F παρέχει πληροφορίες σχετικά με την προβλεπτική ικανότητα του προτεινόμενου μοντέλου γραμμικής παλινδρόμησης (Bedeian and Mossholder, 1994). Όσο υψηλότερη είναι η τιμή F , τόσο καλύτερη είναι η προβλεπτική ικανότητα του προτεινόμενου μοντέλου γραμμικής παλινδρόμησης.

Τέλος, οι τιμές p -values κάθε εξίσωσης παλινδρόμησης χρησιμοποιήθηκαν για να εκτιμηθεί εάν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ εξαρτημένων και ανεξάρτητων μεταβλητών (W3C School30, 2022). Μια χαμηλή τιμή p -value κάτω από την καθορισμένη ανοχή σημαντικότητας υποδεικνύει μια σχέση μεταξύ των εξαρτημένων και ανεξάρτητων μεταβλητών (στην περίπτωσή μας, εννοούμε τις μετρικές ιστού). Αντιφατικά, υψηλότερες τιμές p -values πάνω από την καθορισμένη ανοχή σημαντικότητας εκφράζουν μια μη στατιστικά σημαντική συσχέτιση μεταξύ των εξαρτημένων και ανεξάρτητων μεταβλητών. Για τη συγκεκριμένη μελέτη, η ανοχή p -value παλινδρόμησης ορίστηκε σε $< 0,05$, όπως ακριβώς πρότειναν οι Maneejuk και Yamaka (2021) και ο Cohen (2013).

Κεφάλαιο 4. Αποτελέσματα – Ευρήματα / Επιτεύγματα

Στο παρόν κεφάλαιο βρίσκονται συγκεντρωμένα όλα τα αποτελέσματα- ευρήματα, τα οποία παρήχθησαν μετά από τις αναλύσεις που πραγματοποιήσαμε. Πιο συγκεκριμένα, υπάρχουν τα αποτελέσματα των τεσσάρων βασικών μετρικών (Total Website Score, SEO, Speed, Security) όπως αυτές ορίστηκαν από το Checkbot. Οι βαθμολογίες καταλήγουν στην σύγκριση των αποτελεσμάτων μεταξύ Google και Bing ώστε να διαπιστωθεί ποια μηχανή αναζήτησης εμφάνισε σχετικότερα αποτελέσματα αλλά και το ποσοστό αλληλοκάλυψης, ή με άλλα λόγια το βαθμό κατά τον οποίο οι δύο μηχανές ανέκτησαν τον ίδιο σύνδεσμο. (Overlapping URLs). Επιπλέον, παρουσιάζονται τα ευρήματα της παλινδρόμησης των υπομέτρικών των τριών κύριων αξόνων απόδοσης (SEO, Speed, Security) δείχνοντας το ποσοστό που επηρεάζει ο καθένας την εκάστοτε ιστοσελίδα. Όλα τα αποτελέσματα βρίσκονται σε μορφή πινάκων και διαγραμμάτων.

4.1 Αποτελέσματα περιγραφικής στατιστικής ανάλυσης

Πίνακας 2. Μέσος όρος της απόδοσης των ιστοσελίδων των δέκα πρώτων αποτελεσμάτων ανάκτησης στην μηχανή Bing

Search Terms	Total Website Score	SEO	Speed	Security
Australia vs India	72.4	75	79.4	63.2
India vs England	70.8	74.4	75.4	62.4
IPL	74	76.6	82.2	64.6
NBA	65	68.2	68	59
Euro 2021	74.8	76.6	76.2	76.4
Copa América	72.2	74.2	76.4	69.2
India vs New Zealand	71.4	69.8	80.6	71.4
T20 World Cup	74.6	75.4	84.6	67.4
Squid Game	75.2	74.8	85.2	70
DMX	69.2	71.2	74.6	64.4

Πίνακας 3. Μέσος όρος της απόδοσης των ιστοσελίδων των δέκα πρώτων αποτελεσμάτων ανάκτησης στη μηχανή Google

Search Terms	Total Website Score	SEO	Speed	Security
Australia vs India	71.6	72.6	78.2	67.2
India vs England	73	75.4	79.4	64.2
IPL	68.6	66	81.4	64.8
NBA	75	77.2	82.2	67.8
Euro 2021	77.2	76.8	85.2	76.6
Copa América	73	78.2	75.4	61
India vs New Zealand	71	69.4	81.6	66.8
T20 World Cup	73	72.4	81.8	67.6
Squid Game	78	78	82	79.8
DMX	77	76.4	87	76.2

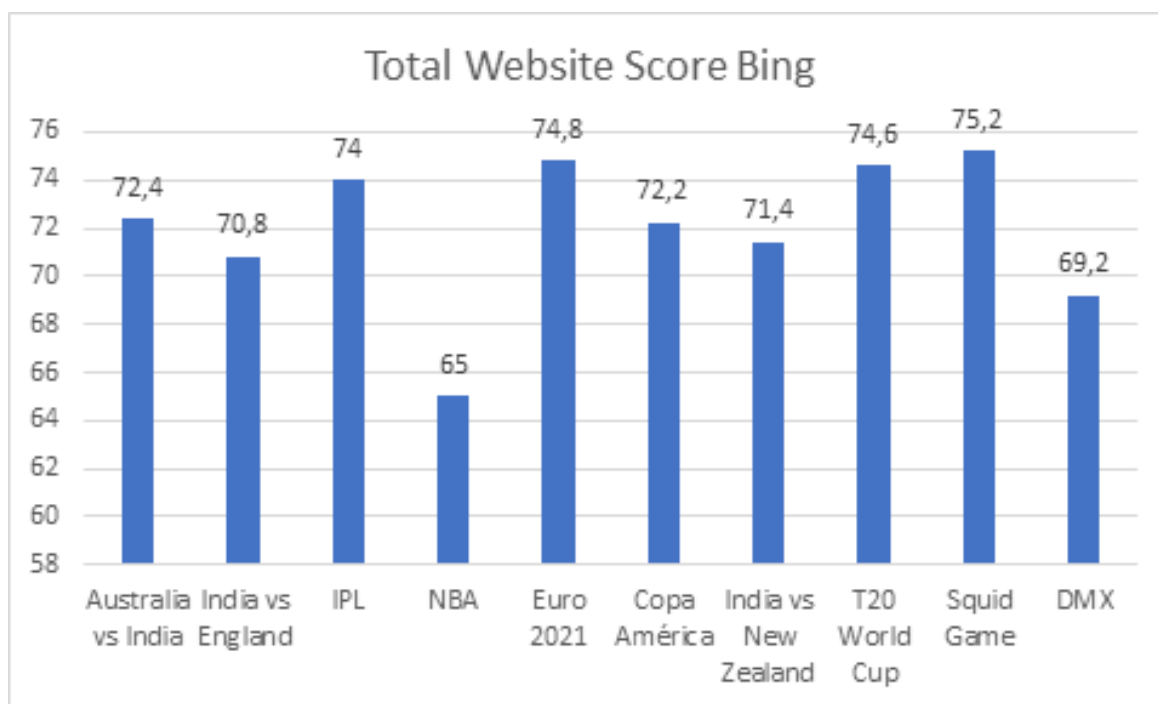
Οι δύο παραπάνω πίνακες συνετέλεσαν στη δημιουργία 4 διαγραμμάτων για την Bing και άλλων 4 για την Google τα οποία ακολουθούν στην επόμενη ενότητα. Κάθε ένα από τα διαγράμματα αξιολογεί τα αποτελέσματα της κάθε μηχανής αναζήτησης σε σχέση με την κάθε μία μετρική που επιλέξαμε. Όπως αναφέρθηκε και στο μεθοδολογικό μέρος της εργασίας, τα δεδομένα προήλθαν από το Checkbot.

Πίνακας 4. Σύγκριση. Μέσος όρος των μετρικών απόδοσης ανάμεσα στις δύο μηχανές.

Metrics	Google	Bing	Διαφορά
Total Website Score	73.64	71.9	1.73
SEO	74.24	73.6	0.62
Speed	81.42	78.2	3.16
Security	69.2	66.8	2.4

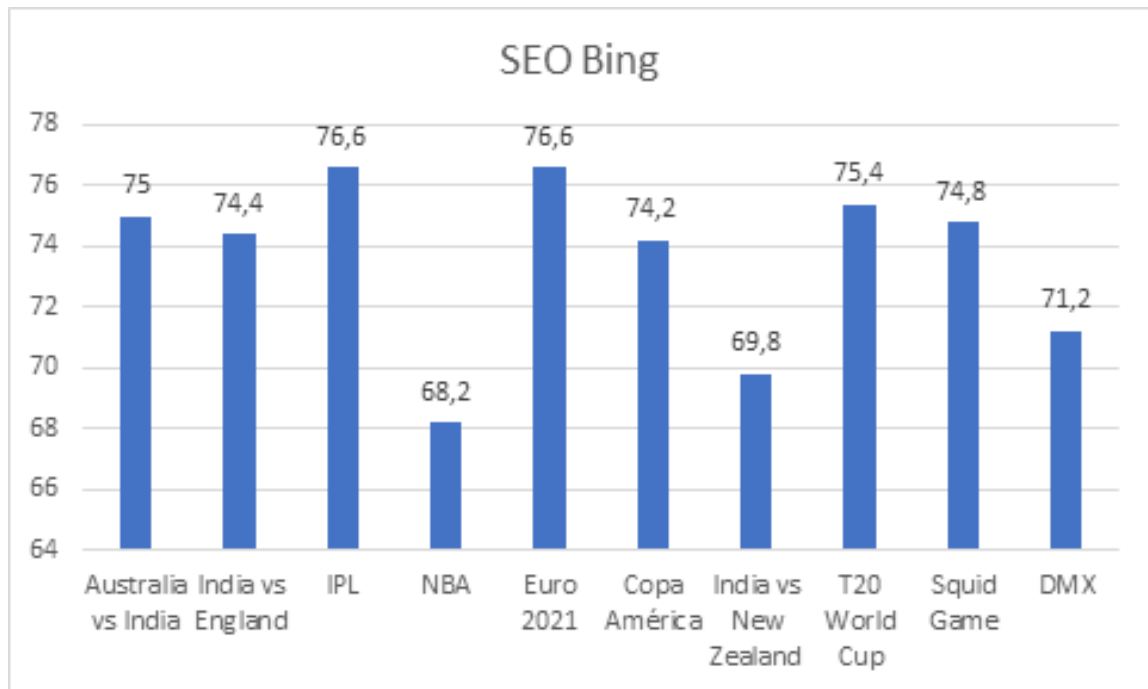
Ο παραπάνω πίνακας δείχνει το μέσο όρο της κάθε μετρικής που συγκέντρωσε η Bing και η Google με βάση τα δεδομένα που συλλέχθηκαν το Checkbot. Πιο συγκεκριμένα, στη μετρική Total Website Score, η Google εμφάνισε αριθμητική διαφορά ύψους 1.73 σε σχέση με τη Bing. Η ίδια λογική επικρατεί και στη μετρική SEO, όπου πάλι προηγείται η Google με διαφορά 0.62. Επιπλέον, η Google είναι ανώτερη και στη μετρική Speed, αφού η μεταξύ τους διαφορά με τη Bing είναι 3.16. Τέλος, στη μετρική Security, η Google προηγείται έναντι της Bing με διαφορά 2.4. Με βάση τα παραπάνω δεδομένα γίνεται αντιληπτό, ότι τα αποτελέσματα, που ανέκτησε η Google συγκέντρωσαν καλύτερη βαθμολογία, καθώς προηγείται και στις τέσσερις βασικές μετρικές έναντι της Bing. Ωστόσο, η μεταξύ τους διαφορά δεν είναι πολύ μεγάλη.

Διαγράμματα Μηχανής Αναζήτησης Bing



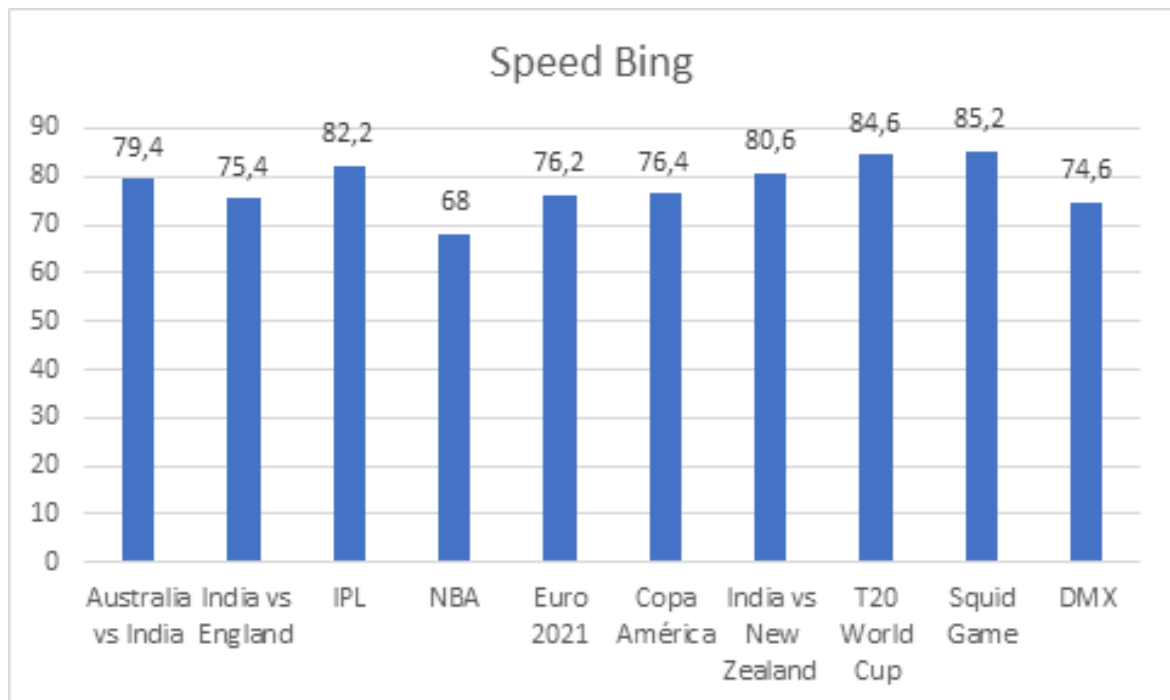
Εικόνα 11. Βαθμολογίες Total Website Score Bing

Στο παραπάνω διάγραμμα, παρατηρούμε την συνολική βαθμολογία των πέντε πρώτων URLs που εμφάνισε η Bing, όπως αυτή αποδόθηκε από το Checkbot για κάθε έναν από τους 10 όρους, τους οποίους επιλέξαμε εξαρχής μέσω της πλατφόρμας του Google Trends. Μεγαλύτερη βαθμολογία παρουσιάζουν τα αποτελέσματα του όρου Squid Game με 75,2% και Euro 2021 με 74,8%. Αντίθετα, μικρότερη βαθμολογία παρουσιάζουν ιστοσελίδες που περιείχαν τον όρο NBA με ποσοστό 65%.



