



**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

**Μέθοδοι βελτιστοποίησης στις μηχανές αναζήτησης
(SEO) ιστοσελίδων οργανισμών πληροφόρησης**

Συγγραφέας

Δημήτριος Μανώλας (ΑΜ: 16042)

Επιβλέπων: Δημήτριος Κουής

Αθήνα, Σεπτέμβριος 2020

Ευχαριστίες – Αφιερώσεις

Θέλω να ευχαριστήσω τους επιβλέποντες καθηγητές μου τον κ. Κουή Δημήτριο και τον κ. Δρίβα Ιωάννη, που μου έδωσαν την ευκαιρία να ασχοληθώ με τη βελτιστοποίηση των ιστοσελίδων οργανισμών πληροφόρησης στις μηχανές αναζήτησης, ώστε να κάνω μια εμπειριστατωμένη έρευνα γύρω από το συγκεκριμένο θέμα και να αναπτύξω την πτυχιακή μου. Τους ευχαριστώ για τις χρήσιμες συμβουλές και παρατηρήσεις καθ' όλη τη διάρκεια εκπόνησης της εργασίας.

Επίσης, θα ήθελα να ευχαριστήσω τους γονείς μου για τη συμπαράσταση που μου παρείχαν κατά την διάρκεια των σπουδών μου στο Τμήμα Αρχειονομίας, Βιβλιοθηκονομίας & Συστημάτων Πληροφόρησης.

Σεπτέμβριος 2020

Μανώλας Δημήτριος

Περίληψη στα ελληνικά

Οι ιστοσελίδες αποτελούν πλέον ένα σημαντικό μέρος των οργανισμών πληροφόρησης. Μέσω αυτών παρουσιάζουν στο διαδίκτυο το υλικό και τις υπηρεσίες τους. Οι μηχανές αναζήτησης είναι το μέσο που ενώνει τους χρήστες με το επιθυμητό υλικό, επομένως είναι πολύ σημαντική η εμφάνιση των ιστοσελίδων στα πρώτα αποτελέσματα των αναζητήσεων. Η βελτιστοποίηση της σειράς εμφάνισης στις μηχανές αναζήτησης των ιστοσελίδων (γνωστό και ως SEO – Search Engine Optimization), με στόχο τον ευκολότερο εντοπισμό τους στα αποτελέσματα, επηρεάζει θετικά ή αρνητικά την ανακαλυψιμότητα (discoverability) και την ευρεσιμότητα (findability) του ψηφιακού περιεχομένου. Απώτερος στόχος της βελτιστοποίησης των αποτελεσμάτων στις μηχανές αναζήτησης είναι η προσέλκυση χρηστών και η ικανοποίηση των πληροφοριακών τους αναγκών με ακρίβεια. Παράλληλα με τη βελτιστοποίηση στις μηχανές αναζήτησης πρέπει να βελτιώνονται η πλοήγηση και γενικότερα η εμπειρία χρήσης των ιστοσελίδων. Σε αυτή τη κατεύθυνση θα εξεταστούν και τα θέματα δομής και διευκόλυνσης χρήσης (user experience) των ιστοσελίδων, καθώς διαδραματίζουν κομβικό ρόλο στην κατάταξη. Επομένως, μέσα από τη μελέτη και ανάλυση των δεδομένων κατάστασης των ιστοσελίδων με τη χρήση διαδικτυακών εργαλείων προτείνονται οι απαραίτητες διορθώσεις τόσο για τη βελτίωση ορισμένων χαρακτηριστικών που αφορούν τη δομή, τη ταχύτητα και την ασφάλεια, καθώς και σε ότι αφορά την ελαχιστοποίηση των ενεργειών που χρειάζεται να πραγματοποιούν οι χρήστες ή το χρόνου αναμονής, κατά τη διάρκεια της πλοήγησης, αυξάνοντας την προσβασιμότητα στις πηγές. Ως πεδίο εφαρμογής θα χρησιμοποιηθούν ιστοσελίδες βιβλιοθηκών, αρχειακών οργανισμών, μουσείων, ερευνητικών κέντρων και πανεπιστημιακών ιδρυμάτων από την Ελλάδα και το εξωτερικό.

Λέξεις Κλειδιά: Οργανισμοί πληροφόρησης, Βελτιστοποίηση ιστοσελίδων, Search Engine Optimization (SEO), Ευρεσιμότητα περιεχομένου (Content findability)

Περίληψη στα αγγλικά

Websites are an important part of information organizations. Through them organizations presented their content, news and related information to the public. Search engines are the medium that help users to discover the desired content. In this sense, it is very important that the most related web pages to the users' queries should appear in the first in the results. The optimization of the order of appearance in the search engines of the web pages (also known as SEO - Search Engine Optimization) improves locating the appropriate results and affects positively (or negatively) the discoverability and the findability of the digital content. The ultimate goal of search engine optimization is to attract users and meet their information needs accurately. Along with search engine optimization, the navigation and the user experience of the websites in general must be improved. In this direction, the issues of structure and usability (user experience) of the websites will be examined, as they play a key role in the ranking. Therefore, through the study and analysis of the websites with the use of internet tools, the necessary enhancements are proposed both for the improvement of some characteristics regarding the structure, the speed and the security, as well as for the minimization of the actions that are needed by users to access the resources. Websites of libraries, archival organizations, museums, research centers and universities from Greece and abroad will be used as field of application.

