



Πτυχιακή Εργασία

Κατάταξη Αποτελεσμάτων στις Μηχανές Αναζήτησης. Μια Συγκριτική Μεθοδολογική Προσέγγιση μεταξύ της Google και της Bing

Νταραράς Βασίλειος - Παναγιώτης & Ντίμο Γεώργιος



Επιβλέπων: Δημήτριος Κουής
Τριμελής Επιτροπή: Δημήτριος Κουής,
Σαράντος Καπιδάκης και Ιωάννης Τριανταφύλλου

Αθήνα, Ιούλιος 2023

Περιεχόμενα Παρουσίασης

I. Εισαγωγή

- Θεωρητικό Υπόβαθρο και Σχετικές Μελέτες

II. Σκοπός και Επιμέρους Στόχοι Μελέτης

III. Μεθοδολογία Έρευνας

- Το Σύνολο των Βημάτων της Μεθοδολογίας
- Επίπεδο 1 – Προσδιορισμός της Πηγής Ανάκτησης Δεδομένων
- Επίπεδο 2 – Ανάκτηση και Προ-επεξεργασία
- Επίπεδο 3 – Μέθοδοι Ανάλυσης

IV. Ενδεικτικά Αποτελέσματα

V. Συζήτηση – Συμπεράσματα Έρευνας

- Θεωρητική Συνεισφορά
- Πρακτική συνεισφορά
- Προεκτάσεις Έρευνας

I. Εισαγωγή - Θεωρητικό Υπόβαθρο και Σχετικές Μελέτες

➤ Σημαντικότητα μηχανών αναζήτησης

- Βέλτιστη δυνατή ποσότητα αποτελεσμάτων, στον βέλτιστο δυνατό χρόνο, στην βέλτιστη ποιότητα (ποιότητα = κατάταξη)
- Κάθε χρήστης υποβάλλει τουλάχιστον 4 όρους αναζήτησης σε Google ή Bing καθημερινά



Εισαγωγή - Θεωρητικό Υπόβαθρο και Σχετικές Μελέτες

Μελέτες σύγκρισης μηχανών αναζήτησης

- **Ο Lewandowski, κ.ά (2006):**
 - Χρησιμοποίησαν τις μηχανές αναζήτησης Google, Yahoo και MSN
 - Ερεύνησαν τη συχνότητα με την οποία, οι μηχανές αναζήτησης ενημερώνουν τους δείκτες τους
 - Συμπέραναν ότι η ποιότητα των διαφόρων ευρετηρίων των μηχανών αναζήτησης ποικίλλει
 - ότι θα πρέπει να χρησιμοποιούνται περισσότερες από μία μηχανές κατά την αναζήτηση τρέχοντος περιεχομένου.
- **Οι Höchstötter και Lewandowski (2009):**
 - Χρησιμοποίησαν τις μηχανές αναζήτησης Google, Yahoo, Live.com και Ask
 - Ερεύνησαν 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
 - η συγκεκριμένη μελέτη βασίστηκε σε 42.758 στοιχεία
- **Οι Edosomwan (2010):**
 - Χρησιμοποίησαν τις μηχανές αναζήτησης Google, Yahoo, AlltheWeb, Gigablast, Zworks, AltaVista και Bing/MSN
 - Ερευνήσαν την επίδοση της ανάκτησης πληροφοριών των μηχανών αναζήτησης χρησιμοποιώντας 10 διαφορετικά ερωτήματα διαφορετικής πολυπλοκότητας
 - Η μελέτη δείχνει ότι η Google είναι η καλύτερη μηχανή αναζήτησης σε ακρίβεια (70%) και χρόνο απόκρισης (2 δευτερόλεπτα).

II. Σκοπός και Επιμέρους Στόχοι Μελέτης

Σκοπός:

- Κατανόηση των παραγόντων που επηρεάζουν την απόδοση των ιστοσελίδων σε Google και Bing
- Ανακαλύπτοντας αυτούς τους παράγοντες σε μια ιστοσελίδα, βελτιώνονται και οι ίδιες οι μηχανές (Search Engine Optimization)