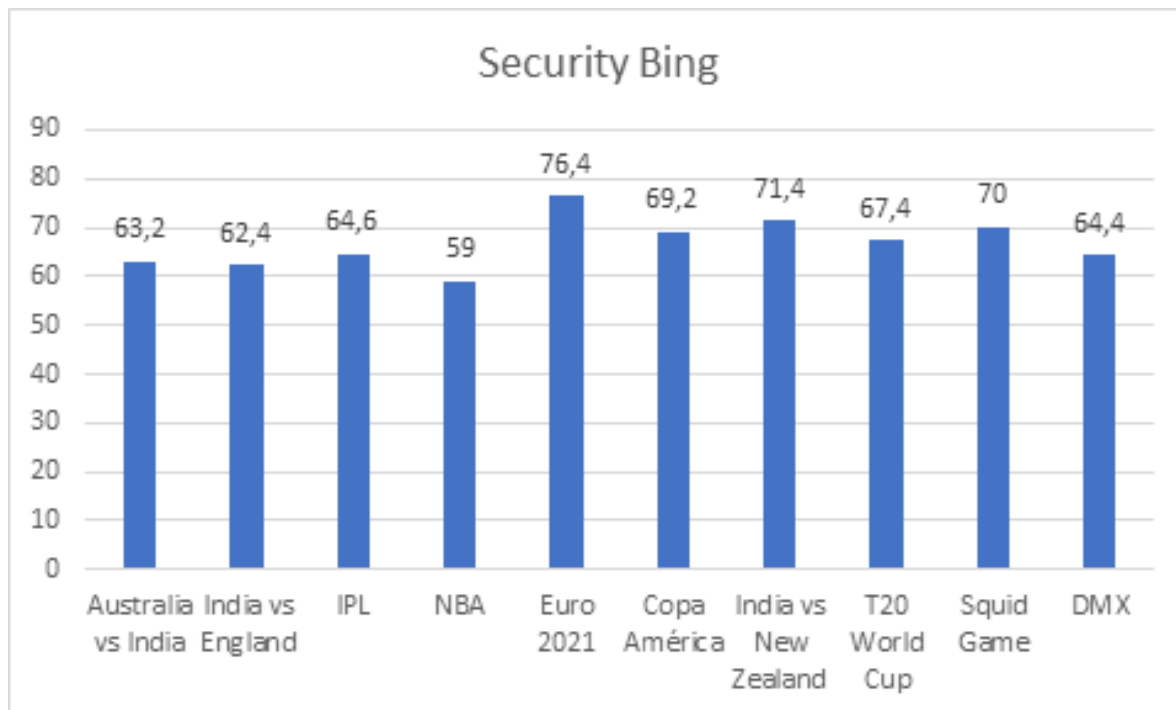
Εικόνα 12. Βαθμολογίες SEO Bing

Το παραπάνω διάγραμμα, απεικονίζει το ποσοστό της μετρικής SEO, που συγκέντρωσαν τα πέντε πρώτα URLs που εμφάνισε η Bing για κάθε όρο, όπως αποδόθηκε από το Checkbot. Τα αποτελέσματα των όρων IPL και Euro 2021 συγκέντρωσαν τη μεγαλύτερη βαθμολογία με ποσοστό 76,6. Αντιθέτως, τα αποτελέσματα με τη μικρότερη βαθμολογία ήταν των NBA με ποσοστό 68,2 και India vs New Zealand με ποσοστό 69,8.



Εικόνα 13. Βαθμολογίες Speed Bing

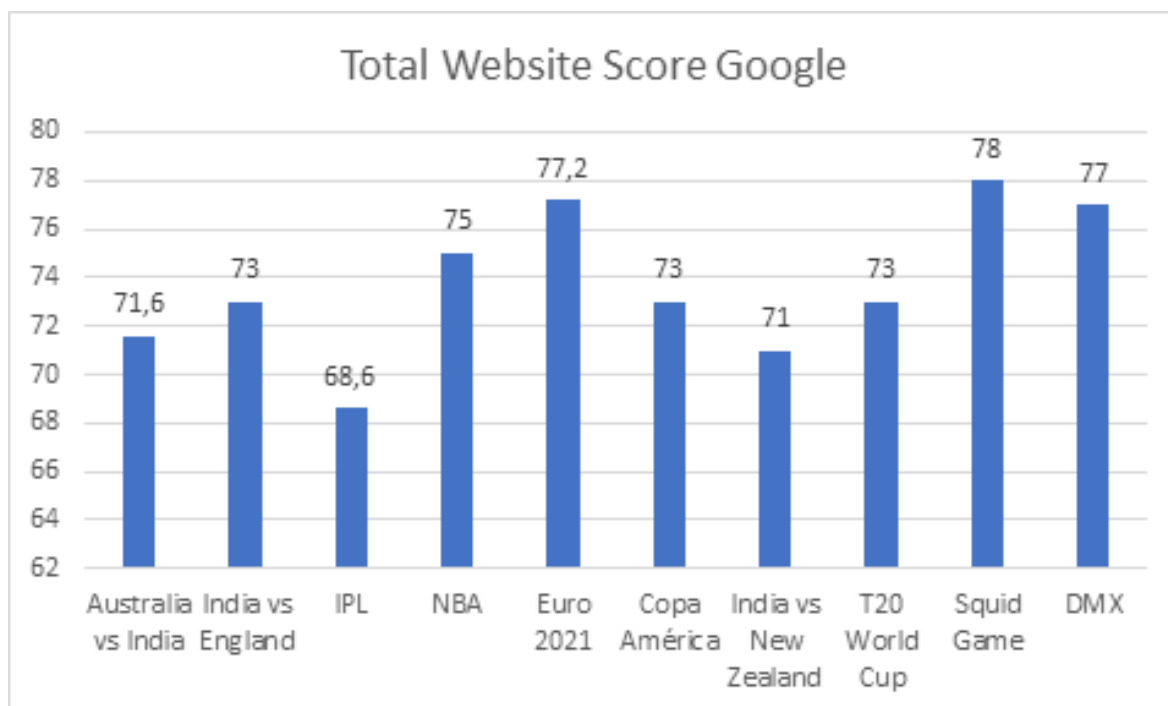
Το παραπάνω διάγραμμα, δείχνει την μετρική της ταχύτητας που είχαν τα πέντε URLs που εμφάνισε πρώτα η Bing για κάθε όρο, σύμφωνα με το Checkbot. Μεγαλύτερο ποσοστό παρουσιάζουν τα αποτελέσματα των όρων Squid Game με 85,2% και T20 World Cup με 84,6%, ενώ μικρότερο ποσοστό συγκέντρωσαν τα αποτελέσματα του όρου NBA 68%.



Εικόνα 14. Βαθμολογίες Security Bing

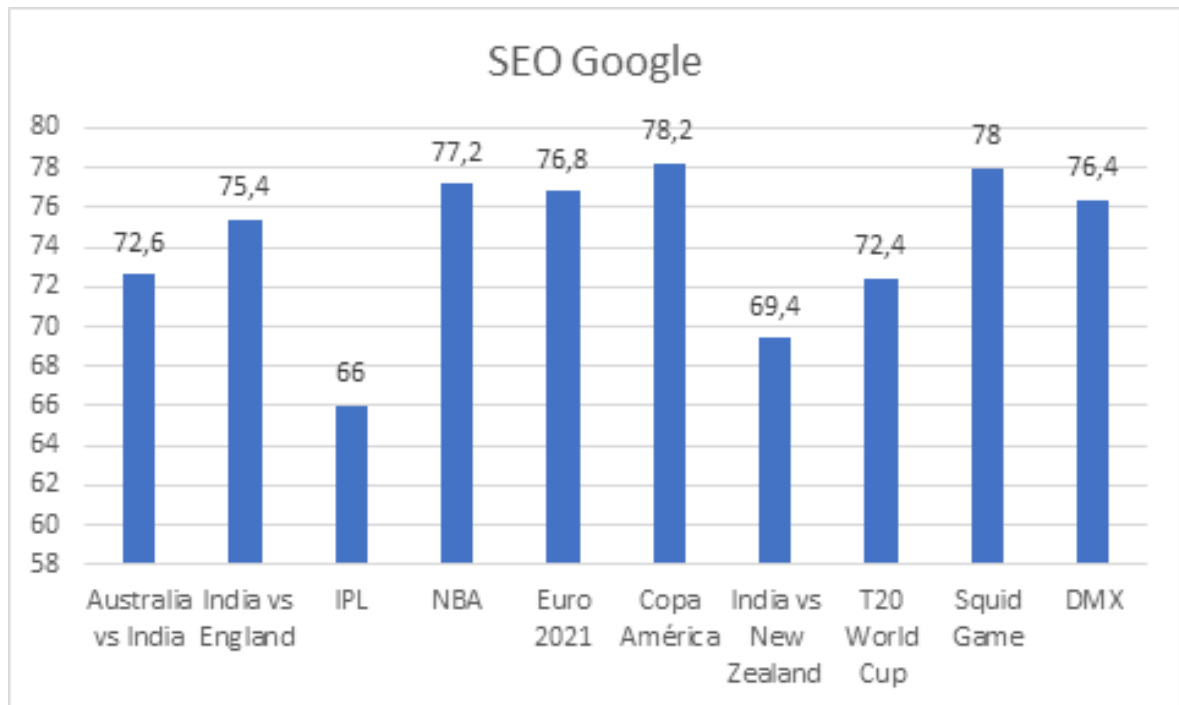
Το παραπάνω διάγραμμα, δείχνει την μετρική της ασφάλειας, που είχαν τα πέντε URLs που εμφάνισε πρώτα η Bing για κάθε όρο, σύμφωνα με το Checkbot. Μεγαλύτερο ποσοστό συγκέντρωσαν Οι ιστοσελίδες που περιείχαν τους όρους Euro 2021 με 76,4% και India vs New Zealand με 71,4%, ενώ μικρότερο ποσοστό συγκέντρωσαν τα αποτελέσματα που περιείχαν τους όρους NBA με 59% και India vs England με 62,4.

4.1.1 Διαγράμματα Μηχανής Αναζήτησης Google



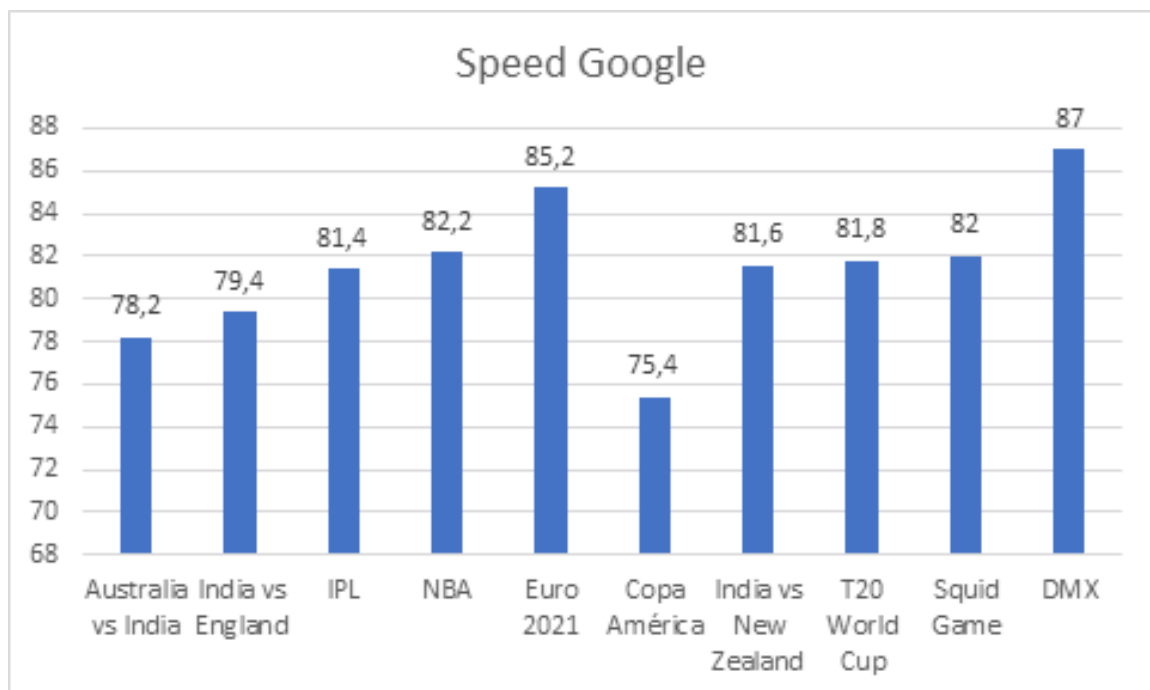
Εικόνα 15. Βαθμολογίες Total Website Score Google

Στο παραπάνω διάγραμμα, παρατηρούμε την συνολική βαθμολογία των πέντε πρώτων URLs που εμφάνισε η Google, όπως αυτή αποδόθηκε από το Checkbot για κάθε έναν από τους 10 όρους, τους οποίους επιλέξαμε εξ αρχής. Μεγαλύτερο ποσοστό συγκέντρωσαν τα αποτελέσματα των όρων Squid Game με ποσοστό 78% και Euro 2021 με 77,2. Αντιθέτως, τα αποτελέσματα των όρων IPL με 68,6 και India vs New Zealand με 71% ήταν τα χαμηλότερα βαθμολογικά.