Keywords: Information organizations, Websites' optimization, Search Engine Optimization (SEO), Content findability

Περιεχόμενα

Ευχαριστίες – Αφιερώσεις	2
Περίληψη στα ελληνικά	3
Περίληψη στα αγγλικά	4
Κεφάλαιο 1. Εισαγωγή	8
1.1. Στόχοι της πτυχιακής εργασίας - Διάρθρωση	9
Κεφάλαιο 2. Θεωρητικό μέρος - Βιβλιογραφική επισκόπηση.....	11
2.1. Οργανισμοί πληροφόρησης και ιστοσελίδες πρόσβασης	11
2.2. Μηχανές αναζήτησης και βελτιστοποίηση ιστοσελίδων στις μηχανές αναζήτησης (SEO)	13
2.3. Δεδομένα συμπεριφοράς χρηστών.....	14
2.4. Η σημασία της βελτιστοποίησης των μηχανών αναζήτησης για τους οργανισμούς πληροφόρησης.....	15
2.5. Παρόμοιες προσπάθειες.....	16
2.6. Συμπεράσματα	20
Κεφάλαιο 3. Μεθοδολογία	21
3.1. Συνοπτική περιγραφή της εφαρμογής της μεθοδολογίας	21
3.2. Εργαλεία SEO.....	23
3.2.1. Checkbot.....	23
3.2.1.1. Μετρικές SEO	24
3.2.1.2. Μετρικές σχετικές με την ασφάλεια	27
3.2.1.3 Μετρικές σχετικές με την ταχύτητα	28
3.2.1.4. Τελικά σκορ	29
3.3. Άντληση δεδομένων χρηστών.....	30
3.3.1. SimilarWeb	30
3.4. Επιλογή δείγματος ιστοτόπων	31
3.4.1. Προσδιορισμός τύπου CMS - Whatcms.org.....	31
3.5. Στατιστικοί παράμετροι.....	32
3.6. Χρήση του JASP	33
Κεφάλαιο 4. Αποτελέσματα	35
4.1. Αποτελέσματα Metadata performance	35
4.2. Αποτελέσματα Architecture Structure performance	36
4.3. Αποτελέσματα Security performance	39

4.4. Αποτελέσματα Overall performance & size	41
4.4.1 Αποτελέσματα Overall performance & size με βάση τον τύπο CMS	43
4.5. Αποτελέσματα Behavioral Analytics	45
Κεφάλαιο 5. Ανακεφαλαίωση – Συμπεράσματα – Προτάσεις	47
Βιβλιογραφικές Αναφορές	51

Πίνακας Γραφημάτων

Γράφημα 1. Μέσος όρος και απόκλιση για το περιεχόμενο της ιστοσελίδας (Avoid thin content pages) – Σύγκριση ελληνικών και ξένων ιστοτόπων	37
Γράφημα 2. Μέσος όρος και απόκλιση για τη χρήση των παραμέτρων στις διευθύνσεις (Avoid URL parameters) – Σύγκριση ελληνικών και ξένων ιστοτόπων	37
Γράφημα 3. Μέσος όρος και απόκλιση για τα αρχεία robots.txt – Σύγκριση ελληνικών και ξένων ιστοτόπων	38
Γράφημα 4. Μέσος όρος και απόκλιση για τους χάρτες πλοήγησης (Set sitemap locations) – Σύγκριση ελληνικών και ξένων ιστοτόπων	39
Γράφημα 5. Μέσος όρος και απόκλιση για τη χρήση του https στις διευθύνσεις (Use HTTPS) – Σύγκριση ελληνικών και ξένων ιστοτόπων	40
Γράφημα 6. Μέσος όρος και απόκλιση για το συνολικό SEO (SEO) – Σύγκριση ελληνικών και ξένων ιστοτόπων	41
Γράφημα 7. Μέσος όρος και απόκλιση για τη συνολική ασφάλεια (Security) – Σύγκριση ελληνικών και ξένων ιστοτόπων	42
Γράφημα 8. Μέσος όρος και απόκλιση για το τις παραμέτρους του overall performance – Σύγκριση μεταξύ κατηγοριών CMS	44

Πίνακας Πινάκων

Πίνακας 1 Στοιχεία απόδοσης των μεταδεδομένων.....	36
Πίνακας 2 - Στοιχεία απόδοσης "αρχιτεκτονικής" των ιστοσελίδων.....	39
Πίνακας 3 - Στοιχεία απόδοσης της ασφάλειας.....	40
Πίνακας 4 - Στοιχεία της συνολικής απόδοσης & του μεγέθους των ιστοσελίδων.....	43
Πίνακας 5 - Στοιχεία της συνολικής απόδοσης & του μεγέθους των συστημάτων διαχείρισης περιεχομένου	45
Πίνακας 6 - Στοιχεία της απόδοσης δεδομένων συμπεριφοράς χρηστών.....	46

Κεφάλαιο 1. Εισαγωγή

Οι μέθοδοι που χρησιμοποιούνται στη πρόσβαση, την ανάκτηση και τη χρήση της πληροφορίας έχουν αλλάξει δραματικά από την εμφάνιση και τη καθιέρωση των μηχανών αναζήτησης (search engines). Πριν λίγες δεκαετίες, μία τυπική αναζήτηση της επιθυμητής πληροφορίας θα απαιτούσε την επίσκεψη του κοινού στη τοπική βιβλιοθήκη, τη συζήτηση με τον αρμόδιο βιβλιοθηκονόμο και τη προσεκτική αναζήτηση στον κατάλογο. Η έλευση των μηχανών αναζήτησης, ως το νέο μέσο αναζήτησης πληροφοριών, έφερε την επανάσταση στη διαδικασία αναζήτησης, εντοπισμού και ανάκτησης των πληροφοριών. Η επικράτηση των μηχανών αναζήτησης ως το κύριο σύστημα ανάκτησης πληροφοριών οφείλεται στην εύκολη και γρήγορη πρόσβαση σε πληθώρα πηγών, μειώνοντας το χρόνο και την ανάγκη της διαμεσολάβησης του επιστήμονα πληροφόρησης. Το 92% όλων των διαδικτυακών αναζητήσεων περιλαμβάνει τη χρήση των μηχανών για την εύρεση πληροφοριών (Purcell, 2011). Το 2010 διεξάγονταν πάνω από ένα δισεκατομμύριο αναζητήσεις ημερησίως μόνο στο Google, χωρίς να υπολογιστούν οι υπόλοιπες μηχανές αναζήτησης, όπως το Bing, η Yahoo κ.ά. (McGee, 2010), ενώ σήμερα ο αριθμός αυτός έχει ξεπεράσει τα 3,5 δισεκατομμύρια¹. Ο κεντρικός ρόλος λοιπόν που έχουν αποκτήσει οι μηχανές αναζήτησης στη διαδικασία εύρεσης και ανάκτησης έχει ως αποτέλεσμα την ανάγκη κατάταξης των ιστοσελίδων στα πρώτα αποτελέσματα (Search Engine Result Pages – SERPs) για ευκολότερη πρόσβαση.