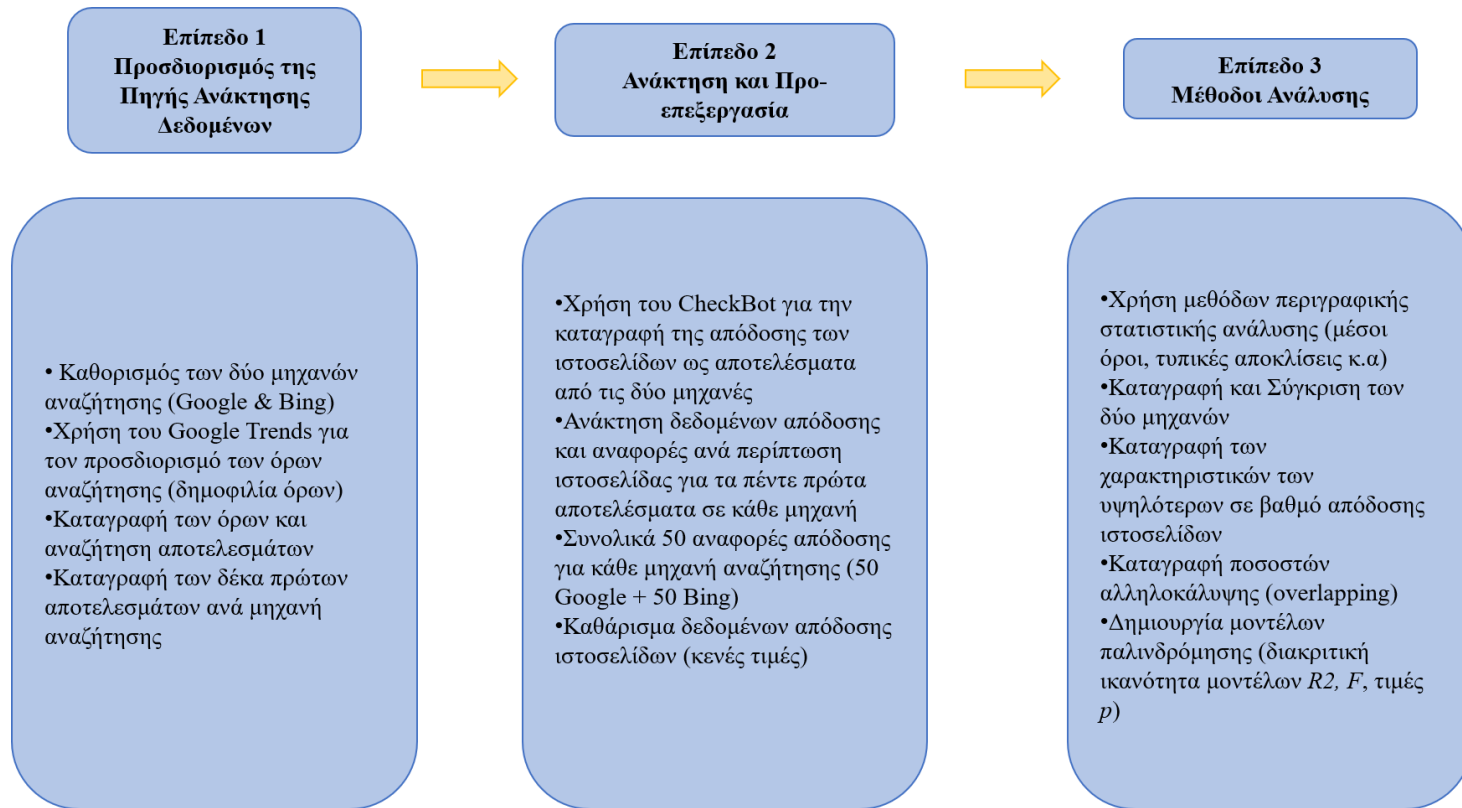
Επιμέρους Στόχοι:

- Σύγκριση των αποτελεσμάτων Google και Bing μέσω αξιολόγησης πλήθους μετρικών.
- Βαθμός αλληλοκάλυψης των αποτελεσμάτων μεταξύ των μηχανών αναζήτησης.
 - Μία ίδια σειρά κατάταξης μιας ιστοσελίδας σημαίνει ότι τηρεί τις βέλτιστες πρακτικές που προτείνουν και οι ίδιες οι μηχανές.
- Δημιουργία μεθοδολογικής προσέγγισης ανάκτησης, επεξεργασίας, ανάλυσης και ερμηνείας των αποτελεσμάτων ανά μηχανή αναζήτησης και προτεραιοποίησης μετρικών με μεγαλύτερη απήχηση