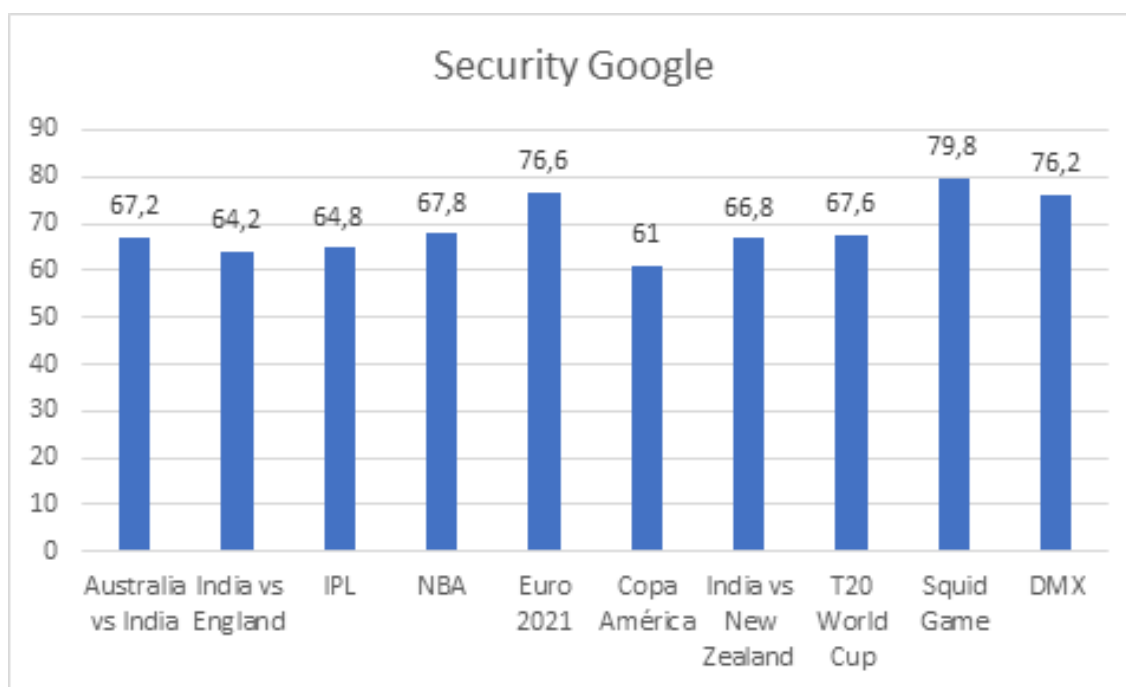
Εικόνα 16. Βαθμολογίες SEO Google

Στο παραπάνω διάγραμμα παρατηρούμε τη μετρική της SEO των πέντε πρώτων URLs που εμφάνισε η Google, όπως αυτή αποδόθηκε από το Checkbot για κάθε έναν από τους 10 όρους, τους οποίους επιλέξαμε εξαρχής. Τα μεγαλύτερα ποσοστά ήταν τα αποτελέσματα των όρων Cora America με 78,2% και Squid Game με 78%, ενώ το μικρότερο ποσοστό, τα αποτελέσματα του όρου IPL με 66%.



Εικόνα 17. Βαθμολογίες Speed Google

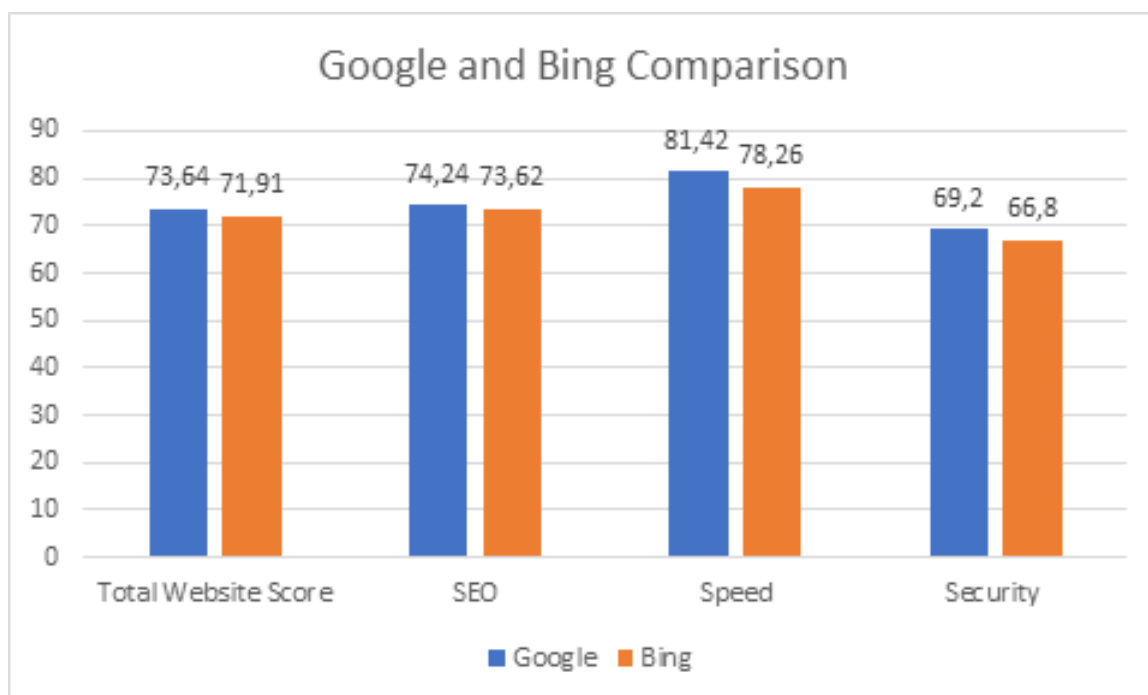
Στο παραπάνω διάγραμμα, παρατηρούμε την μετρική της ταχύτητας των πέντε πρώτων URLs που εμφάνισε η Google, όπως αυτή αποδόθηκε από το Checkbot για κάθε έναν από τους 10 όρους, τους οποίους επιλέξαμε εξαρχής. Τα αποτελέσματα του όρου DMX παρουσιάζουν το μεγαλύτερο ποσοστό, μεταξύ των 10 όρων, με 87%. Ενώ Τα αποτελέσματα του όρου Cora America ήταν τα χαμηλότερα στη συγκεκριμένη σύγκριση με ποσοστό 75,4.



Εικόνα 18. Βαθμολογίες Security Google

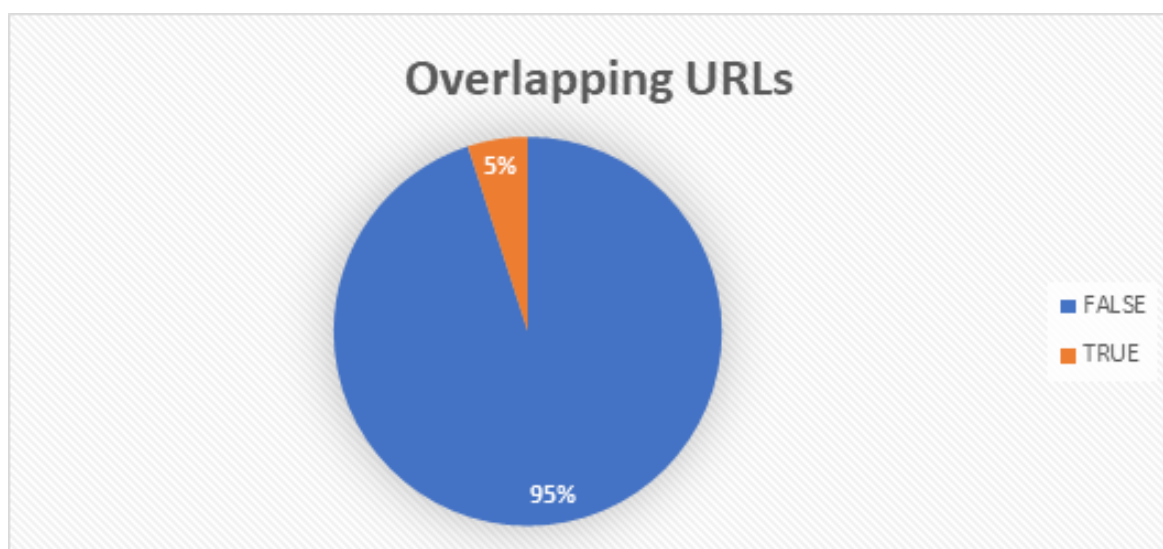
Στο παραπάνω διάγραμμα, παρατηρούμε την μετρική της ασφάλειας των πέντε πρώτων URLs που εμφάνισε η Google, όπως αυτή αποδόθηκε από το Checkbot για κάθε έναν από τους 10 όρους, τους οποίους επιλέξαμε εξαρχής. Οι ιστοσελίδες που περιείχαν τον όρο Squid Game εμφάνισαν το υψηλότερο ποσοστό, πιο συγκεκριμένα το ποσοστό της τάξης του 79,8%. Αντίθετα, οι ιστοσελίδες που περιείχαν τον όρο Cora America έλαβαν τη μικρότερη βαθμολογία με 61%.

4.1.2 Σύγκριση διαγραμμάτων Μηχανών Google και Bing



Εικόνα 19. Διάγραμμα σύγκρισης των βασικών μετρικών Google και Bing

Το συγκεκριμένο διάγραμμα δείχνει την σύγκριση των δύο μηχανών αναζήτησης Google και Bing, όσον αφορά τις 4 βασικές μετρικές του Checkbot. Τα παραπάνω αποτελέσματα δημιουργήθηκαν από τους μέσους όρους των πέντε πρώτων συνδέσμων κάθε όρου για καθεμία μηχανή αναζήτησης. Παρατηρείται προβάδισμα της Google έναντι της Bing και στις 4 κατηγορίες. Η μεγαλύτερη διαφορά βαθμολογίας παρατηρείται στην μετρική ταχύτητα με διαφορά 3,16% υπέρ της Google, ενώ η μικρότερη διαφορά σημειώνεται στη κατηγορία του SEO με 0,62% υπέρ της Google.



Εικόνα 20. Πίτα ποσοστού αλληλοκάλυψης των αποτελεσμάτων Google και Bing

Η παραπάνω πίνα απεικονίζει το ποσοστό αλληλοκάλυψης (results overlapping) αποτελεσμάτων των δύο μηχανών αναζήτησης στα 5 πρώτα αποτελέσματα που εμφάνισαν για κάθε όρο. Παρατηρείται αλληλοκάλυψη αποτελεσμάτων σε ποσοστό 5%. Πιο απλά, από τα 100 αποτελέσματα που εμφάνισαν ξεχωριστά οι δύο μηχανές αναζήτησης, τα 5 αποτελέσματα ήταν ακριβώς τα ίδια, δηλαδή εμφάνισαν ίδιο σύνδεσμο και ίδια σειρά κατάταξης (URL).

4.2 Αποτελέσματα Ανάλυσης Παλινδρόμησης

Στο συγκεκριμένο υποκεφάλαιο παρουσιάζονται σε μορφή πίνακα, τα αποτελέσματα της παλινδρόμησης, ανά μετρική (SEO, Speed, Security). Κάθε ένας από τους 3 πίνακες δείχνει τις υπομετρικές, οι οποίες έχουν το μεγαλύτερο αντίκτυπο στην βελτίωση της ιστοσελίδας στις συγκεκριμένες μετρικές και άρα της συνολικής βαθμολογίας της ιστοσελίδας. Για την δημιουργία των πινάκων παλινδρόμησης των τριών μετρικών, που αναφέρθηκαν, χρησιμοποιήθηκαν οι υπομετρικές της μετρικής Total Website Score. Αν χρησιμοποιούσαμε άλλη μετρική ως βάση, θα είχαν παραχθεί άλλα δεδομένα.

4.2.1 Αποτελέσματα Παλινδρόμησης για τον Άξονα του SEO

Μετρικές	Συντελεστής μεταβολής	R ²	Προσαρμοσμένο R ²	Τιμή F	Τιμή στ. Σημαντικότητας p
Total Website Score (Constant)	57.366	0.187	0.170	11.045	0.002
Use Unique Titles	0.186				
Total Website Score (Constant)	70.907	0.022	0.002	1.074	0.305
Use H1 Headings	0.036				
Total Website Score (Constant)	69.234	0.126	0.107	6.889	0.012
Use H1 heading per page	0.075				
Total Website Score (Constant)	72.473	0.005	-0.0016	0.222	0.640
Use Optimal Length H1 Headings	0.017				

Total Website Score (Constant)	64.409	0.471	0.460	42.805	<.001
<i>Use Unique Headings</i>	0.149				
Total Website Score (Constant)	69.732	0.051	0.031	2.560	0.116
<i>Set Page Descriptions</i>	0.050				
Total Website Score (Constant)	74.855	0.008	-0.012	0.405	0.527
<i>Use Optimal Length Descriptions</i>	-0.022				
Total Website Score (Constant)	64.561	0.220	0.203	13.516	<.001
<i>Set Canonical URLs</i>	0.111				
Total Website Score (Constant)	50.975	0.219	0.202	13.424	<.001
<i>Avoid Duplicate Page Content</i>	0.241				
Total Website Score (Constant)	68.607	0.140	0.122	7.797	0.007
<i>Set Alt Texts</i>	0.077				
Total Website Score (Constant)	86.154	0.222	0.206	13.732	<.001
<i>Set Mobile Scaling</i>	-0.139				
Total Website Score (Constant)	69.388	0.025	0.004	1.217	0.275
<i>Use Short URLs</i>	0.054				

Πίνακας 5. Αποτελέσματα παλινδρόμησης για τον άξονα του SEO

Αναφορικά με τον πίνακα παρατηρούνται τα εξής αποτελέσματα παλινδρόμησης. Αρχικά, σχετικά την επιρροή που ασκεί η μετρική Use Unique Titles στην συνολική απόδοση της ιστοσελίδας, αναφέρεται ότι το μοντέλο πρόβλεψης χαρακτηρίζεται από στατιστική σημαντικότητα με τιμή p της τάξης του 0.002. Παράλληλα, το μοντέλο διακατέχεται από ικανοποιητική στατιστική ικανότητα πρόβλεψης (prediction capability) εμφανίζοντας τιμή R^2 0.187 και τιμή F 11.045. Ειδικότερα για κάθε μία μονάδα που οι διαχειριστές των συγκεκριμένων ιστοσελίδων πρόκειται να βελτιώσουν την μετρική Use Unique Titles, η συνολική απόδοση της ιστοσελίδας θα μεταβληθεί κατά 0.186.