Οι απαιτήσεις κατά το σχεδιασμό των ιστοσελίδων με τέτοιο τρόπο, ώστε να εμφανίζονται στα πρώτα αποτελέσματα όταν οι χρήστες αναζητούν μία συγκεκριμένη λέξη ή φράση, ορίζονται ως βελτιστοποίηση των ιστοσελίδων στις μηχανές αναζήτησης ή Search Engine Optimization (SEO) (Cahill & Chalut, 2009). Ο ρόλος του SEO είναι η υψηλότερη δυνατή κατάταξη στην εμφάνιση των οργανικών αποτελεσμάτων για μία συγκεκριμένη λέξη – κλειδί (keyword). Το SEO είναι μία ευρέως χρησιμοποιούμενη πρακτική, η οποία παρέχει υψηλό επίπεδο ορατότητας και πρόσβασης σε ιστοσελίδες στο διαδίκτυο. Η εσωτερική βελτιστοποίηση μιας ιστοσελίδας (On-page SEO) περιλαμβάνει το σχεδιασμό της, τα μεταδεδομένα, τις λέξεις - κλειδιά, τα ονόματα των σελίδων, τις εικόνες, τους συνδέσμους, το περιεχόμενο της κάθε σελίδας, τον πίνακα πλοήγησης (sitemap), το RSS Feeds κ.ά. Η εξωτερική βελτιστοποίηση μιας ιστοσελίδας (Off-page SEO) περιλαμβάνει τη προσθήκη της ιστοσελίδας στον πίνακα πλοήγησης, τη χρήση των μέσων κοινωνικής δικτύωσης και τη χρήση των συνδέσμων από άλλους βελτιστοποιημένους ιστοτόπους με σχετικό περιεχόμενο (Yalçin & Köse, 2010). Η γενικότερη ιδέα του SEO μπορεί να εξεταστεί υπό το πρίσμα των διαδικασιών της αναζήτησης και της ανάκτησης των πληροφοριών. Η αναζήτηση των πληροφοριών σχετίζεται με τη πρόσβαση και την εξερεύνηση των πηγών μέσω των μηχανών αναζήτησης. Από την άλλη πλευρά, η ανάκτηση περιλαμβάνει την αποθήκευση, τη προβολή, τον εντοπισμό της σχετικής πληροφορίας. Η εφαρμογή των μεθόδων

¹ Πηγή: <https://www.internetlivestats.com/google-search-statistics/>

βελτιστοποίησης πρέπει να συμβαδίζει με τις οδηγίες των μηχανών αναζήτησης, αλλά και τις πληροφοριακές ανάγκες και συμπεριφορές των χρηστών. Ο τρόπος παρουσίασης και εμφάνισης των αποτελεσμάτων επηρεάζουν σημαντικά τον τρόπο αναζήτησης. Οπότε όσο οι μηχανές αναζήτησης συνεχίζουν να έχουν σημαντικό ρόλο στη πληροφόρηση, το φαινόμενο της βελτιστοποίησης των ιστοσελίδων (SEO) πρέπει να εξεταστεί για την ανάδειξη του περιεχομένου των οργανισμών πληροφόρησης.

1.1. Στόχοι της πτυχιακής εργασίας - Διάρθρωση

Ο στόχος αυτής της εργασίας είναι να εξετάσει το φαινόμενο της βελτιστοποίησης των ιστοσελίδων οργανισμών πληροφόρησης στις μηχανές αναζήτησης, ως έναν μηχανισμό για τη βελτίωση της ανακαλυψιμότητας του ψηφιακού περιεχομένου στο διαδίκτυο. Εξετάζονται τεχνικά χαρακτηριστικά της δομής των ιστοσελίδων, καθώς και στοιχεία που αφορούν τη ταχύτητα, την ασφάλεια και κατ' επέκταση την εμπειρία περιήγησης.

Τα στοιχεία που συλλέχθηκαν, χρησιμοποιούνται για τον εντοπισμό των δυνατών και αδύνατων σημείων των επιμέρους ιστοσελίδων και τη βελτιστοποίησή τους σε σχέση με τα κριτήρια / προδιαγραφές που θέτουν οι μηχανές αναζήτησης. Η βελτιστοποίηση των στοιχείων θα τις βοηθήσει να καταταγούν σε υψηλότερες θέσεις και να προσφέρουν πληροφορίες σε περισσότερους χρήστες, εκπληρώνοντας με το καλύτερο δυνατό τρόπο τον σκοπό ύπαρξής τους. Εξετάζονται ορισμένα μόνο από τα χαρακτηριστικά ιστοσελίδων, τα οποία είναι τα εξής: ο τίτλος της σελίδας, το περιεχόμενο της ιστοσελίδας, η χρήση των παραμέτρων στις διευθύνσεις, τα αρχεία robots.txt, οι χάρτες πλοήγησης, οι ανιχνευμένες διευθύνσεις ιστοσελίδων, η χρήση του https στις διευθύνσεις, οι σελίδες ανά επίσκεψη συγκριτικά με το συνολικό SEO, τη συνολική ταχύτητα και τη συνολική ασφάλεια.