Μεθοδολογία έρευνας

III. Μεθοδολογία Έρευνας I

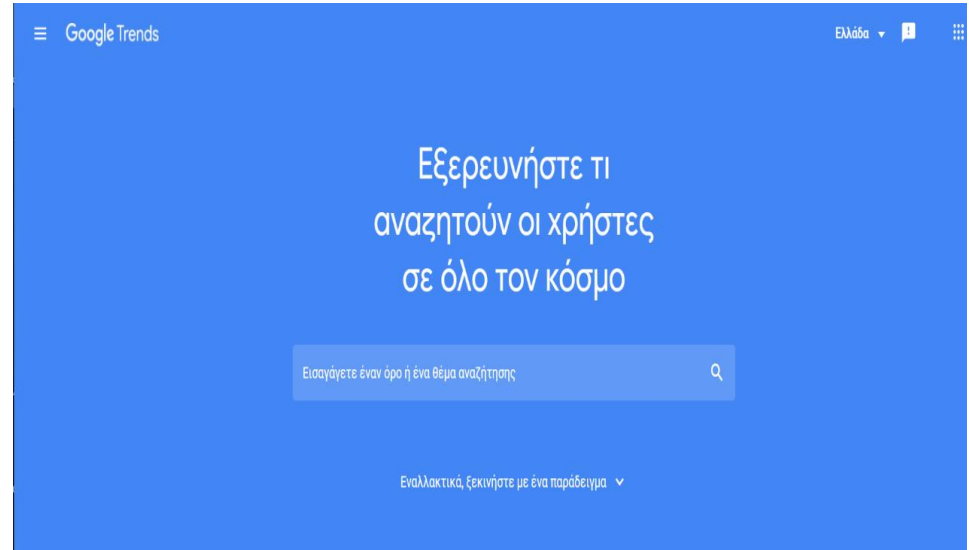
Το Σύνολο των Βημάτων της Μεθοδολογίας



Μεθοδολογία Έρευνας II

Επίπεδο 1 – Προσδιορισμός της Πηγής Ανάκτησης Δεδομένων

- I. Καθορισμός των δύο μηχανών αναζήτησης (Google & Bing)
- II. Χρήση του Google Trends για τον προσδιορισμό των δημοφιλών όρων αναζήτησης
- III. Καταγραφή των όρων και αναζήτηση αποτελεσμάτων
- IV. Καταγραφή των δέκα πρώτων αποτελεσμάτων ανά μηχανή αναζήτησης

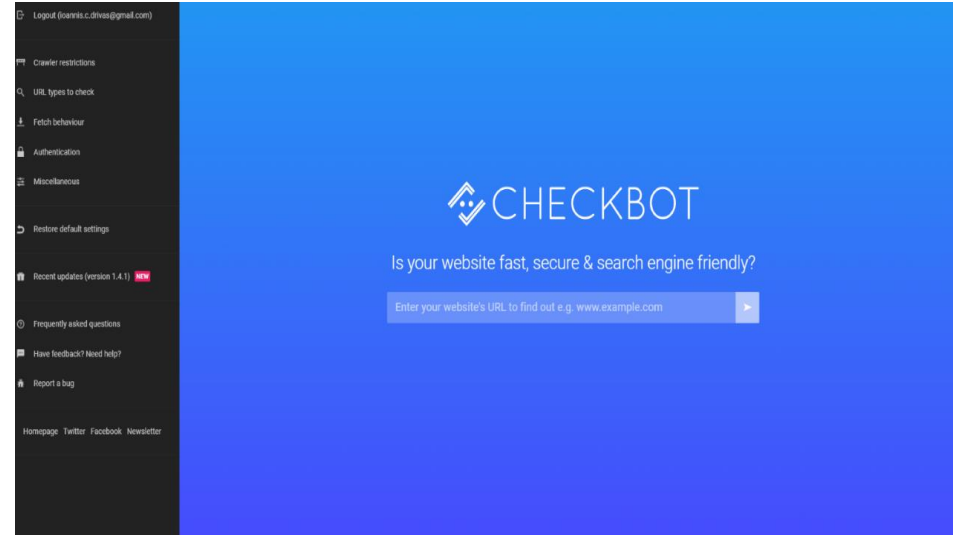


Περιβάλλον Google Trends

Μεθοδολογία Έρευνας III

Επίπεδο 2 – Ανάκτηση και Προ-επεξεργασία

- I. Χρήση του CheckBot για την καταγραφή της απόδοσης των ιστοσελίδων ως αποτελέσματα από τις δύο μηχανές
- II. Ανάκτηση δεδομένων απόδοσης και αναφορές ανά περίπτωση ιστοσελίδας για τα πέντε πρώτα αποτελέσματα σε κάθε μηχανή
- III. Συνολικά 50 αναφορές απόδοσης για κάθε μηχανή αναζήτησης (50 Google + 50 Bing)
- IV. Καθάρισμα δεδομένων απόδοσης ιστοσελίδων (Π.χ. κενές τιμές)