Αντίθετα, η μετρική Use H1 Headings, δεν εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή p 0.305, τιμή που ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Έτσι παρόλο που η πρόβλεψη εμφανίζει ότι η συνολική απόδοση της ιστοσελίδας μπορεί να μεταβληθεί κατά 0.036 για κάθε μία μονάδα που αυξάνεται η μετρική Use H1 Headings, οι διαχειριστές δεν μπορούν να βασιστούν στην συγκεκριμένη πρόβλεψη μιας και αυτή δεν εμφανίζει στατιστική αξιοπιστία.

Στη μετρική Use H1 heading per page διαπιστώνουμε, ότι αυτή εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή p 0.012, τιμή που δε ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Η πρόβλεψη, βασισμένη στη τιμή F , εμφανίζει ότι η συνολική απόδοση της ιστοσελίδας μπορεί να μεταβληθεί κατά 0.075 για κάθε μία μονάδα που αυξάνετε η μετρική Use H1 Headings. Οι διαχειριστές μπορούν να βασιστούν στην συγκεκριμένη πρόβλεψη, καθώς εμφανίζει στατιστική αξιοπιστία.

Η μετρική Use Optimal Length, δεν εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή p 0.640, τιμή που ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Έτσι παρόλο που η πρόβλεψη εμφανίζει ότι η συνολική απόδοση της ιστοσελίδας μπορεί να μεταβληθεί κατά 0.017 για κάθε μία μονάδα που αυξάνεται η μετρική Use H1 Headings, οι διαχειριστές δεν μπορούν να βασιστούν στην συγκεκριμένη πρόβλεψης μιας και αυτή δεν εμφανίζει στατιστική αξιοπιστία.

Αντιθέτως, η μετρική Use Unique Headings εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή p <.001. Παράλληλα, το μοντέλο διακατέχεται από ικανοποιητική στατιστική ικανότητα πρόβλεψης (prediction capability) εμφανίζοντας τιμή R^2 0.460 και τιμή F 42.805. Ειδικότερα για κάθε μία μονάδα που οι διαχειριστές των συγκεκριμένων ιστοσελίδων βελτιώσουν την μετρική Use Unique Titles, η συνολική απόδοση της ιστοσελίδας θα μεταβληθεί κατά 0.149.

Η μετρική Use Optimal Length, δεν εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή p 0.640. Έτσι παρόλο που η πρόβλεψη εμφανίζει, ότι η συνολική απόδοση της ιστοσελίδας μπορεί να μεταβληθεί κατά 0.050 για κάθε μία μονάδα που αυξάνεται η μετρική Use H1 Headings, οι διαχειριστές δεν μπορούν να βασιστούν στην συγκεκριμένη πρόβλεψης μιας και αυτή δεν εμφανίζει στατιστική αξιοπιστία.

Ακόμα, χειρότερη παρουσιάζεται η μετρική Use Optimal Length Descriptions. Πιο αναλυτικά, ο βαθμός στατιστικής σημαντικότητας είναι υπερβολικά υψηλός με τιμή 0.527, η οποία ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Η εν λόγω μετρική δεν παρουσιάζει ικανοποιητική στατιστική ικανότητα πρόβλεψης με τιμή R^2 0.008 και F 0.405. Τέλος, ο Συντελεστής μεταβολής έχει αρνητική τιμή -0,022 με αποτέλεσμα οι διαχειριστές να μην θεωρούν τη συγκεκριμένη μετρική σημαντική για την ανάπτυξη της ιστοσελίδας.

Στο άλλο άκρο παρουσιάζεται η μετρική Set Canonical URLs, καθώς εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή $p < .001$. Παράλληλα, το μοντέλο διακατέχεται από ικανοποιητική στατιστική ικανότητα πρόβλεψης (prediction capability) εμφανίζοντας τιμή R^2 0.220 και τιμή F 13.516. Ειδικότερα, για κάθε μία μονάδα που οι διαχειριστές των συγκεκριμένων ιστοσελίδων βελτιώσουν την μετρική Use Unique Titles, η συνολική απόδοση της ιστοσελίδας θα μεταβληθεί κατά 0.111.

Παρόμοια δεδομένα παρουσιάζει η μετρική Avoid Duplicate Page Content καθώς, εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή $p < .001$, η οποία δεν ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Παράλληλα, το μοντέλο διακατέχεται από ικανοποιητική στατιστική ικανότητα πρόβλεψης (prediction capability) εμφανίζοντας τιμή R^2 0.219 και τιμή F 13.424. Ειδικότερα, για κάθε μία μονάδα που οι διαχειριστές των συγκεκριμένων ιστοσελίδων βελτιώσουν την μετρική Use Unique Titles, η συνολική απόδοση της ιστοσελίδας θα μεταβληθεί κατά 0.241.

Η μετρική Set Alt Texts παρουσιάζει τιμή p 0.007, όπου διαπιστώνουμε ότι και αυτή εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με την τιμή P να μην ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Το μοντέλο διακατέχεται από ικανοποιητική στατιστική ικανότητα πρόβλεψης παρουσιάζοντας τιμή R^2 0.122 και τιμή F 11.045. Για κάθε μια μονάδα που οι διαχειριστές των συγκεκριμένων ιστοσελίδων βελτιώσουν την μετρική Set Alt Texts η συνολική απόδοση θα μεταβληθεί κατά 0.077.

Όσον αφορά τον αντίκτυπο της μετρικής Set Mobile Scaling στην συνολική απόδοση των ιστοσελίδων, το μοντέλο πρόβλεψης χαρακτηρίζεται από τιμή $P < .001$ η οποία είναι στατιστικά σημαντική. Ταυτόχρονα, το μοντέλο έχει επαρκή στατιστική ικανότητα πρόβλεψής με τιμή R^2 0.222 και τιμή F 13.732. Συγκεκριμένα, για κάθε μια μονάδα βελτίωσης

της μετρικής Set Mobile Scaling από τους διαχειριστές αυτών των ιστοσελίδων, η συνολική απόδοση θα μεταβληθεί κατά -0.139.

Η μετρική Use Short URLs δεν εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή P 0.275 η οποία είναι πάνω από το όριο στατιστικής σημαντικότητας 0.05. Επομένως, παρόλο που η πρόβλεψη δείχνει ότι μια αύξηση κατά μία μονάδα στη μετρική Use Short URLs μπορεί να μεταβάλει την απόδοση της ιστοσελίδας κατά 0.054 οι διαχειριστές δεν μπορούν να βασιστούν στη συγκεκριμένη πρόβλεψη, καθώς δεν δείχνει κάποια στατιστική αξιοπιστία.

4.2.2 Αποτελέσματα Παλινδρόμησης για τον Άξονα της Ταχύτητας

Μετρικές	Συντελεστής μεταβολής	R ²	Προσαρμοσμένο R ²	Τιμή F	Τιμή σ. Σημαντικότητας p
Total Website Score (Constant)	61.623	0.027	0.007	1.350	0.251
Deeply Nested URLs	0.125				
Total Website Score (Constant)	73.429	0.001	-0.020	0.040	0.841
Valid HTML	0.006				
Total Website Score (Constant)	71.544	0.005	-0.015	0.261	0.612
Avoid Broken Internal Links	0.024				
Total Website Score (Constant)	67.877	0.005	-0.016	0.222	0.640
Use Compression	0.059				
Total Website Score (Constant)	53.419	0.202	0.186	12.161	0.001
Use minification	0.228				
Total Website Score (Constant)	72.402	0.015	-0.005	0.739	0.394
Avoid Excessive Inline JavaScript	0.022				
Total Website Score (Constant)	68.812	0.373	0.360	28.603	<.001

Avoid Render-Blocking JavaScript	0.109					
Total Website Score (Constant)	73.640	0.000	-0.021	0.008	0.929	
Avoid excessive inline CSS	0.003					
Total Website Score (Constant)	69.681	0.055	0.035	2.793	0.101	
Avoid resource redirects	0.050					

Πίνακας 6. Αποτελέσματα παλινδρόμησης για τον άξονα της ταχύτητας

Η μετρική Deeply Nested URLs δεν παρουσίασε ικανοποιητική στατιστική σημαντικότητα με τιμή p 0.251, υποδεικνύοντας τιμή R^2 0.027 και τιμή F 1.350. Οι προβλέψεις με βάση την τιμή F δείχνουν ότι για κάθε αύξηση κατά 1 μονάδα της μετρικής Deeply Nested URLs, η απόδοση μεταβάλλεται κατά 0.125. Ωστόσο, οι διαχειριστές δεν θα πρέπει να βασίζονται σε αυτή την πρόβλεψη, καθώς δεν αποδεικνύει στατιστική αξιοπιστία.

Ακόμα χειρότερα εμφανίζεται η μετρική Valid HTML με τιμή P 0.841 πάνω από το όριο στατιστικής σημαντικότητας 0.05. Η μετρική δεν κρίνεται στατιστικά σημαντική με τιμή R^2 0.001 και τιμή F 0.040. Τέλος για κάθε 1 μονάδα που αυξάνεται η μετρική Valid HTML η απόδοση της θα μεταβληθεί κατά 0.006. Οι διαχειριστές δεν πρέπει να βασιστούν στην μετρική μιας και αυτή δεν παρουσιάζει στατιστική σημαντικότητα.

Η μετρική Avoid Broken Internal Links δεν παρουσιάζει και αυτή ικανοποιητική στατιστική σημαντικότητα με τιμή P 0.612 η οποία είναι πάνω από το στατιστικό όριο σημαντικότητας 0.05. Εμφανίζει τιμή R^2 0.005 και τιμή F 0.261. Επομένως, παρόλο που η πρόβλεψη δείχνει, ότι η αύξηση κατά 1 μονάδα στην μετρική Avoid Broken Internal Links μεταβάλλει την απόδοση της ιστοσελίδας κατά 0.024 οι διαχειριστές δεν μπορούν να βασιστούν στην συγκεκριμένη πρόβλεψη μιας και αυτή δεν εμφανίζει στατιστική αξιοπιστία.

Η μετρική Use Compression δεν παρουσιάζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή 0.640. Εμφανίζει τιμή R^2 0.005 και τιμή F 0.222. Για κάθε 1 μονάδα που αυξάνεται η μετρική Use Compression η απόδοση της ιστοσελίδας μπορεί να μεταβληθεί κατά 0.059. Οι διαχειριστές δεν πρέπει να βασιστούν στην συγκεκριμένη μετρική, διότι δεν παρουσιάζεται κάποια στατιστική σημαντικότητα.

Αντιθέτως, στην μετρική Use Minification παρατηρούμε τιμή P 0.001, η οποία είναι στατιστική σημαντική και κάτω από το όριο του 0.05. Με την τιμή R^2 να είναι 0.202 και την τιμή F 12.161 παρατηρούμε, ότι το μοντέλο διακατέχεται από ικανοποιητική στατιστική ικανότητα πρόβλεψης. Συγκεκριμένα, για κάθε μια μονάδα, που οι διαχειριστές των συγκεκριμένων ιστοσελίδων βελτιώσουν την μετρική Use Minification η συνολική απόδοση της ιστοσελίδας θα μεταβληθεί κατά 0.228.

Η μετρική Avoid Excessive Inline JavaScript δεν εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή p 0.394, τιμή που ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Έτσι, παρόλο που η πρόβλεψη εμφανίζει, ότι η συνολική απόδοση της ιστοσελίδας μπορεί να μεταβληθεί κατά 0.022 για κάθε μία μονάδα που αυξάνεται η μετρική Use H1 Headings, οι διαχειριστές δεν μπορούν να βασιστούν στην συγκεκριμένη πρόβλεψη μιας και αυτή δεν εμφανίζει στατιστική αξιοπιστία.

Αντιθέτως, η μετρική Avoid Render-Blocking JavaScript εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή p <.001. Παράλληλα η πρόβλεψη παρουσιάζει, ότι η συνολική απόδοση της ιστοσελίδας μπορεί να μεταβληθεί κατά 0.109 για κάθε μία μονάδα που αυξάνεται η μετρική Avoid Render-Blocking JavaScript. Οι διαχειριστές, λόγω της στατιστικής σημαντικότητας αλλά και της πρόβλεψης, η οποία παρουσιάζεται από τις τιμές R^2 τιμή 0.373 και F με τιμή 28.603, μπορούν να βασιστούν στην εν λόγω μετρική για την βελτίωση της ιστοσελίδας τους.