Η δομή της εργασίας αποτελείται από το θεωρητικό μέρος - βιβλιογραφική ανασκόπηση, στο οποίο εξετάζονται θέματα που αφορούν τις ιστοσελίδες πρόσβασης και τα είδη των οργανισμών. Στο συγκεκριμένο κεφάλαιο αναλύεται ο σκοπός και οι λειτουργίες του κάθε είδους οργανισμού, ενώ στη συνέχεια γίνεται αναφορά στο ρόλο των μηχανών αναζήτησης στη βελτιστοποίηση των ιστοσελίδων των οργανισμών. Παρακάτω γίνεται λόγος για τα δεδομένα συμπεριφοράς χρηστών, τα οποία υποδηλώνουν την ποιότητα της εμπειρίας περιήγησης. Η ανάλυση των μετρικών αυτής της κατηγορίας ενισχύει τη θέση των οργανισμών πληροφόρησης ως προς τη χρησιμότητα των ιστοσελίδων τους. Στη συνέχεια, αναφέρεται η σημασία των μηχανών αναζήτησης και η μεγάλη επιρροή που έχουν στη σειρά εμφάνισης των αποτελεσμάτων. Παρατίθενται επίσης παρόμοιες προσπάθειες, στις οποίες εξετάζονται ορισμένα χαρακτηριστικά των ιστοσελίδων. Αυτό το κεφάλαιο καταλήγει στο συμπέρασμα ότι δεν υπάρχει κάποιος προκαθορισμένος τρόπος με τον οποίο θα πρέπει να ενεργήσουν οι οργανισμοί πληροφόρησης.

Στο τρίτο κεφάλαιο παρουσιάζεται η μεθοδολογία που ακολουθήθηκε κατά τη διάρκεια εκπόνησης της εργασίας. Εντοπίστηκαν και εξετάστηκαν τα κύρια χαρακτηριστικά που οδηγούν σε υψηλή κατάταξη, ταχύτητα και ασφάλεια, καθώς και τα επικρατέστερα συστήματα διαχείρισης περιεχομένου. Η εφαρμογή των εργαλείων για την καταμέτρηση των παραμέτρων εξέτασε τη βελτιστοποίηση του περιεχομένου για πιο εύκολη, ακριβή και γρήγορη ανάκτηση. Τα βήματα για τη διεξαγωγή των αποτελεσμάτων που περιγράφονται είναι τα εξής: πρώτα προσδιορίστηκαν οι μετρικές που θα εξεταστούν και επιλέχθηκαν τα

κατάλληλα εργαλεία συλλογής και ανάλυσης δεδομένων. Έπειτα, περιγράφεται αναλυτικά η διαδικασία συγκέντρωσης των στοιχείων με τη παράλληλη χρήση των εργαλείων. Τέλος, αξιολογήθηκαν τα αποτελέσματα και πραγματοποιήθηκε η εξαγωγή των συμπερασμάτων. Περιγράφηκαν αναλυτικά τα εργαλεία και τα στοιχεία τους, ξεκινώντας από το Checkbot και τις μετρικές SEO, ταχύτητας και ασφάλειας. Έγινε αναφορά επίσης στα δεδομένα συμπεριφοράς των χρηστών και το εργαλείο SimilarWeb, ενώ στη συνέχεια αναλύθηκε και το διαδικτυακό εργαλείο προσδιορισμού των συστημάτων διαχείρισης περιεχομένου whatcms.org. Αναφέρονται οι συγκεκριμένες μετρικές που επιλέχθηκαν από το μεγάλο σύνολο των στοιχείων Checkbot, καθώς και τα στατιστικά στοιχεία. Η σημασία της χρήσης του JASP σημειώθηκε στη συνέχεια λόγω της μεγάλης συμβολής του στη διαδικασία ανάλυσης και ερμηνείας των μετρικών.

Στο τέταρτο κεφάλαιο γίνεται αναλυτική παρουσίαση των αποτελεσμάτων με τα δεδομένα όλων των μετρικών, κατηγοριοποιώντας τις ως εξής: μετρικές απόδοσης των μεταδεδομένων (Metadata performance), μετρικές απόδοσης της αρχιτεκτονικής δομής (Architecture structure performance), μετρικές απόδοσης τη ασφάλειας (Security performance), μετρικές συνολικής απόδοσης & μεγέθους (Overall performance & size), μετρικές για τη συμπεριφορά των χρηστών (Behavioral Analytics) και μετρικές για τα συστήματα διαχείρισης περιεχομένου (CMS). Χρησιμοποιούνται επίσης πίνακες και γραφήματα για τη καλύτερη παρουσίαση και κατανόηση των αποτελεσμάτων. Στο πέμπτο κεφάλαιο γίνεται ανακεφαλαίωση της εργασίας και διεξάγονται συμπεράσματα έπειτα από την αξιολόγηση των αποτελεσμάτων. Προτείνονται ορισμένες πρακτικές που έχουν εντοπιστεί στη βιβλιογραφία και πρόκειται να βοηθήσουν τους οργανισμούς πληροφόρησης στη βελτιστοποίηση των ιστοτόπων τους.

Κεφάλαιο 2. Θεωρητικό μέρος - Βιβλιογραφική επισκόπηση

Παρά το σημαντικό αντίκτυπο της βελτιστοποίησης των ιστοσελίδων στην εμφάνιση των αποτελεσμάτων, το ενδιαφέρον εφαρμογής της από τους οργανισμούς πληροφόρησης δεν είναι μεγάλο. Οι υπάρχουσες μελέτες που παρατίθενται παρακάτω αφορούν τη βελτιστοποίηση ιστοσελίδων γενικότερα, ενώ υπάρχουν και ορισμένες στο πλαίσιο των οργανισμών πληροφόρησης. Στη βιβλιογραφική επισκόπηση που ακολουθεί αναφέρονται οι ρόλοι και ο σκοπός του κάθε φορέα, η συμβολή των μηχανών αναζήτησης στη διαδικασία εντοπισμού και ανάκτησης της πληροφορίας, καθώς και η σημασία των δεδομένων συμπεριφοράς των χρηστών. Επίσης, παρατίθενται παρόμοιες έρευνες με κύριο θέμα τη βελτιστοποίηση ιστοσελίδων και την ενίσχυση της εμπειρία χρήσης (user experience).

2.1. Οργανισμοί πληροφόρησης και ιστοσελίδες πρόσβασης

Στην παρούσα πτυχιακή εργασία το κύριο αντικείμενο μελέτης είναι οι ιστοσελίδες οργανισμών πληροφόρησης, όπως οι περιπτώσεις των ακαδημαϊκών βιβλιοθηκών, οργανισμών διαχείρισης αρχείων, των μουσείων, των εθνικών βιβλιοθηκών, των ερευνητικών κέντρων και των πανεπιστημιακών ιδρυμάτων. Η ιστοσελίδα του κάθε οργανισμού πληροφόρησης έχει τα δικά της ιδιαίτερα χαρακτηριστικά, εξυπηρετεί συγκεκριμένο κοινό, προσφέρει διαφορετικό περιεχόμενο και υπηρεσίες εκπληρώνοντας τις αντίστοιχες πληροφοριακές ανάγκες των εν δυνάμει χρηστών.