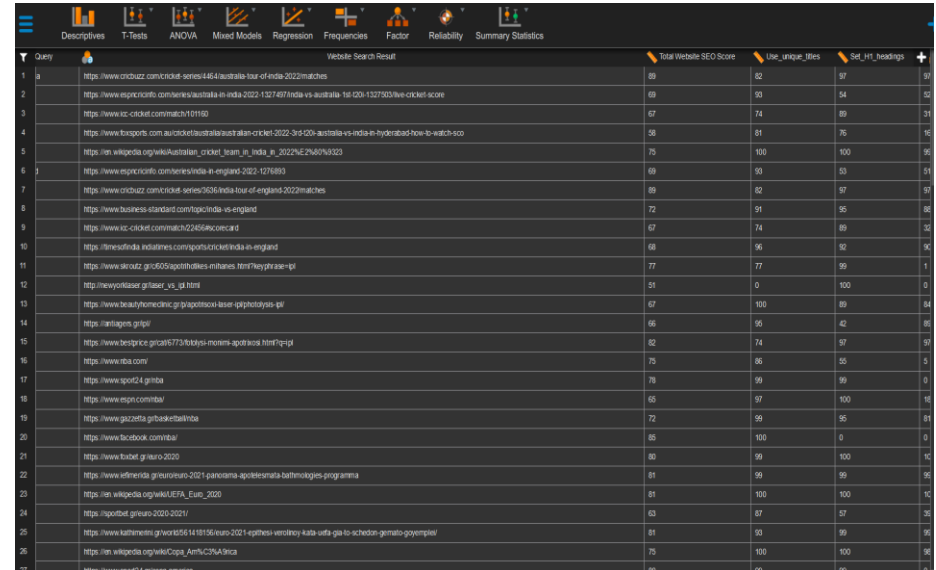


Περιβάλλον Checkbot

Μεθοδολογία Έρευνας IV

Επίπεδο 3 – Μέθοδοι Ανάλυσης

- I. Χρήση μεθόδων περιγραφικής στατιστικής ανάλυσης (μέσοι όροι, τυπικές αποκλίσεις κ.α)
- II. Καταγραφή και Σύγκριση των δύο μηχανών
- III. Καταγραφή των χαρακτηριστικών των υψηλότερων σε βαθμό απόδοσης ιστοσελίδων
- IV. Καταγραφή ποσοτών αλληλοκάλυψης (overlapping)
- V. Δημιουργία μοντέλων παλινδρόμησης (διακριτική ικανότητα μοντέλων R2, F, τιμές p)



| Query | Total Website SEO Score | User_unique_visits | Ref_Hit_headings |
|-------|-------------------------|--------------------|------------------|
| 1 | 89 | 82 | 97 |
| 2 | 69 | 93 | 54 |
| 3 | 67 | 74 | 89 |
| 4 | 98 | 81 | 76 |
| 5 | 75 | 100 | 100 |
| 6 | 69 | 93 | 53 |
| 7 | 89 | 82 | 97 |
| 8 | 72 | 51 | 95 |
| 9 | 67 | 74 | 89 |
| 10 | 68 | 95 | 92 |
| 11 | 77 | 77 | 99 |
| 12 | 51 | 0 | 100 |
| 13 | 67 | 100 | 89 |
| 14 | 66 | 95 | 42 |
| 15 | 82 | 74 | 97 |
| 16 | 75 | 86 | 55 |
| 17 | 78 | 99 | 99 |
| 18 | 65 | 97 | 100 |
| 19 | 72 | 99 | 95 |
| 20 | 85 | 100 | 0 |
| 21 | 80 | 99 | 100 |
| 22 | 81 | 99 | 99 |
| 23 | 81 | 100 | 100 |
| 24 | 63 | 87 | 87 |
| 25 | 81 | 93 | 99 |
| 26 | 75 | 100 | 100 |
| 27 | 66 | 66 | 66 |

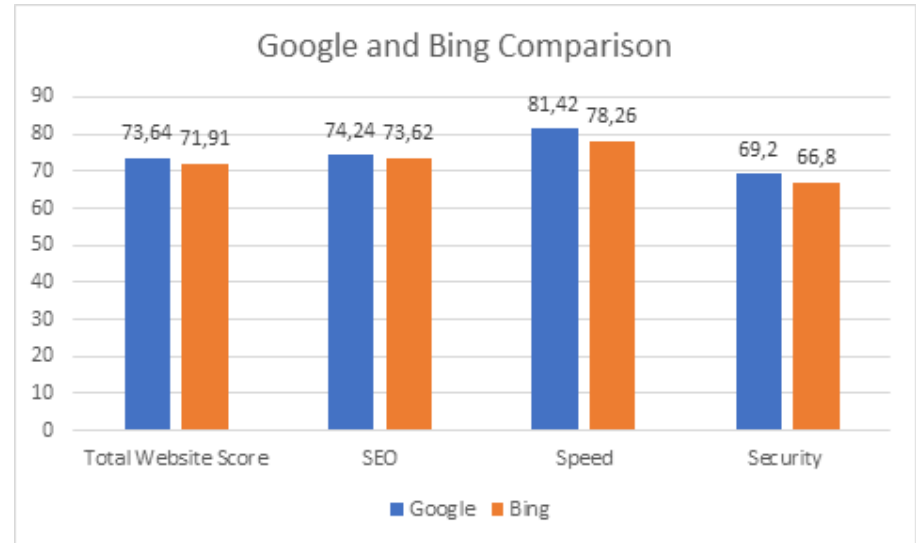
Περιβάλλον επεξεργασίας δεδομένων του JASP