Η μετρική Avoid excessive inline CSS δεν παρουσίασε ικανοποιητική στατιστική σημαντικότητα με τιμή p 0.929, ξεπερνώντας σημαντικά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Η μόλις 0.003 μεταβολή σε συνδυασμό με την σχεδόν ανύπαρκτη στατιστική ικανότητα πρόβλεψης παρουσιάζοντας τιμή R^2 0.000 και τιμή F 0.008 δεν μπορεί να ωθήσει τους διαχειριστές στο να την επιλέξουν και να βασιστούν σε αυτή.

Τέλος, η μετρική Avoid resource redirects δεν εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή p 0.305, τιμή που ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.101. Έτσι, παρόλο που η πρόβλεψη εμφανίζει, ότι η συνολική απόδοση της ιστοσελίδας μπορεί να μεταβληθεί κατά 0.050 για κάθε μία μονάδα που αυξάνεται η μετρική

Avoid resource redirects, οι διαχειριστές δεν μπορούν να βασιστούν στην συγκεκριμένη πρόβλεψη μιας και αυτή δεν εμφανίζει στατιστική αξιοπιστία.

4.2.3 Αποτελέσματα Παλινδρόμησης για τον Άξονα της Ασφαλείας

Μετρικές	Συντελεστής μεταβολής	R ²	Προσαρμοσμένο R ²	Τιμή F	Τιμή σ. Σημαντικότητας p
Total Website Score (Constant)	24.149	0.161	0.143	9.205	0.004
Use HTTPS	0.500				
Total Website Score (Constant)	-79.285	0.057	0.038	2.910	0.095
Avoid Mixed Content	1.538				
Total Website Score (Constant)	67.064	0.212	0.196	12.952	<.001
Use HSTS	0.089				
Total Website Score (Constant)	66.103	0.562	0.553	61.640	<.001
Use Content Sniffing Protection	0.134				
Total Website Score (Constant)	70.015	0.363	0.350	27.331	<.001
Use Clickjack Protection	0.112				
Total Website Score (Constant)	70.851	0.269	0.254	17.653	<.001
Use XSS Protection	0.098				
Total Website Score (Constant)	74.640	0.005	-0.016	0.240	0.626
Hide Server Version Data	-0.014				

Πίνακας 7. Αποτελέσματα παλινδρόμηση για τον άξονα της ασφαλείας

Η μετρική Use HTTPS εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή $p < .004$. Παράλληλα, το μοντέλο διακατέχεται από ικανοποιητική στατιστική ικανότητα πρόβλεψης (prediction capability) εμφανίζοντας τιμή $R^2 0.143$ και τιμή $F 9.205$. Ειδικότερα, για κάθε μία μονάδα που οι διαχειριστές των συγκεκριμένων ιστοσελίδων βελτιώσουν την μετρική Use Unique Titles, η συνολική απόδοση της ιστοσελίδας θα μεταβληθεί κατά 0.500.

Αντιθέτως, Η μετρική Avoid Mixed Content δεν εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή $p 0.095$. Εμφανίζει τιμή $R^2 0.057$ και τιμή $F 2.910$. Η πρόβλεψη, βασισμένη στη τιμή F εμφανίζει, ότι η συνολική απόδοση της ιστοσελίδας μπορεί να μεταβληθεί κατά 1.538 για κάθε μία μονάδα που αυξάνεται η μετρική Use H1 Headings. Ωστόσο, οι διαχειριστές δεν μπορούν να βασιστούν στην συγκεκριμένη πρόβλεψη μιας και αυτή δεν εμφανίζει στατιστική αξιοπιστία.

Η μετρική Use HSTS παρουσιάζει τιμή $p < .001$, όπου διαπιστώνουμε ότι και αυτή εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με την τιμή P να μην ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Το μοντέλο διακατέχεται από ικανοποιητική στατιστική ικανότητα πρόβλεψης παρουσιάζοντας τιμή $R^2 0.196$ και τιμή $F 12.952$. Για κάθε μια μονάδα που οι διαχειριστές των συγκεκριμένων ιστοσελίδων βελτιώσουν την μετρική Use HSTS η συνολική απόδοση θα μεταβληθεί κατά 0.089.

Η ίδια λογική παρουσιάζεται και στη μετρική η μετρική Use Content Sniffing Protection, η οποία παρουσιάζει τιμή $p < .001$, όπου διαπιστώνουμε ότι και αυτή εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με την τιμή P να μην ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Το μοντέλο διακατέχεται από ικανοποιητική στατιστική ικανότητα πρόβλεψης παρουσιάζοντας τιμή $R^2 0.562$ και τιμή $F 61.640$. Για κάθε μια μονάδα, που οι διαχειριστές των συγκεκριμένων ιστοσελίδων βελτιώσουν την μετρική Use HSTS η συνολική απόδοση θα μεταβληθεί κατά 0.134.

Σχετικά με την επιρροή που ασκεί η μετρική Use Clickjack Protection στην συνολική απόδοση της ιστοσελίδας, αναφέρεται ότι το μοντέλο πρόβλεψης χαρακτηρίζεται από στατιστική σημαντικότητα με τιμή p της τάξης του 0.01. Το μοντέλο διακατέχεται από ικανοποιητική στατιστική ικανότητα πρόβλεψης παρουσιάζοντας τιμή $R^2 0.363$ και τιμή $F 27.331$. Για κάθε

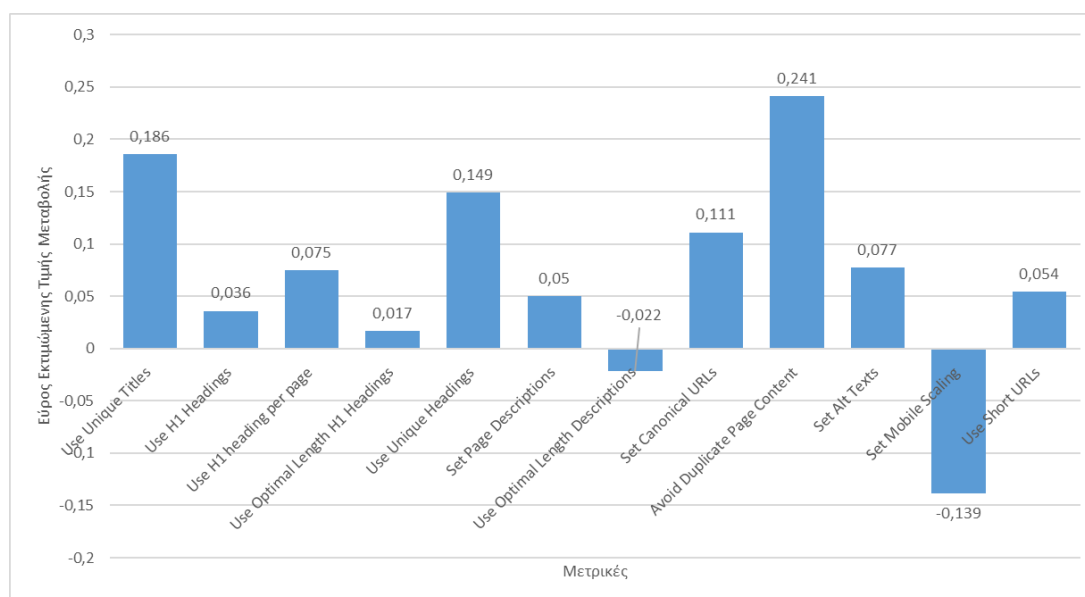
για μονάδα, που οι διαχειριστές των συγκεκριμένων ιστοσελίδων βελτιώσουν την μετρική Use HSTS η συνολική απόδοση θα μεταβληθεί κατά 0.112.

Προς την ίδια κατεύθυνση κινείται η μετρική Use XSS Protection, η οποία παρουσιάζει τιμή $p < .001$, όπου διαπιστώνουμε ότι και αυτή εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με την τιμή P να μην ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Με τιμή R^2 0.226 και τιμή F 17.653 γίνεται αντιληπτή η ικανοποιητική στατιστική ικανότητα πρόβλεψης (prediction capability). Πιο συγκεκριμένα, για κάθε μία μονάδα που οι διαχειριστές των συγκεκριμένων ιστοσελίδων βελτιώσουν την μετρική Use XSS Protection, η συνολική απόδοση της ιστοσελίδας θα μεταβληθεί κατά 0.098.

Τέλος, η μετρική Hide Server Version Data δεν εμφανίζει ικανοποιητικό βαθμό στατιστικής σημαντικότητας με τιμή p 0.626, τιμή που ξεπερνά το στατιστικό όριο σημαντικότητας της τάξης του 0.05. Η παραπάνω μετρική, σε συνδυασμό με τις μετρικές R^2 με τιμή 0.005 και F με τιμή 0.240, οδηγούν τους διαχειριστές στο να μην βασιστούν σε αυτή τη μετρική καθώς η πρόβλεψη δεν δείχνει κάποια στατιστική αξιοπιστία.

4.2.4 Διαγράμματα Παλινδρόμησης και Εκτιμώμενες Τιμές Μεταβολής

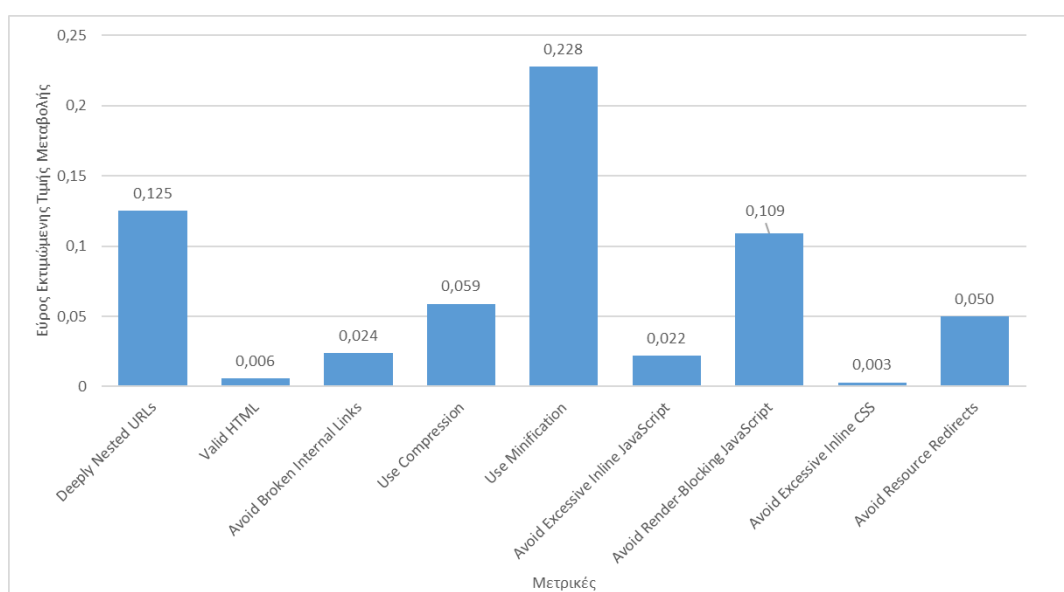
Παρακάτω ακολουθούν τα διαγράμματα των υπομετρικών των τριών βασικών μετρικών (SEO, Speed, Security) που χρησιμοποιήσαμε από τις προτάσεις της Google.Γ



Εικόνα 21. Εκτιμώμενες τιμές μεταβολής για κάθε μετρική στον άξονα SEO

Στο παραπάνω διάγραμμα, εμφανίζονται όλες οι μετρικές που επιλέξαμε να χρησιμοποιήσουμε κρίνοντας, ότι επηρεάζουν το SEO των ιστοσελίδων και πραγματοποιήσαμε μετρήσεις που αναφέρονται στη παλινδρόμηση. Η μετρική με το υψηλότερη τιμή βελτίωσης είναι η Avoid Duplicate Page Content με τιμή 0,241 για κάθε μία μονάδα που προσθέτουν οι διαχειριστές, ενώ η μετρική με τη μικρότερη βελτίωση είναι η Set Mobile Scaling με τιμή -0,139.

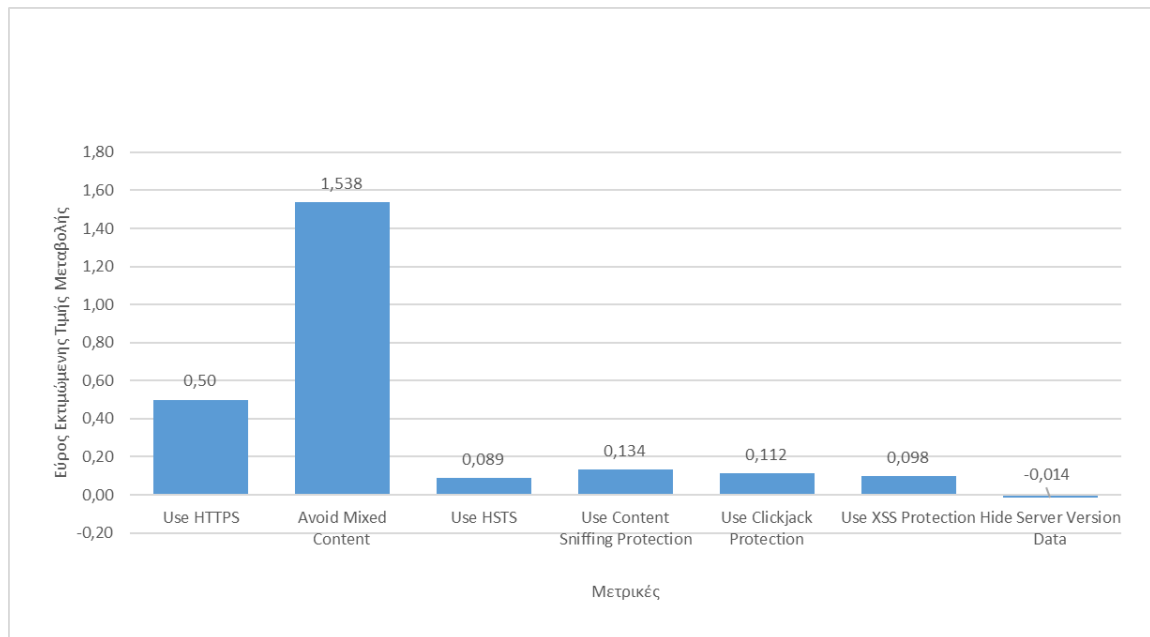
Speed Regression



Εικόνα 22 Εκτιμώμενες τιμές μεταβολής για κάθε μετρική στον άξονα Speed

Στο παραπάνω διάγραμμα, εμφανίζονται όλες οι μετρικές, που επιλέξαμε να χρησιμοποιήσουμε κρίνοντας, ότι επηρεάζουν την ταχύτητα των ιστοσελίδων και πραγματοποιήσαμε μετρήσεις που αναφέρονται στη παλινδρόμηση. Η μετρική με το υψηλότερη τιμή βελτίωσης είναι η Avoid Duplicate Page Content με τιμή 0,228 για κάθε μία μονάδα που προσθέτουν οι διαχειριστές, ενώ η μετρικές με τη μικρότερη βελτίωση είναι η Valid HTML με τιμή -0,006 και η Avoid Excessive Inline CSS με τιμή -0,003.