Οι ακαδημαϊκές βιβλιοθήκες συνδέονται με το αντίστοιχο πανεπιστημιακό ίδρυμα στο οποίο ανήκουν και εξυπηρετούν δύο κύριους σκοπούς: την υποστήριξη των προγραμμάτων σπουδών των προπτυχιακών και μεταπτυχιακών φοιτητών και την υποστήριξη της ερευνητικής διαδικασίας. Η εξυπηρέτηση των αναγκών του κοινού είναι εξειδικευμένη, έχει υψηλό κόστος και συνήθως περιλαμβάνει πρόσκτηση υλικού μέσα από συμφωνίες με διεθνείς εκδοτικούς οίκους και μεγάλες βιβλιογραφικές βάσεις δεδομένων. Το περιεχόμενο της εκάστοτε ακαδημαϊκής βιβλιοθήκης ορίζεται από τα τμήματα που περιλαμβάνονται στο πανεπιστημιακό ίδρυμα και κατευθύνεται προς τη προώθηση της έρευνας. Η οργάνωση των συλλόγων γίνεται με βάση το επιστημονικό πεδίο. Στις ιστοσελίδες τους υπάρχει συνήθως κατάλογος ανοιχτής πρόσβασης (OPAC³) με τη δυνατότητα αναζήτησης με βάση τον τίτλο, τον συγγραφέα ή το θέμα. Η σύνθετη αναζήτηση με την εφαρμογή πολλαπλών πεδίων είναι η καλύτερη επιλογή για ακριβέστερα αποτελέσματα. Δίνονται επίσης πληροφορίες σχετικά με τις παρεχόμενες υπηρεσίες και σύνδεσμοι με το σύνολο των ενιαίων καταλόγων με συνδεδεμένες πηγές ενισχύοντας με αυτό τον τρόπο τη διαλειτουργικότητα. Η ύπαρξη πολλαπλών πηγών περιεχομένου στις ιστοσελίδες των βιβλιοθηκών σε συνδυασμό με τις ειδικές εφαρμογές (όπως βιβλιογραφικοί κατάλογοι και τα ψηφιακά αποθετήρια) μειώνουν την ευχρηστία και την ταχύτητα, αλλά είναι απολύτως αναγκαία για τη πρόσβαση στο κατάλληλο υλικό.

Οι αρχειακοί οργανισμοί είναι οργανισμοί πληροφόρησης, οι οποίοι παρέχουν πρόσβαση σε ιστορικά αρχεία στο ευρύτερο κοινό. Ο σκοπός της λειτουργίας τους είναι η

³ Online Public Access Catalog – Δημόσιος διαδικτυακός κατάλογος ανοιχτής πρόσβασης

διατήρηση του υλικού στο χρόνο σε φυσική ή ηλεκτρονική μορφή για την εξυπηρέτηση των μελλοντικών ερευνητών. Οι κύριοι χρήστες των υπηρεσιών τους είναι ερευνητές και ιστορικοί ή άτομα με εξειδικευμένα ενδιαφέροντα. Στην πτυχιακή εργασία συμπεριλαμβάνονται εθνικά αρχεία, ιστορικά, οπτικοακουστικά, χαρτογραφικά, τραπεζικών ιδρυμάτων και ακαδημαϊκά. Το κάθε αρχείο ανάλογα με την εξειδίκευση του εξυπηρετεί διαφορετικές ομάδες και επαγγελματίες. Το υλικό οργανώνεται σε πρώτο στάδιο με βάση το παραγωγό, έπειτα βάση της χρονολογίας ή του θέματος. Επίσης, υπάρχει διανοητικός και φυσικός διαχωρισμός σε υποκατηγορίες όπως είναι οι σειρές, οι υποσειρές, οι φάκελοι και οι υποφάκελοι. Στις ιστοσελίδες τους υπάρχει συνήθως κατάλογος με τη κατηγοριοποίηση του υλικού. Η ύπαρξη των καταλόγων διαδικτυακά διευκολύνει το έργο των ενδιαφερόμενων, ενώ η παροχή πληροφοριών για τις υπηρεσίες και τη φυσική θέση τεκμηρίων ή αρχειακών συνόλων είναι ανεκτίμητης αξίας. Η πλοήγηση στις ιστοσελίδες είναι πιο απλή και προσιτή χωρίς επιπρόσθετες λεπτομέρειες κάνοντας το πιο γρήγορο. Προσφέρουν ευρετήριο με ένα μέρος των ψηφιακών συλλογών να διατίθεται μέσω διαδικτύου.

Τα μουσεία ανήκουν επίσης στους οργανισμούς πληροφόρησης, αλλά έχουν ως απώτερο σκοπό τη συλλογή και διατήρηση φυσικών ιστορικών ή πολιτιστικών αντικειμένων, μνημείων ή χώρων. Τα αντικείμενα αυτά έχουν πολιτισμική, καλλιτεχνική, ιστορική και επιστημονική σημασία και η προβολή τους εξυπηρετεί το ευρύ κοινό. Στη παρούσα πτυχιακή εργασία περιλαμβάνονται εθνικά, ιστορικά, καλλιτεχνικά και ειδικού ενδιαφέροντος μουσεία. Το κάθε ένα από αυτά ανταποκρίνεται στις πληροφοριακές ανάγκες των ατόμων με το αντίστοιχο επιστημονικό υπόβαθρο και ενδιαφέρον. Η οργάνωση των συλλογών γίνεται με βάση τη χρονολογία ή τη θεματική ενότητα. Τα μουσεία έχουν κυρίως τουριστικό και πολιτιστικό χαρακτήρα αλλά και εκπαιδευτικό. Στις ιστοσελίδες τους υπάρχουν συνήθως προγράμματα εικονικής περιήγησης, φωτογραφικό υλικό με περιγραφές των αντικειμένων και βάσεις δεδομένων. Η πλοήγηση στις ιστοσελίδες τους θα πρέπει να είναι πιο ευχάριστη και η εμπειρία να βρίσκεται σε υψηλότερα επίπεδα σε αντίθεση με τους υπόλοιπους οργανισμούς. Η παροχή πληροφοριών για τις υπηρεσίες και η προβολή των εκθεμάτων διαδικτυακά δημιουργεί πιο άμεση σχέση με το κοινό.