Ενδεικτικά Αποτελέσματα

IV. Ενδεικτικά Αποτελέσματα I

- 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 5 πρώτων συνδέσμων κάθε όρου για καθεμία μηχανή αναζήτησης.
- Παρατηρείται προβάδισμα της Google έναντι της Bing και στις 4 κατηγορίες



Ενδεικτικά Αποτελέσματα II

Σύγκριση μηχανών αναζήτησης



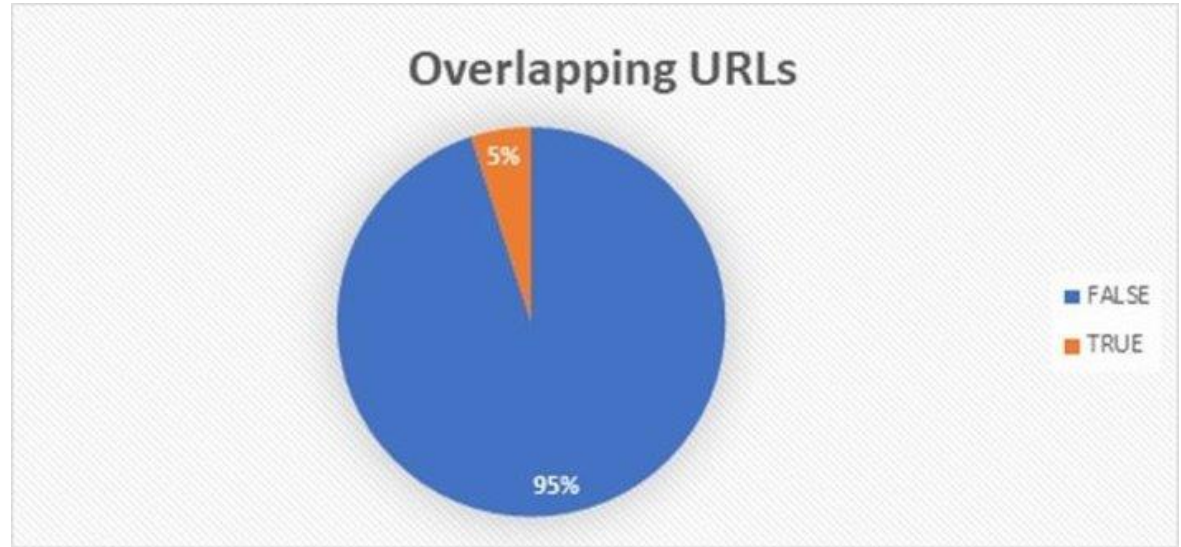
| Metrics | Google | Bing | Διαφορά |
|---------------------|--------|-------|---------|
| Total Website Score | 73.64 | 71.91 | 1.73 |
| SEO | 74.24 | 73.62 | 0.62 |
| Speed | 81.42 | 78.26 | 3.16 |
| Security | 69.2 | 66.8 | 2.4 |

Μέσος όρος των μετρικών απόδοσης ιστοσελίδων που ανακτήθηκαν μεταξύ των δύο μηχανών.

Ενδεικτικά Αποτελέσματα III

Αλληλοκάλυψη

- Πίτα ποσοστού αλληλοκάλυψης των αποτελεσμάτων Google και Bing
- Στα 100 αποτελέσματα που εμφάνισαν ξεχωριστά οι δύο μηχανές αναζήτησης, τα 5 αποτελέσματα ήταν ακριβώς τα ίδια, δηλαδή εμφάνισαν ίδιο σύνδεσμο και ίδια σειρά κατάταξης



Ενδεικτικά Αποτελέσματα IV

Αποτελέσματα Παλινδρόμησης (Μεταβλητές SEO)