Security Regression



Εικόνα 23. Εκτιμώμενες τιμές μεταβολής για κάθε μετρική στον άξονα Security

Στο παραπάνω διάγραμμα, εμφανίζονται όλες οι μετρικές, που επιλέξαμε να χρησιμοποιήσουμε κρίνοντας, ότι επηρεάζουν την ασφάλεια των ιστοσελίδων και πραγματοποιήσαμε μετρήσεις που αναφέρονται στη παλινδρόμηση. Η μετρική με το υψηλότερη τιμή βελτίωσης είναι η Avoid Mixed Content με τιμή 1,538 για κάθε μία μονάδα που προσθέτουν οι διαχειριστές, ενώ η μετρικές με τη μικρότερη βελτίωση είναι η Use HSTS με τιμή -0,089 και η Hide Server Version Data με τιμή -0,014.

Κεφάλαιο 5. Συζήτηση – Συμπεράσματα – Μελλοντικές επεκτάσεις

Το τελευταίο κεφάλαιο εμπεριέχει συζήτηση επί της μεθοδολογίας και των αποτελεσμάτων που ανακτήθηκαν χωρίζοντας το πρακτικό από το θεωρητικό μέρος. Με αυτόν τον τρόπο, παρουσιάζονται και συγκρίνονται τα ευρήματα από την βιβλιογραφική έρευνα αλλά και την έρευνα που πραγματοποιήθηκε, ώστε να στελεχωθεί η ερευνητική μας προσπάθεια. Επιπροσθέτως, το κεφάλαιο αυτό, συνεχίζει με τα βασικά συμπεράσματα, τα οποία προέκυψαν από τα αποτελέσματα που εξαγάγαμε. Στη συνέχεια, ξεδιπλώνονται τα προβλήματα που παρουσιάστηκαν κατά την εκπόνηση της εργασίας. Τέλος, παραθέτουμε τις προτάσεις μας για τις μελλοντικές ερευνητικές προεκτάσεις της παρούσας μελέτης.

5.1 Συζήτηση επί της μεθοδολογίας

Για την επίτευξη των στόχων της συγκεκριμένης έρευνας δημιουργήθηκε σχετική μεθοδολογία βοηθώντας να εξαχθούν τα δεδομένα, στα οποία βασίστηκε το πρακτικό μέρος της έρευνας. Η μεθοδολογία χωρίζεται σε δύο μέρη:

- Το θεωρητικό μέρος, κατά το οποίο, συγκεντρώθηκαν, σχετικές με το θέμα, έρευνες/επιστημονικά άρθρα/μελέτες.
- Το πρακτικό μέρος, που αφορά την ερευνητική διαδικασία, κατά την οποία πραγματοποιήθηκε εξαγωγή και σχολιασμός δεδομένων/αποτελεσμάτων.

Κατά το στάδιο της βιβλιογραφικής επισκόπησης (κεφάλαιο 2), τα σημεία αξιολόγησης και επιλογής των δημοσιεύσεων ήταν η κοινή ερευνητική θεματική, κοινή ή παρόμοια μεθοδολογία υλοποίησης πειραμάτων, το ερευνητικό υπόβαθρο των συγγραφέων καθώς και η χρονολογία έκδοσης. Όλα τα παραπάνω, συνετέλεσαν στο φιλτράρισμα και στη συνέχεια στην επιλογή συγκεκριμένων ερευνών, πάνω στις οποίες βασίστηκε και αναπτύχθηκε περαιτέρω η ερευνητική μας προσπάθεια.

Στο στάδιο της μεθοδολογίας (κεφάλαιο 3), κατόπιν έρευνας και αξιολόγησης πραγματοποιήθηκε η συγκέντρωση του ερευνητικού δείγματος που αναφέρθηκε και παραπάνω. Δηλαδή, δεδομένα συμπεριφοράς στις μηχανές αναζήτησης από το Google Trends (δημοτικότητα όρων), καθώς και στοιχεία για την απόδοση των ιστοσελίδων μέσω του εργαλείου checkbot.

Στο πρακτικό μέρος της έρευνας, επιλέχθηκαν μέσω του Google Trends οι 10 δημοφιλέστεροι όροι, που επισκέφτηκαν οι χρήστες παγκοσμίως το έτος 2021. Στη συνέχεια, οι όροι αναζητήθηκαν μέσω των μηχανών αναζήτησης (Google & Bing) και τα αποτελέσματα τους αξιολογήθηκαν με βάση τις μετρικές που αναφέρονται στο κεφάλαιο 4.1 από την επέκταση του Google (Checkbot).

Θεωρητική Συνεισφορά:

Η έρευνα που πραγματοποιήθηκε αποτελεί μία από τις πιο πρόσφατες ερευνητικές προσπάθειες να συγκριθούν οι Google & Bing, όσον αφορά τον βαθμό καταλληλότητας τους ως προς την ευρετηρίαση όρων. Ένα βήμα παρακάτω, η μελέτη εξέτασε το βαθμό σχετικότητας της απόδοσης των εξεταζόμενων ιστοσελίδων ως προς την αυξομείωση της κατάταξης που μπορούν να λάβουν στις δύο μηχανές αναζήτησης. Προγενέστερη μελέτη που διαθέτει παρόμοιο τρόπο σκέψης είναι ο Britvic (2014, May), κ.ά. συγκρίνοντας τα αποτελέσματα που εμφάνισαν οι δυο μηχανές αναζήτησης. Η μεθοδολογία που αναπτύχθηκε, όπως αυτή αναλύθηκε, είναι βασισμένη στους όρους και τις συνθήκες της παγκόσμιας αγοράς που προσφέρουν οι μηχανές αναζήτησης. Κάνοντας ξεκάθαρο τον παραπάνω ισχυρισμό, η μελέτη ανέπτυξε μια μεθοδολογία εύρεσης των όρων χρησιμοποιώντας συγκεκριμένο εργαλείο κατάταξης δημοτικότητας των όρων για την επιλογή τους (Google Trends), αποκλίνοντας σε μεγάλο βαθμό -αν όχι ολιστικά- την υποκειμενική επιλογή και εξέταση τους.

Επιπλέον, τα σύγχρονα εργαλεία (Google Trends/Checkbot/JASP) που χρησιμοποιήθηκαν για την επιλογή, εξαγωγή, επεξεργασία και ανάλυση των δεδομένων διαφοροποιούν την εν λόγω έρευνα από τις προγενέστερες συγκριτικές μελέτες. Η πλειοψηφία των συγκριτικών μελετών έχει, ως βασικό άξονα σύγκρισης την Google αλλά διαφοροποιούνται οι υπόλοιπες μηχανές αναζήτησης. Εξάιρεση στη παραπάνω πρόταση αποτελούν ο Britvic, κ.ά. & French και Fagan (2019), οι οποίοι συγκρίνουν εξ ολοκλήρου μόνο αυτές τις δύο μηχανές, ενώ οι Barbar και Ismail προχώρησαν στη σύγκριση Google, Bing & Yahoo. Φυσικά αξίζει να σημειωθεί ότι ο μεγαλύτερος αριθμός περιπτώσεων μηχανών αναζήτησης προς εξέταση, ακούσια μπορεί να επιφέρει και την μείωση του αριθμού των αξόνων που ελέγχονται ως προς την απόδοση των ιστοσελίδων (Drivas et al. 2021; Onaifo and Rasmussen, 2013). Με άλλα λόγια, όσες πιο πολλές οι μηχανές αναζήτησης, τόσο μεγαλύτερες οι πιθανότητες η εξέταση τους να παρουσιάζει πιο γενικευμένα αποτελέσματα.

Ακόμα, μία επισήμανση σχετίζεται με το ποσοστό αλληλοκάλυψης μεταξύ των μηχανών. Πιο συγκεκριμένα, αναλύοντας τα δεδομένα των πρώτων πέντε αποτελεσμάτων του κάθε όρου που μας εμφάνισε η Google διαπιστώνουμε, ότι το ποσοστό αλληλοκάλυψης (results

overlapping) μεταξύ των δύο μηχανών αναζήτησης ήταν μόλις 5%. Δηλαδή, μόνο οι πέντε από τις εκατό ιστοσελίδες εμφανίστηκαν ακριβώς στην ίδια θέση αποτελεσμάτων κατάταξης στις δύο μηχανές αναζήτησης. Το γεγονός αυτό, εγείρει ερωτήματα σχετικά με τον τρόπο και την αξιοπιστία της αξιολόγησης του SEO από τη Google & Bing. Επιπλέον, αξίζει να σημειωθεί ότι στα αποτελέσματα των 8 από τους 10 όρων, το πρώτο αποτέλεσμα που εμφάνισε δεν ήταν εκείνο με τη μεγαλύτερη βαθμολογία ως προς τα τεχνικά χαρακτηριστικά.

Ακολούθως, το χαμηλότερο ποσοστό website score ιστοσελίδας που ανακτήθηκε στα 5 πρώτα αποτελέσματα ήταν 51%, ενώ το μεγαλύτερο ήταν 89%. Συνεπώς, η ύπαρξη περιπτώσεων με μικρή βαθμολογία που εμφανίστηκαν να καταλαμβάνουν κάποια από τις πρώτες θέσεις σε Google & Bing είναι γεγονός. Η δυσανάλογη θέση της κατάταξης σε σχέση με την βαθμολογία μπορεί να οφείλεται στην συμπεριφορά των χρηστών, η οποία σύμφωνα με συγκεκριμένες ερευνητικές προσπάθειες τείνει να επηρεάζει την αυξομείωση της κατάταξης μιας ιστοσελίδας (Drivas et al., 2020; Mavridis and Symeonidis, 2015).

Πιο αναλυτικά, στην περίπτωση της Google, αξιολογεί την επίσκεψη των χρηστών σε ιστοσελίδες, ελέγχοντας την ευκολία της περιήγησης, το ποσοστό εγκατάλειψης και επιστροφή στην αρχική σελίδα της Google, προσφέροντας παράλληλα την δυνατότητα στους διαχειριστές να έχουν και εκείνοι δικαίωμα πρόσβασης στις επιδόσεις των ιστοσελίδων μέσω της ψηφιακής υπηρεσίας Google Analytics. Στο ίδιο μήκος κύματος κυμαίνεται η Bing, καθώς και αυτή δίνει έμφαση στη σχετικότητα της πληροφορίας-ιστοσελίδας και αξιοπιστίας του περιεχομένου ελέγχοντας τη γνησιότητα των πληροφοριών. Τέλος, προσμετρά την συμπεριφορά των χρηστών κατά την πλοήγηση τους στην εκάστοτε ιστοσελίδα σε συνάφεια με τον έλεγχο της Google που αναλύθηκε παραπάνω.

Πρακτική Συνεισφορά

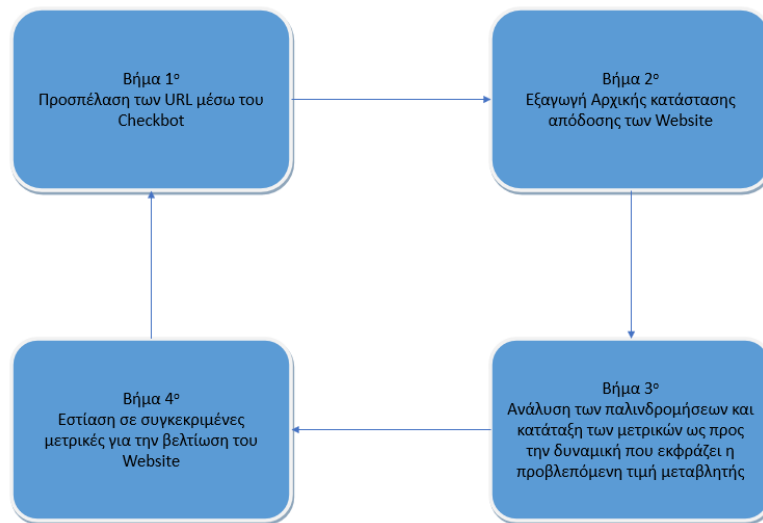
Οι τρεις βασικοί παράγοντες-μετρικές κάθε ιστοσελίδας είναι οι εξής:

- SEO-Βελτιστοποίηση μηχανής αναζήτησης
- Speed-Ταχύτητα
- Security- Ασφάλεια

Κάθε μία από αυτές τις μετρικές διαθέτει υπομετρικές, οι οποίες με τη σειρά τους επηρεάζουν το περιβάλλον, το περιεχόμενο, την ταχύτητα αλλά και την ασφάλεια μιας ιστοσελίδας. Συνολικά εξετάσαμε (SEO=12 Metrics | Speed=9 Metrics | Security=7) 28 μετρικές. Τα δεδομένα που εξήχθησαν μας υποδεικνύουν συγκεκριμένες υπομετρικές, οι οποίες όταν μεταβάλλεται η τιμή τους, επηρεάζουν τη συνολική βαθμολογία και κατά συνέπεια τη κατάταξη της ιστοσελίδας.