Οι εθνικές βιβλιοθήκες ιδρύονται από το κράτος στο οποίο ανήκουν και έχουν πολιτιστικό και ενωτικό χαρακτήρα, καθώς σ' αυτές διασώζεται όλο το υλικό παραγωγής στα πλαίσια της επικράτειας. Σκοπός τους είναι η διάσωση του πνευματικού, καλλιτεχνικού και επιστημονικού περιεχομένου για τις επόμενες γενιές. Εξυπηρετεί όλους τους πολίτες της χώρας και τους ξένους επισκέπτες, αλλά στον εσωτερικό χώρο της. Ο δανεισμός είναι πιο σπάνιος διότι το υλικό θεωρείται υψηλής πολιτιστικής κληρονομιάς και αξίας. Περιλαμβάνουν συχνά σπάνια και πολύτιμα κειμήλια ή έργα μοναδικής σημαντικότητας. Η οργάνωση των συλλογών γίνεται κάθε φορά με βάση το θέμα. Στις ιστοσελίδες τους υπάρχει κατάλογος ανοικτής πρόσβασης (OPAC) με τη κατηγοριοποίηση του υλικού. Η δυνατότητα αναζήτησης με βάση τον τίτλο, τον συγγραφέα ή το θέμα είναι επίσης παρούσα. Ο μεγάλος όγκος των τεκμηρίων εμφανίζεται διαδικτυακά, όπως και οι παρεχόμενες υπηρεσίες.

Οι ιστοσελίδες των ερευνητικών κέντρων και των πανεπιστημίων, αν και δεν ανήκουν στη κατηγορία των οργανισμών πληροφόρησης, παρέχουν όμως πρόσβαση σε μεγάλο όγκο δεδομένων προς χρήστες που κατ' εξοχήν χρησιμοποιούν τις ιστοσελίδες βιβλιοθηκών, των αρχείων και των μουσείων (π.χ. καθηγητές, ερευνητές, φοιτητές κ.λπ.).

Τα ερευνητικά κέντρα είναι αυτόνομοι οργανισμοί που πραγματοποιούν έρευνα σε διάφορα επιστημονικά πεδία. Οι ερευνητές αποτελούν το προσωπικό των ερευνητικών κέντρων και είναι οι φορείς διεξαγωγής της έρευνας. Ο κύριος σκοπός των ερευνητικών κέντρων είναι η υποστήριξη της ερευνητικής διαδικασίας και η προώθηση της γνώσης. Τα ερευνητικά κέντρα οργανώνουν τις συλλογές και το υλικό τους με βάση τα εκάστοτε επιστημονικά πεδία. Στις ιστοσελίδες τους υπάρχει συνήθως κατάλογος ανοιχτής πρόσβασης με τη κατηγοριοποίηση του υλικού για την ευκολότερη και ακριβέστερη αναζήτηση και ανάκτηση.

Τα πανεπιστήμια είναι ιδρύματα τριτοβάθμιας εκπαίδευσης, τα οποία εξυπηρετούν τους προπτυχιακούς και μεταπτυχιακούς φοιτητές, το προσωπικό και τα μέλη ΔΕΠ⁴. Ο κύριος σκοπός τους είναι εκπαίδευση των φοιτητών, η επαγγελματική και τεχνική κατάρτιση τους και η εξέλιξη της έρευνας. Στις ιστοσελίδες τους υπάρχει πληθώρα πληροφοριών για όλα τα θέματα που μπορεί να απασχολήσει κάποιο από τα μέλη τους, καθώς και ο διαχωρισμός με βάση τις σχολές και τα τμήματα. Υπάρχει πληθώρα πληροφοριών σχετικά με τις υπηρεσίες και το περιεχόμενο των επιστημονικών αντικειμένων. Η έκταση των ιστοσελίδων τους είναι συνήθως μεγαλύτερη συγκριτικά με τις υπόλοιπες κατηγορίες οργανισμών πληροφόρησης.

2.2. Μηχανές αναζήτησης και βελτιστοποίηση ιστοσελίδων στις μηχανές αναζήτησης (SEO)

Οι μηχανές αναζήτησης έχουν γίνει το βασικό εργαλείο για την ανάκτηση πληροφοριών διαδικτυακά με ολοένα και περισσότερους χρήστες να τις χρησιμοποιούν για να ικανοποιήσουν τις πληροφοριακές τους ανάγκες. Με τη συνεχή αύξηση του αριθμού των σελίδων στο διαδίκτυο είναι δύσκολο κανείς να αποκτήσει και να διατηρήσει υψηλή θέση στην εμφάνιση των αποτελεσμάτων. Οι διαχειριστές των ιστοσελίδων οφείλουν να τις προσαρμόζουν με βάση τις αναζητήσεις που πραγματοποιούν οι χρήστες. Ο ανταγωνισμός για την κατάταξη στα πρώτα αποτελέσματα θα αυξάνεται με γρήγορο ρυθμό και οι χρήστες θα συνεχίσουν να χρησιμοποιούν τις μηχανές αναζήτησης ως το κύριο μέσο πρόσβασης στην πληροφορία.