- Για κάθε μια ποσοστιαία μονάδα που βελτιώνεται η μετρική Use Unique Titles, η συνολική βελτίωση της ιστοσελίδας (Total Website Score) αυξάνεται κατά 0.186
- Η μετρική Use Unique Headings εμφανίζει στατιστική σημαντικότητα ($p < .001$) και το μοντέλο έχει ικανότητα πρόβλεψης ($R^2 = 0.460$). Κάθε βελτίωση της μετρικής κατά μια μονάδα αυξάνει τη συνολική απόδοση της ιστοσελίδας κατά 0.149
- Για κάθε μια ποσοστιαία μονάδα που βελτιώνεται η μετρική Avoid Duplicate Page Content, η συνολική βελτίωση της ιστοσελίδας (Total Website Score) αυξάνεται κατά 0.241
- Για την μετρική Set Canonical URLs βλέπουμε ότι για κάθε μονάδα που βελτιώνεται η απόδοση της ιστοσελίδας μεταβληθεί κατά 0.111

| Metric | Predicted value of change | Adjusted R2 | P value |
|--|---------------------------|-------------|---------|
| Total Website Score (Constant) Use Unique Titles | 57.366 0.186 | 0.170 | 0.002 |
| Total Website Score (Constant) Use Unique Headings | 64.409 0.149 | 0.460 | <.001 |
| Total Website Score (Constant) Avoid Duplicate Page Content | 50.975 0.241 | 0.202 | <.001 |
| Total Website Score (Constant) Set Canonical URLs | 64.561 0.111 | 0.203 | <.001 |

Ενδεικτικά Αποτελέσματα V

Αποτελέσματα Παλινδρόμησης (Μεταβλητές Speed)

- Για κάθε μια ποσοστιαία μονάδα που βελτιώνεται η μετρική Use minification, η συνολική βελτίωση της ιστοσελίδας (Total Website Score) αυξάνεται κατά 0.228
- Η μετρική Avoid Render-Blocking JavaScript αυξάνεται κατά 0.109 κάθε φορά που έχουμε βελτιώσει κατά μια ποσοστιαία μονάδα
- Η μετρική Deeply Nested URLs δεν παρουσίασε στατιστική σημαντικότητα καθώς η τιμή P είναι 0.251.

| Metric | Predicted value of change | Adjusted R2 | P value |
|--|---------------------------|-------------|--------------|
| Total Website Score (Constant) Use minification | 53.419 0.228 | 0.186 | 0.001 |
| Total Website Score (Constant) Avoid Render-Blocking JavaScript | 68.812 0.109 | 0.360 | <.001 |
| Total Website Score (Constant) Deeply Nested URLs | 61.623 0.125 | 0.007 | 0.251 |

Ενδεικτικά Αποτελέσματα VI

Αποτελέσματα Παλινδρόμησης (Μεταβλητές Security)

- Για κάθε μια ποσοστιαία μονάδα που βελτιώνεται η μετρική Use HTTPS, η συνολική βελτίωση της ιστοσελίδας (Total Website Score) αυξάνεται κατά 0.500
- Για κάθε μια ποσοστιαία μονάδα που βελτιώνεται η μετρική Use Content Sniffing Protection, η συνολική βελτίωση της ιστοσελίδας αυξάνεται κατά 0.134
- Η μετρική Use Clickjack Protection αυξάνεται κατά 0.112 κάθε φορά που έχουμε βελτιώσει κατά μια ποσοστιαία μονάδα

| Metric | Predicted value of change | Adjusted R2 | P value |
|--|---------------------------|-------------|---------|
| Total Website Score (Constant) Use HTTPS | 24.149 0.500 | 0.143 | 0.004 |
| Total Website Score (Constant) Use Content Sniffing Protection | 66.103 0.134 | 0.553 | <.001 |
| Total Website Score (Constant) Use Clickjack Protection | 70.015 0.112 | 0.350 | <.001 |



Συζήτηση – Συμπεράσματα Έρευνας

V. Συζήτηση – Συμπεράσματα Έρευνας

Συνεισφορά Μελέτης - Θεωρητική Συνεισφορά

- Η έρευνα που πραγματοποιήθηκε αποτελεί μία από τις πιο πρόσφατες προσπάθειες να συγκριθούν οι Google & Bing.
- Τα σύγχρονα εργαλεία (Google Trends/Checkbot/JASP) που χρησιμοποιήθηκαν για την επιλογή, εξαγωγή, επεξεργασία και ανάλυση των δεδομένων διαφοροποιούν την έρευνα από τις προγενέστερες συγκριτικές μελέτες.
- Αξιοσημείωτο αποτέλεσμα: Σε ορισμένες περιπτώσεις η υψηλή απόδοση των ιστοσελίδων δεν ευθυγραμμίστηκε με την υψηλή τους κατάταξη. Ή αλλιώς κάποιες σελίδες με χαμηλή απόδοση ανήλθαν ψηλά στη κατάταξη.