Πιο συγκεκριμένα, για τον παράγοντα SEO, όπως εμφανίζει και το σχεδιάγραμμα παλινδρόμησης υπάρχουν μετρικές που ενέχουν μεγαλύτερης προτεραιοποίησης έναντι άλλων, όπως είναι η Avoid Duplicate Page Content με τιμή 0,241 και ακολουθούν οι Use Unique Titles με τιμή 0,186 και Use Unique Headings με τιμή 0,149. Αντιθέτως, οι υπομετρικές που επηρεάζουν την ταχύτητα της ιστοσελίδας είναι η Use Minification με τιμή 0,228 ενώ ακολουθούν οι Deeply Nested URLs με 0,125 και η Avoid Render-Blocking JavaScript με τιμή 0,109. Τέλος, οι υπομετρικές με τη μεγαλύτερη συνεισφορά στην ασφάλεια είναι η Avoid Mixed Content με τιμή 1,538 και η Use HTTPS με τιμή 0,50. Είναι σημαντικό να τονιστεί ότι η προτεραιοποίηση μετρικών που έγινε θα ήταν διαφορετική αν είχαν χρησιμοποιηθεί άλλα δεδομένα. Πιο συγκεκριμένα, με βάση τα ανακτηθέντα δεδομένα και την ανάλυση τους, εξήχθησαν οι συγκεκριμένες μετρικές που ενέχουν προτεραιοποίησης. Σε άλλο δείγμα ιστοσελίδων, μπορεί να εμφανίζονταν πρώτες άλλες μετρικές. Με άλλα λόγια, στην έρευνα μας, δεν υποδεικνύουμε ποιες είναι η μετρικές που χαίρουν προτεραιοποίησης ως προς τη βελτίωσή τους με de facto τρόπο. Αντιθέτως, ένα βήμα πιο πριν, παρουσιάζουμε μια μεθοδολογία η οποία δείχνει τον τρόπο στους διαχειριστές να ανακαλύψουν τις μετρικές που χαίρουν προτεραιοποίησης στη βελτίωση έναντι άλλων.

Επιπλέον, η μεθοδολογία της συγκεκριμένης έρευνας αποτελεί ένα άρτιο παράδειγμα συγκριτικής προτυποποίησης (benchmarking). Η συγκριτική προτυποποίηση είναι ένας σύγχρονος και καθιερωμένος όρος, κατά τον οποίο, μία συνθήκη- κατάσταση θεωρείται επιτυχημένη μέχρι την αντικατάστασή της από μία καλύτερη. Moriarty, J. P. (2008). Ακολουθεί το σχήμα της μεθοδολογίας που χρησιμοποιήθηκε στην ερευνά μας και που προτείνουμε στον εκάστοτε δημιουργό ή διαχειριστή ιστοσελίδων.



Εικόνα 24. Βήματα Ανάλυσης Απόδοσης και Βελτίωσης Ιστοσελίδων

Το παραπάνω σχήμα εκτός από τα βήματα που πρέπει να ακολουθούνται, δείχνει την συνέπεια και την επανάληψη της διαδικασίας (iterative process) με σκοπό την συνεχή βελτίωση και τελειοποίηση στον βαθμό που είναι εφικτό, της ιστοσελίδας ώστε να αποδώσει τα μέγιστα για τον λόγο που διατίθεται στο διαδικτυακό κοινό. Πιο συγκεκριμένα, να καλύψει τις βασικές απαιτήσεις που καταγράφει μια μηχανή αναζήτησης, δηλαδή την εύρεση της πληροφορίας στο βέλτιστο δυνατό χρόνο, όγκο και ποιότητα.

Στον αντίποδα του SEO, καταγράφεται η διαφήμιση επί πληρωμή (Paid advertising), ως ένα παράλληλο μοντέλο ψηφιακού μάρκετινγκ. Ευρέως διαδεδομένη μέθοδος διαφήμισης επί πληρωμής στο διαδίκτυο, είναι το Pay-per-Click, κατά το οποίο ο διαφημιστής πληρώνει τον κάτοχο της ιστοσελίδας όταν γίνεται κλικ στην διαφήμιση. Αν και φαίνεται αποτελεσματική η μέθοδος, αξίζει να σημειωθεί, ότι χωρίς τη σταθερή και συνεχή βελτιστοποίηση της ιστοσελίδας, προς διαφήμιση, δεν θα κατατάσσεται σε καλή-υψηλή θέση στην κατάταξη των μηχανών αναζήτησης, όπως είναι η Google. Και αυτό πολύ απλά, γιατί αν τα επίπεδα ευχρηστίας μιας ιστοσελίδας είναι χαμηλά, τότε η πιθανότητα ο επί πληρωμή χρήστης να εισέλθει σε αυτή και να φύγει κατευθείαν χωρίς διάδραση, αυξάνονται σημαντικά (Drivas et al. 2021). Άρα, γίνεται αντιληπτό η σημασία του SEO και στο κομμάτι της προώθησης της ιστοσελίδας μέσω διαφημιστικών μεθόδων και προσπαθειών.

Η βελτίωση του SEO των ιστοσελίδων έχει με τη σειρά του σαν αποτέλεσμα την δημιουργία καλύτερου AI-Generated content ή αλλιώς περιεχομένου που δημιουργείται από μια αλγοριθμική διαδικασία και όχι από ανθρώπους-συγγραφείς. Για την παραγωγή του

“αυτοδημιούργητου” περιεχομένου, ο εκάστοτε χρήστης εισάγει λέξεις-κλειδιά και ένα θέμα στο ερώτημα του προς το σύστημα τεχνητής νοημοσύνης-AI (Artificial Intelligence) που χρησιμοποιεί. Στη συνέχεια, το AI περιηγείται στο διαδίκτυο αναζητώντας -σχετικές με το ερώτημα- πληροφορίες και ανατροφοδοτεί τον χρήστη με το περιεχόμενο που δημιούργησε. Αυτός ο τύπος περιεχομένου έχει γίνει όλο και πιο διαδεδομένος στο SEO λόγω της ικανότητάς του να παράγει μεγάλες ποσότητες περιεχομένου γρήγορα και ανέξοδο από οικονομικής σκοπιάς. Ωστόσο, μπορεί να έχει αρνητικό αντίκτυπο στο SEO με διάφορους τρόπους. Bailyn (2023)

Αρχικά, το AI-Generated content μπορεί να επηρεάσει αρνητικά την κατάταξη στις μηχανές αναζήτησης. Οι μηχανές αναζήτησης, όπως η Google, χρησιμοποιούν πολύπλοκους αλγόριθμους για να προσδιορίσουν τη συνάφεια (relevance) την ποιότητα περιεχομένου (content quality) αλλά και την μοναδικότητα (uniqueness) των ιστοσελίδων. Επιπροσθέτως, συχνά είναι χαμηλής ποιότητας και δεν διαθέτει την τεχνογνωσία και την αυθεντία που αναζητούν οι μηχανές αναζήτησης κατά την κατάταξη των σελίδων. Ως αποτέλεσμα, οι ιστοσελίδες με τον συγκεκριμένο τρόπο παραγωγής ενδέχεται να κατατάσσονται χαμηλότερα στα αποτελέσματα αναζήτησης από ό,τι οι σελίδες με ποιοτικό και πρωτότυπο περιεχόμενο.

Είναι σημαντικό να αναφερθεί, ότι το περιεχόμενο των AI μοντέλων είναι συχνά κακογραμμένο και δυσανάγνωστο. Το γεγονός αυτό, μπορεί να οδηγήσει σε υψηλά ποσοστά αναπήδησης και χαμηλή δέσμευση του χρήστη στην αλληλεπίδραση του με το περιεχόμενο. Ως εκ τούτου είναι λογικό, οι χρήστες να είναι πιο πιθανό να παραμείνουν σε μια ιστοσελίδα αν το περιεχόμενο είναι καλογραμμένο, ενημερωτικό και ελκυστικό. Το auto-generative content συχνά δεν είναι τίποτα από αυτά. Τα παραπάνω είναι πιθανόν να οδηγήσουν τον χρήστη σε μία δύσχρηστη και ανούσια εν τέλει εμπειρία χωρίς να ικανοποιεί την πληροφοριακή, ψυχαγωγική ή/και καταναλωτική του ανάγκη (Wu et al. 2023).

Παραδείγματα της λογικής του AI-Generated content στην πραγματική ζωή είναι περιγραφές προϊόντων που παράγονται αυτόματα με βάση τις προδιαγραφές του προϊόντος ή κριτικές που παράγονται αυτόματα με βάση λέξεις-κλειδιά. Αυτοί οι τύποι περιεχομένου είναι συχνά χαμηλής ποιότητας και στερούνται των αποχρώσεων και των λεπτομερειών που μπορούν να προσφέρουν οι ανθρώπινοι συγγραφείς (Cao et al. 2023).

Για να αποφευχθεί η δημιουργία ελλιπούς και εσφαλμένου περιεχομένου πρέπει να ενταθούν οι προσπάθειες για SEO, ώστε να γίνει εφικτή η παραγωγή υψηλής ποιότητας, πρωτότυπου περιεχομένου που είναι ενημερωτικό, ελκυστικό και σχετικό με το κοινό-στόχο τους. Θα πρέπει επίσης να αποφεύγουν τη χρήση εργαλείων που παράγουν περιεχόμενο

αυτόματα ή που βασίζονται σε μεγάλο βαθμό στη συμπλήρωση λέξεων-κλειδιών. Συμπερασματικά, το AI-Generated content μπορεί να έχει αρνητικό αντίκτυπο στο SEO λόγω της χαμηλής ποιότητας και της κακής εμπειρίας του χρήστη. Οι ερευνητές του SEO θα πρέπει να δίνουν προτεραιότητα στην ποιότητα έναντι της ποσότητας όταν πρόκειται για την παραγωγή περιεχομένου για τους ιστότοπους τους. Με τον τρόπο αυτό, μπορούν να βελτιώσουν την κατάταξη στις μηχανές αναζήτησης και να παρέχουν στους χρήστες μια καλύτερη εμπειρία.

5.2 Βασικά Συμπεράσματα

Παρατηρώντας τα ευρήματα μας, γίνεται αντιληπτή η ανάγκη της συνεχούς βελτίωσης του SEO των ιστοσελίδων αλλά και των μεθόδων ανάκτησης των αποτελεσμάτων των μηχανών αναζήτησης ακόμα κι όταν μιλάμε για την δημοφιλέστερη μηχανή αναζήτησης της Google αλλά και την αμέσως επόμενη, Bing. Πιο αναλυτικά, η Google είχε καλύτερη βαθμολογία από την Bing στις τέσσερις βασικές μετρικές που ορίσαμε (Total Website Score, SEO, Speed, Security). Ωστόσο, η διαφορά των δύο δεν ήταν τόσο μεγάλη, όσο περιμέναμε. Εντύπωση επίσης, προκαλεί το ποσοστό αλληλοκάλυψης των αποτελεσμάτων των δύο μηχανών αναζήτησης, το οποίο ανέρχεται μόλις στο 5%.

5.3 Προβλήματα έρευνας

Η έρευνα μας παρά την επίτευξη σημαντικών αποτελεσμάτων, αντιμετώπισε ζητήματα στην βιβλιογραφική στελέχωση της. Πιο συγκεκριμένα, κατά την αναζήτηση προγενέστερων ερευνών μελετών σχετικά με τη σύγκριση Bing και Google εντοπίσαμε μικρό αριθμό σχετικών επιστημονικών άρθρων. Ως εκ τούτου, η ανάπτυξη της μεθοδολογίας μας ξεκίνησε να υπάρχει έστω μία ξεκάθαρη μέθοδος με προδιαγεγραμμένα βήματα μεθοδολογία για την αρχική καθοδήγηση από τα βιβλιογραφικά πεπραγμένα του σχετικού αντικειμένου για την σύγκριση μεταξύ των δύο μηχανών αναζήτησης. Οι δυσκολίες που αναφέρθηκαν αμβλύθηκαν και εν τέλει αντιμετωπίστηκαν με προσωπική έρευνα και αναζήτηση των κατάλληλων μεθόδων επεξεργασίας και ανάλυσης των δεδομένων καθώς και εργαλείων που υποστήριξαν τη μελέτη μας. Αφενός στο πλαίσιο της ανάπτυξης των τεσσάρων βημάτων (βλέπε εικόνα 14), αφετέρου κατά την διαδικασία της ανάλυσης των δεδομένων μέσω μεθόδων περιγραφικής στατιστικής και ανάλυσης παλινδρομήσεων.

Είναι αναγκαίο, επίσης, να τονιστεί η δυσκολία στην επιλογή των όρων αναζήτησης, ώστε να υπάρχει αντικειμενική προσέγγιση, χωρίς υποκειμενικές επιλογές από την πλευρά μας. Το γεγονός αυτό, οδήγησε στην επιλογή των όρων από το Google Trends προσθέτοντας, με

αυτόν τον τρόπο, αξιοπιστία αλλά και αντικειμενικότητα στον ερευνητικό μας σκοπό. Πιο συγκεκριμένα, επιλέχθηκαν οι όροι που ήρθαν πρώτοι σε αναζήτηση από τους χρήστες σε παγκόσμιο επίπεδο το 2021.