Πολύ σημαντικό ρόλο στη βελτιστοποίηση ιστοσελίδων στις μηχανές αναζήτησης έχουν οι σωστές λέξεις κλειδιά (keywords). Σε μία έρευνα από τους Dimitroff και Zhang (2005) για τη χρησιμότητα των μεταδεδομένων στα αποτελέσματα των μηχανών αναζήτησης, σημειώνεται ότι η επανάληψη των κατάλληλων λέξεων κλειδιών μπορεί να βοηθήσει σημαντικά στην εμφάνιση των αποτελεσμάτων. Ωστόσο, η χρήση τους θα πρέπει να γίνει σε συγκεκριμένα πεδία και με τη συνιστώμενη συχνότητα για τη βέλτιστη ορατότητα (visibility) στα αποτελέσματα. Οι μεγαλύτερες μηχανές αναζήτησης, όπως είναι

⁴ ΔΕΠ: Διδακτικό και Ερευνητικό Προσωπικό

η Google, η Yahoo και το Bing, ανανεώνουν τους οδηγούς χρήσης και τις πολιτικές τους με απώτερο σκοπό τη καλύτερη εξυπηρέτηση των χρηστών. Τα αρχεία αυτά παρέχουν σημαντικές πληροφορίες, πρακτικές βελτιστοποιήσεις στις μηχανές αναζήτησης, τον τρόπο κατάταξης των ιστοσελίδων και τις μη επιθυμητές μεθόδους. Κάποια από τα στοιχεία των ιστοσελίδων είναι ευρέως γνωστά ότι βοηθούν στην υψηλότερη κατάταξη, αλλά είναι επιθυμητό να γνωρίζει κανείς τους γενικούς κανόνες που δημοσιεύει η εκάστοτε μηχανή αναζήτησης. Σύμφωνα με τους συγγραφείς, οι μη κερδοσκοπικοί οργανισμοί μπορούν να επωφεληθούν όπως και οι επιχειρήσεις από τη βελτιστοποίηση των ιστοσελίδων τους.

Η ευρετηρίαση που πραγματοποιούν οι μηχανές αναζήτησης είναι μία διαδικασία κατά την οποία συλλέγονται και αποθηκεύονται δεδομένα του περιεχομένου των ιστοσελίδων. Οι μηχανές αναζήτησης τα ανακτούν γρήγορα και με ακρίβεια σε σχέση με τις αναζητήσεις των χρηστών (Malaga, 2008). Οι ευρετηριασμένες σελίδες (Crawled URLs⁵) είναι ένας σημαντικός παράγοντας για τη καλύτερη κατάταξη. *“Οι ιστοσελίδες που εμφανίζουν τις σωστές λέξεις κλειδιά σε τρία πεδία μεταδεδομένων στο κώδικα (πεδίο τίτλου, θέματος και περιγραφής) επιτυγχάνουν καλύτερη απόδοση ορατότητας στα αποτελέσματα”* (Zhang & Dimitroff, 2005). Τα μεταδεδομένα περιγραφής είναι χρήσιμα διότι δίνουν τη δυνατότητα στις μηχανές να αποδώσουν σημασιολογικές έννοιες στο περιεχόμενο ιστοσελίδων και να τις κατατάσσουν ανάλογα.

2.3. Δεδομένα συμπεριφοράς χρηστών

Τα δεδομένα συμπεριφοράς χρηστών υποδεικνύουν την ποιότητα της εμπειρίας χρήσης (user experience) των ιστοσελίδων. Οι διάφορες μετρικές που περιλαμβάνονται, υποδεικνύουν το βαθμό σχετικότητας της αναζήτησης με το περιεχόμενο. Οι μετρικές είναι οι εξής: η μέση διάρκεια επίσκεψης των χρηστών (Visit Duration), ο μέσος αριθμός των σελίδων ανά επίσκεψη (Pages per visit), το ποσοστό εξόδου από την ιστοσελίδα (Bounce Rate), το ποσοστό των επισκεπτών από τα «οργανικά» αποτελέσματα των μηχανών αναζήτησης (Percent of Organic Traffic). Με τη πληθώρα ιστοσελίδων στο διαδίκτυο, η βελτιστοποίηση στις μηχανές αναζήτησης μπορεί να επιτευχθεί μέσα από την ανάλυση των δεδομένων συμπεριφοράς χρηστών και τη ταυτόχρονη αξιολόγηση του περιεχομένου των ιστοσελίδων (ποιότητα του περιεχομένου). Η απόδοση των ιστοσελίδων με βάση τα στοιχεία μεταδεδομένων, της αρχιτεκτονικής και της ασφάλειας σε συνδυασμό με την εμπειρία χρήσης μπορούν να μετατραπούν σε χρήσιμες πληροφορίες και να βελτιώσουν την ορατότητα και την ευρεσιμότητα (findability) τους στο διαδίκτυο.

Τα δεδομένα συμπεριφοράς χρηστών επεκτείνουν τις ευκαιρίες για περαιτέρω ανάλυση και εξαγωγή πολύτιμων συμπερασμάτων, χρήσιμα για τη διαμόρφωση των κατάλληλων μεθόδων βελτιστοποίησης ανά περίπτωση. Επιπλέον, ο όγκος πληροφοριών που περιλαμβάνεται και παρουσιάζεται στις ιστοσελίδες οργανισμών πληροφόρησης όπου είναι σχετικά μεγάλος και σε πολλαπλά μορφότυπα (κείμενο, εικόνες, αρχεία κ.λπ.) ενέχεται να επηρεάζει τα δεδομένα της συμπεριφοράς των χρηστών. Στην εργασία των

⁵ Το crawling είναι η διαδικασία κατά την οποία οι σελίδες ιστοτόπων ευρετηριάζονται από τα Googlebots, (γνωστά και ως robots, bots ή spiders) και προστίθενται στον κατάλογο της Google για πιο γρήγορη και ακριβή ανάκτηση.

Drivas, Sakas κ.ά. (2020) παρουσιάζονται μια σειρά από μετρικές αλληλεπίδρασης των χρηστών με το περιεχόμενο, οι οποίες συνιστούν τα δεδομένα των χρηστών ή όπως είναι γνωστά ως behavioral analytics. Ο μέσος αριθμός των σελίδων ανά επίσκεψη (Pages per visit) είναι η κύρια μετρική που υπολογίζεται στη συγκεκριμένη εργασία. Αναφέρεται στον αριθμό σελίδων που περιηγείται ο χρήστης κάθε φορά που επισκέπτεται την ιστοσελίδα (session). Είναι σημαντικό να σημειωθεί ότι η κάθε επίσκεψη λήγει μετά από 30 λεπτά παραμονής στον ιστότοπο και ξεκινάει καινούρια. Οπότε, το ενδεχόμενο μικρής απόκλισης είναι πιθανό να παρουσιαστεί και να εξεταστεί συγκριτικά και με τα υπόλοιπα στοιχεία.