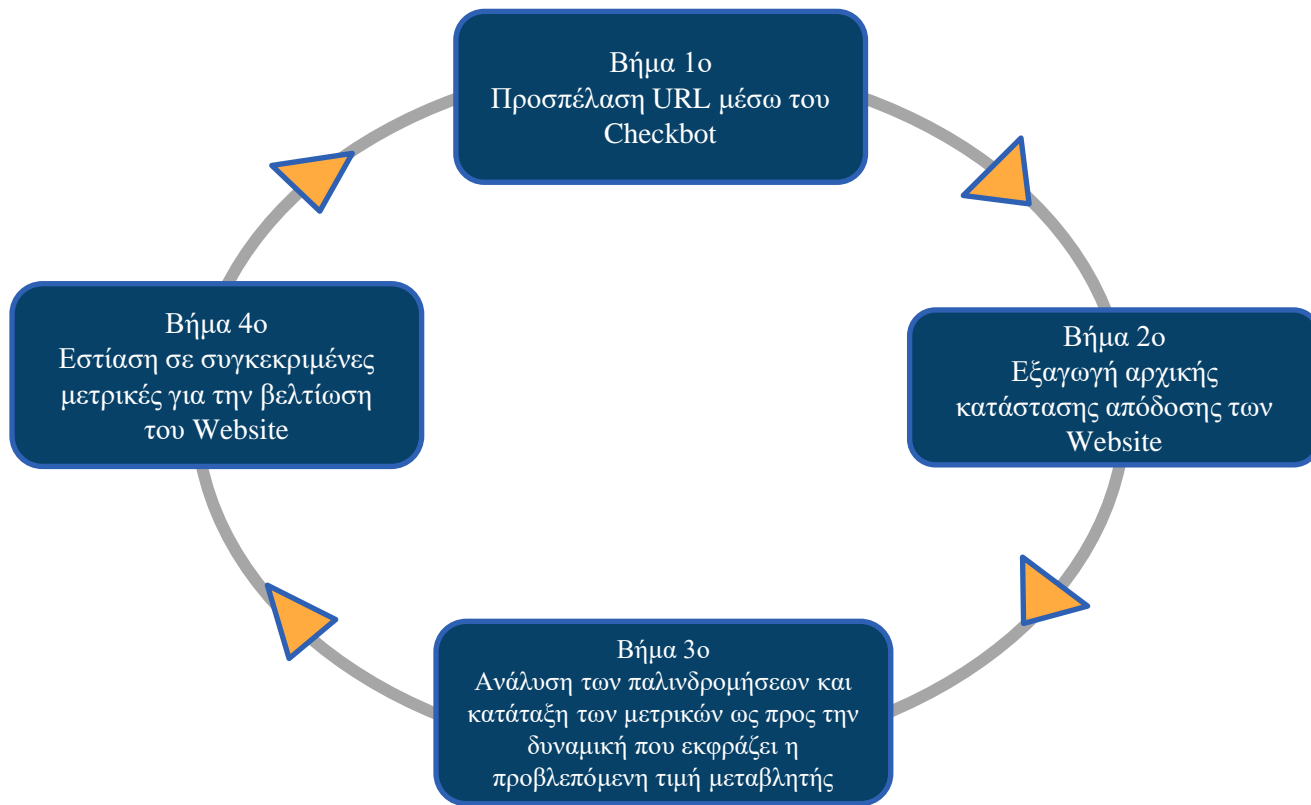
Συζήτηση – Συμπεράσματα Έρευνας

Συνεισφορά μελέτης – Πρακτική Συνεισφορά

- Η μεθοδολογία της συγκεκριμένης έρευνας αποτελεί ένα άρτιο παράδειγμα συγκριτικής προτυποποίησης (benchmarking).
- Το SEO οδηγεί μία ιστοσελίδα σε καλύτερη θέση κατάταξης από το paid advertising. (Aslam & Karjaluoto, 2017)
- Η βελτίωση του SEO παράγει καλύτερο AI-Generated content. (Bailyn, 2023)
- Δημιουργία βημάτων Ανάλυσης Απόδοσης και Βελτίωσης Ιστοσελίδων μέσω της μεθοδολογίας