5.4 Μελλοντικές επεκτάσεις / Πρακτικές Προεκτάσεις της Έρευνας

Η έρευνα μας, κάλυψε στο επίπεδο της ανάλυσης της παλινδρόμησης ένα συγκεκριμένο μέρος, χρησιμοποιώντας ως βάση τα αποτελέσματα των υπομετρικών της μετρικής Total Website Score. Αν και επετεύχθη ο στόχος της έρευνας κρίνουμε, ότι μελλοντικά μπορεί να διευρυνθεί ακόμα περισσότερο με την παλινδρόμηση να πραγματοποιείται με τα δεδομένα των υπόλοιπων υπομετρικών των μετρικών SEO, Speed, Security. Παρόμοιες προσεγγίσεις έχουν ήδη αναπτυχθεί για τον κλάδο των Βιβλιοθηκών, Αρχείων και Μουσείων (Drivas et al. 2021; (Mannheimer et al. 2019).

Αξίζει, επίσης, να αναφερθεί, ότι μελλοντικά η συγκεκριμένη ερευνητική προσπάθεια μπορεί να αποτελέσει βάση για παρόμοιες συγκρίσεις δεδομένων π.χ. αντί για την χρησιμοποίηση των 10 δημοφιλέστερων όρων του 2021, θα μπορούσε να είναι η σύγκριση των 10 δημοφιλέστερων τεχνολογικών εταιριών, καθώς και η σύγκριση των 10 πρώτων εγχώριων τηλεοπτικών καναλιών αξιολογώντας τα αποτελέσματα των μηχανών αναζήτησης, ως μέσο διάδοσης και αναπαραγωγής του περιεχομένου τους.

Αξίζει να σημειωθεί ότι η μεθοδολογία που προτείνεται, θα μπορούσε να αξιοποιηθεί σε ακόμα μεγαλύτερο εύρος δεδομένων. Πιο συγκεκριμένα μέσω της ανάκτησης μεγάλων δεδομένων, τα αποτελέσματα των δύο μηχανών θα μπορούσαν να αξιολογηθούν ως προς τη σχετικότητα τους όπως αντίστοιχα πράττεται σε ετήσια βάση μέσω των συνέδριων του TREC. Ένα βήμα παρακάτω, αυτές οι ιστοσελίδες θα μπορούσαν να αξιολογηθούν και ως προς την απόδοση τους, ενισχύοντας περισσότερο την κατάθεση και παρουσίαση ενδεδειγμένων πρακτικών SEO στον παγκόσμιο Ιστό.

Τέλος, η έρευνα μπορεί να διευρυνθεί στο στάδιο της επιλογής των μηχανών αναζήτησης. Πιο συγκεκριμένα, θεωρούμε εφικτή την πρόσθεση κι άλλων μηχανών αναζήτησης. Με μοναδική επιφύλαξη τον σημαντικά αυξημένο όγκο δεδομένων που θα χρειαστούν να διαχειριστούν οι μελλοντικοί ερευνητές.

Βιβλιογραφικές Αναφορές

- Almkhtar, F., Mahmood, N., & Kareem, S. (2021). Search engine optimization: a review. *Applied Computer Science*, 17(1).
- Arcoya, E. (2021, April 30). Τι είναι το Google Trends. *Actualidad eCommerce*. <https://www.actualidadecommerce.com/el/que-es-google-trends/>
- Arlitsch, K., & O'Brien, P. S. (2012). Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar. *Library Hi Tech*.
- Arlitsch, K., O'Brien, P., & Rossmann, B. (2013). Managing search engine optimization: An introduction for library administrators. *Journal of Library Administration*, 53(2-3), 177-188.
- Bailyn, E. (2021, March 10). How Generative AI Will Impact SEO. LinkedIn. <https://www.linkedin.com/pulse/how-generative-ai-impact-seo-evan-bailyn/>
- Barbar, A., & Ismail, A. (2019, April). Search engine optimization (SEO) for websites. In *Proceedings of the 2019 5th international conference on computer and technology applications* (pp. 51-55).
- Bar-Ilan, J. (2004). Search engine ability to cope with the changing web. In *Web dynamics* (pp. 195-215). Springer, Berlin, Heidelberg.
- Bedeian, A. G., & Mossholder, K. W. (1994). Simple question, not so simple answer: Interpreting interaction terms in moderated multiple regression. *Journal of Management*, 20(1), pp. 159-165
- Britvic, I., Duric, J., & Bužić, D. (2014, May). Comparative analysis of Google and Bing SEO on leading Croatian news portals. In *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 474-478). IEEE.
- Cahill, K., & Chalut, R. (2009). Optimal results: what libraries need to know about Google and search engine optimization. *The Reference Librarian*, 50(3), 234-247.
- Cai, T. T., & Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics*, 34(5), pp. 2159-2179
- Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. arXiv preprint arXiv:2303.04226.

- Chakravarty, R., & Wasan, S. (2015). Webometric Analysis of Library Websites of Higher Educational Institutes (HEIs) of India: A Study through Google Search Engine. *DESIDOC Journal of Library & Information Technology*, 35(5).
- Cohen, J. (2016). The earth is round ($p < .05$). What if there were no significance tests?, In *What If There Were No Significance Tests?* pp. 69-82
- Colton, J. A., & Bower, K. M. (2002). Some misconceptions about R2. *International Society of Six Sigma Professionals, EXTRAOrdinary Sense*, 3(2), 20-22.
- Davalas, A. (2018, November 15). Πώς επηρεάζεται η κατάταξη της google από τη χρησιμότητα και τη συμπεριφορά των επισκεπτών - HTML.GR. HTML.GR.
<https://html.gr/%cf%80%cf%8e%cf%82-%ce%b5%cf%80%ce%b7%cf%81%ce%b5%ce%ac%ce%b6%ce%b5%cf%84%ce%b1%ce%b9-%ce%b7-%ce%ba%ce%b1%cf%84%ce%ac%cf%84%ce%b1%ce%be%ce%b7-%cf%84%ce%b7%cf%82-google-%ce%b1%cf%80%cf%8c-%cf%84%ce%b7/>
- Drivas, I. C., Sakas, D. P., Giannakopoulos, G. A., & Kyriaki-Manessi, D. (2021). Optimization of paid search traffic effectiveness and users' engagement within websites. In *Business Intelligence and Modelling: Unified Approach with Simulation and Strategic Modelling in Entrepreneurship 8th* (pp. 17-30). Springer International Publishing.
- Drivas, I., Kouis, D., Kyriaki-Manessi, D., & Giannakopoulos, G. (2021). Content management systems performance and compliance assessment based on a data-driven search engine optimization methodology. *Information*, 12(7), 259.
- Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological bulletin*, 81(6), 358.
- Edosomwan, J., & Edosomwan, T. O. (2010). Comparative analysis of some search engines. *South African Journal of Science*, 106(11), 1-4.
- Fahimnia, F., & Eltemasi, M. (2021). Comparative analysis of Iranian medical academic libraries websites the base Google SEO component. *The Journal of Academic Librarianship*, 47(4), 102354.
- French, R. B., & Fagan, J. C. (2019). The visibility of authority records, researcher identifiers, academic social networking profiles, and related faculty publications in search engine results. *Journal of Web Librarianship*, 13(2), 156-197.
- Goel, P. K., & DeGroot, M. H. (1980). Only normal distributions have linear posterior expectations in linear regression. *Journal of the American Statistical Association*, 75(372), 895-900.

- Helland, I. S. (1987). On the interpretation and use of R² in regression analysis. *Biometrics*, pp. 61-69
- Höchstötter, N., & Lewandowski, D. (2009). What users see—Structures in search engine results pages. *Information Sciences*, 179(12), 1796-1812.
- JASP - A Fresh Way to Do Statistics. (2023, May 30). JASP - Free and User-Friendly Statistical Software. <https://jasp-stats.org/>
- Kisiel, R. (2010). Dealers get on top of search engine results. *Automotive News*, 84(6408), 24-25.
- Koster, M. (1995). Robots in the Web: threat or treat?. *Oil Spectrum*, 2(9), 8-18
- Krstić, N., & Masliković, D. (2018). Pain points of cultural institutions in search visibility: the case of Serbia. *Library Hi Tech*, 37(3), 496-512.
- Leach, L. F., & Henson, R. K. (2007). The use and impact of adjusted R² effects in published regression research. *Multiple Linear Regression Viewpoints*, 33(1), pp. 1-11
- Lewandowski, D., Wahlig, H., & Meyer-Bautor, G. (2006). The freshness of web search engine databases. *Journal of information science*, 32(2), 131-148.
- Li, X., Wong, W., Lamoureux, E. L., & Wong, T. Y. (2012). Are linear regression techniques appropriate for analysis when the dependent (outcome) variable is not normally distributed?. *Investigative ophthalmology & visual science*, 53(6), 3082-3083.
- Maneejuk, P., & Yamaka, W. (2021). Significance test for linear regression: how to test without Pvalues?. *Journal of Applied Statistics*, 48(5), pp. 827-845
- Mannheimer, S., Clark, J. A., Espeland, J., & Hagerman, K. (2019). Building a Dataset Search for Institutions: Project Update. *Publications*, 7(2), 29.
- Mavridis, T., & Symeonidis, A. L. (2015). Identifying valid search engine ranking factors in a Web 2.0 and Web 3.0 context for building efficient SEO mechanisms. *Engineering Applications of Artificial Intelligence*, 41, 75-91.
- Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, 22(1), 67.
- Moriarty, J. P. (2008). A theory of benchmarking (Doctoral dissertation, Lincoln University).

- Muthoni, J. (2021, June 14). 10 Key Benefits Of SEO For Your Business. Forbes. <https://www.forbes.com/sites/forbesagencycouncil/2021/06/14/10-key-benefits-of-seo-for-your-business/?sh=6e8d37a53fd0>
- Onaifo, D., & Rasmussen, D. (2013). Increasing libraries' content findability on the web with search engine optimization. *Library Hi Tech*.
- Pay Per Click vs. Search Engine Optimization - Google Ads. (n.d.). Pay Per Click vs. Search Engine Optimization - Google Ads. Pay per Click Vs. Search Engine Optimization - Google Ads. <https://ads.google.com/home/resources/seo-vs-ppc/>
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2(1), 21-33.
- Ren, Y., Yang, D., & Diao, X. (2010). Market segmentation strategy in internet market. *Physica A: Statistical Mechanics and its Applications*, 389(8), 1688-1698
- Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of clinical epidemiology*, 98, 146-151.
- Shahzad, A., Nawi, N. M., Sutoyo, E., Naeem, M., Ullah, A., Naqeeb, S., & Aamir, M. (2018). Search engine optimization techniques for malaysian university websites: a comparative analysis on google and bing search engine. *International Journal on Advanced Science, Engineering and Information Technology*, 8(4), 1262-1269.
- Spanos, A. (1995). On normality and the linear regression model. *Econometric Reviews*, 14(2), pp. 195- 203
- Tavosi, M., & Naghshineh, N. (2021, May). An Analysis of Iranian University Library Websites from Standpoint Five Effective Factors on Google SEO: Iranian University Library Websites and Google SEO. In 2021 7th International Conference on Web Research (ICWR) (pp. 306-310). IEEE.
- TechTarget. (n.d.). Pros and cons of AI-generated content. Retrieved May 29, 2023, from <https://www.techtarget.com/whatis/feature/Pros-and-cons-of-AI-generated-content>
- *Text REtrieval Conference (TREC) Home Page*. (n.d.). <https://trec.nist.gov/>
- Von Hippel, P. T. (2005). Mean, median, and skew: Correcting a textbook rule. *Journal of statistics Education*, 13(2).
- Wikipedia contributors. (2023). JASP. Wikipedia. <https://en.wikipedia.org/wiki/JASP>
- Wu, J., Gan, W., Chen, Z., Wan, S., & Lin, H. (2023). Ai-generated content (aigc): A survey. arXiv preprint arXiv:2304.06632.

Παράρτημα

5.5 Δεδομένα έρευνας

Δείγμα δεδομένων της έρευνας στα ανοικτά αποθετήρια ερευνητικών δεδομένων Zenodo και Kaggle.

The screenshot shows the Kaggle interface for the dataset 'Search Engines Comparison and Websites Performance'. The left sidebar contains navigation options like 'Create', 'Home', 'Competitions', 'Datasets', 'Models', 'Code', 'Discussions', 'Learn', 'More', 'Your Work', and 'RECENTLY EDITED'. The main content area includes a search bar, dataset title, and options for 'Data Card', 'Code (0)', 'Discussion (0)', and 'Settings'. There are buttons for 'New Notebook' and 'Download (34 kB)'. The description states the dataset consists of 200 search results from Google and Bing engines. A 'Provide feedback on this dataset' section includes buttons for 'Learning 0', 'Research 0', and 'Application 0'. Below that, there are buttons for 'Well-documented 0', 'Well-maintained 0', 'Clean data 0', 'Original 0', 'High-quality notebooks 0', and 'Other'. At the bottom, there is a 'Data Explorer' section showing the file 'Google_Bing_Comparisons_Websites_Performance.xlsx (38.75 kB)'.

Εικόνα 25: Kaggle, πηγή: <https://www.kaggle.com/datasets/georgiosntimo/search-engines-comparison-and-websites-performance>

The screenshot shows the Zenodo dataset page for 'Search Engines Comparison and Websites Performance'. The header includes the Zenodo logo, a search bar, and buttons for 'Upload' and 'Communities'. The date 'June 30, 2023' is displayed, along with 'Dataset' and 'Open Access' tags. The title 'Search Engines Comparison and Websites Performance' is prominently displayed. Below the title, the author 'Ntimo, Georgios; Ntararas, Vasilios' is listed. The description states the dataset consists of 200 search results from Google and Bing engines. A 'Contact Persons' section lists Vasilis Ntararas, Georgios Ntimo, and Ioannis C. Drivas. A table shows the file 'Google_Bing_Comparisons_Websites_Performance.xlsx' with a size of 38.8 kB and a download button. The table also includes the MD5 hash 'md5:f8e0d6720e9c1306b4a18aa867ad0d9e'. At the bottom, there is a 'Citations' section showing 0 citations.

Εικόνα 26: Zenobo, πηγή: <https://zenodo.org/record/8102700>