Συζήτηση – Συμπεράσματα Έρευνας

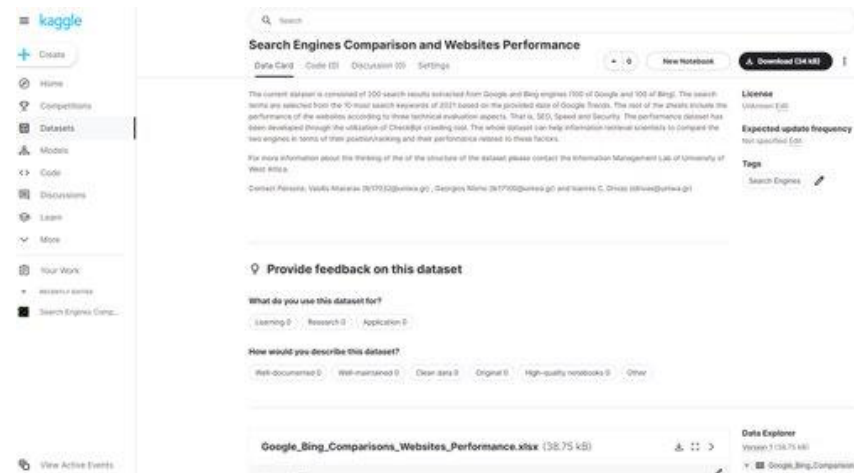
Βήματα Ανάλυσης Απόδοσης και Βελτίωσης Ιστοσελίδων



Συζήτηση – Συμπεράσματα Έρευνας

Προεκτάσεις Έρευνας

- Αύξηση του αριθμού των μηχανών αναζήτησης προς σύγκριση
- Αύξηση του δείγματος έρευνας ως προς τα ερωτήματα στις μηχανές αναζήτησης
- Η μεθοδολογική προσέγγιση που αναλύθηκε θα μπορούσε να χρησιμοποιηθεί ως βάση για άλλες συγκριτικές μελέτες
- Μέτρηση της απόστασης των ίδιων αποτελεσμάτων ανάμεσα στις δύο μηχανές (Distance ranking)



The screenshot shows the Kaggle dataset page for "Search Engines Comparison and Websites Performance". The page includes a search bar, a title, and a description of the dataset. The description states that the dataset is a collection of 100 search results extracted from Google and Bing engines, with 100 of Google and 100 of Bing. The search terms are selected from the 10 most search keywords of 2021 based on the provided data of Google Trends. The rest of the data includes the performance of the websites according to three technical evaluation aspects: Start, SEO, Speed and Security. The performance dataset has been developed through the utilization of Checkit v2 web tool. The whole dataset can help information retrieval scientists to compare the two engines in terms of their positioning and their performance related to these factors.

For more information about the thinking of the of the structure of the dataset please contact the Information Management Lab of University of West Attica.

Contact Person: Vasiliki Alexiou (valexio@uniwa.gr), Georgios Mihos (gmihos@uniwa.gr) and Ioannis C. Oikaridis (ioikaridis@uniwa.gr)

Provide feedback on this dataset

What do you use this dataset for?

Learning 0 Research 0 Application 0

How would you describe this dataset?

Well-documented 0 Well-maintained 0 Clear data 0 Original 0 High-quality notebooks 0 Other

Google_Bing_Comparisons_Websites_Performance.xlsx (38.75 kB)

Data Explorer
Version 2 (38.75 kB)

Δείγμα δεδομένων της έρευνας στα ανοικτό αποθετήριο ερευνητικών δεδομένων Kaggle.

VI. Πηγές Παρουσίασης – Ενδεικτική Βιβλιογραφία

Aslam, B., & Karjaluoto, H. (2017). Digital advertising around paid spaces, E-advertising industry's revenue engine: A review and research agenda. *Telematics and Informatics*, 34(8), 1650-1662.

Edosomwan, J., & Edosomwan, T. O. (2010). Comparative analysis of some search engines. *South African Journal of Science*, 106(11), 1-4.

Lewandowski, D., Wahlig, H., & Meyer-Bautor, G. (2006). The freshness of web search engine databases. *Journal of information science*, 32(2), 131-148.

Höchstötter, N., & Lewandowski, D. (2009). What users see–Structures in search engine results pages. *Information Sciences*, 179(12), 1796-1812.

Σας Ευχαριστούμε Πολύ!

Κατάταξη Αποτελεσμάτων στις Μηχανές Αναζήτησης.

Μια Συγκριτική Μεθοδολογική Προσέγγιση μεταξύ της Google και της Bing

Νταραράς Βασίλειος - Παναγιώτης & Ντίμο Γεώργιος



Επιβλέπων: Δημήτριος Κουής
Τριμελής Επιτροπή: Δημήτριος Κουής,
Σαράντος Καπιδάκης και Ιωάννης Τριανταφύλλου

Αθήνα, Ιούλιος 2023