2.4. Η σημασία της βελτιστοποίησης των μηχανών αναζήτησης για τους οργανισμούς πληροφόρησης

Η πλήρης υιοθέτηση των κατάλληλων τεχνικών βελτιστοποίησης αυξάνει την ευρεσιμότητα στο διαδίκτυο και διασφαλίζει ότι το ψηφιακό περιεχόμενο είναι εύκολα προσβάσιμο από τους χρήστες μέσω των μηχανών αναζήτησης. Αυτό μπορεί να επιτευχθεί μόνο μέσω της υψηλής κατάταξης στα αποτελέσματα. Αρκετές έρευνες έχουν εξετάσει την εφαρμογή μεθόδων βελτιστοποίησης στο πλαίσιο των ιστοσελίδων οργανισμών πληροφόρησης προσφέροντας διαφορετικές προσεγγίσεις επί του θέματος. Οι Cahill και Chalut στην εργασία τους με τίτλο “Optimal Results: What Libraries Need to Know About Google and Search Engine Optimization” (2009) συζητούν για την ανάγκη που έχει προσωπικό των βιβλιοθηκών να γνωρίσει τις βασικές τεχνικές βελτιστοποίησης και το αντίκτυπο που έχουν στην εμφάνιση των αποτελεσμάτων. Θα είναι σε θέση να βοηθήσει τους χρήστες να κατανοήσουν το φαινόμενο και να είναι πιο αποτελεσματικοί κατά τις μελλοντικές τους αναζητήσεις. Η βασική γνώση της λειτουργίας των μηχανών αναζήτησης είναι απαραίτητη υπό το πρίσμα της κοινής υπόθεσης ότι τα αποτελέσματα βασίζονται στην αντικειμενική αξιολόγηση του περιεχομένου των ιστότοπων. Όπως ο Grimmelmann (2011) σωστά σημειώνει, η κατάταξη των αποτελεσμάτων δεν είναι απόλυτα αντικειμενική, καθώς οι μηχανές αναζήτησης έχουν τη δυνατότητα να απαγορεύουν ιστοσελίδες που παραβιάζουν τις πολιτικές τους. Οι μηχανές αναζήτησης έχουν επίσης πλήρη έλεγχο στο σχεδιασμό των αλγορίθμων της ευρετηρίασης (indexing) των σελίδων και συνεπώς μεγάλη επιρροή στη σειρά εμφάνισης.

Η έρευνα από τους Rushton, Kelehan και Strong (2008) φαίνεται να είναι η πιο σχετική με τη βελτιστοποίηση ιστοσελίδων οργανισμών πληροφόρησης. Σ’ αυτή γίνεται περιγραφή των εμπειριών της δημιουργίας και της εφαρμογής του έργου βελτιστοποίησης των ιστοσελίδων ακαδημαϊκών βιβλιοθηκών του Πανεπιστημίου Binghamton. Η σημαντικότητα της συγκεκριμένης έρευνας έγκειται στο γεγονός ότι βασίζεται σε πραγματικές εμπειρίες εφαρμογής των μεθόδων. Αν και η έρευνα περιλαμβάνει τέσσερις μόνο ιστοσελίδες, η περιγραφή των διαδικασιών παρέχει χρήσιμες λεπτομέρειες. Μερικοί από τους επικείμενους στόχους της έρευνας ήταν η αύξηση του αριθμού των σελίδων ανά επίσκεψη (pages per visit), των μοναδικών χρηστών (unique visitors), η αύξηση της συνδεσιμότητας με μη εμπορικές ιστοσελίδες, η υψηλότερη κατάταξη από τους ανταγωνιστές και ο εντοπισμός τεχνικών προβλημάτων που μειώνουν τη βελτιστοποίηση των ιστοσελίδων. Παρόμοιες προσπάθειες έχουν γίνει και για τη βελτιστοποίηση του περιεχομένου των ψηφιακών αποθετηρίων και βιβλιοθηκών από τους DeRidder (2008) και Arlitsch - O’Brien (2012). Ενώ ο DeRidder υποστηρίζει τη χρήση στατικών ιστοσελίδων και

τη δημιουργία ποιοτικών εξερχόμενων συνδέσμων (external links) σε βάσεις δεδομένων, οι Arlitsch - O'Brien προτείνουν τη μετατροπή των στοιχείων που αφορούν τα μεταδεδομένα από Dublin Core σε πιο φιλικά σχήματα ως προς τις μηχανές αναζήτησης (βλέπε παρακάτω σχετικά με την πρωτοβουλία schema.org).

Στην εργασία “Digital libraries: to meet or manage user expectations” οι Bawden και Vilar (2006) δίνουν ιδιαίτερη προσοχή στη μεγάλη επικράτηση των μηχανών αναζήτησης για την εκπλήρωση των πληροφοριακών αναγκών έναντι των βιβλιοθηκών. Χρησιμοποιείται ο όρος “infobesity” (obesity στα αγγλικά σημαίνει παχυσαρκία) για να περιγραφεί η συμπεριφορά των χρηστών, οι οποίοι “καταναλώνουν” πληθώρα ανούσιων και μη σχετικών πληροφοριών κατά την αναζήτηση τους. Η εμφάνιση υπερβολικά μεγάλου αριθμού αποτελεσμάτων περιλαμβάνει χαμηλής ποιότητας υλικό το οποίο είναι δύσκολο να επεξεργαστεί και να γίνει κατανοητό. Οι χρήστες προτιμούν τις μηχανές αναζήτησης έναντι των ιστοσελίδων βιβλιοθηκών διότι τους παρέχεται μεγαλύτερος αριθμός αποτελεσμάτων γρήγορα σε συνδυασμό με απεριόριστη πρόσβαση στις πηγές. Αντίθετα, οι ιστοσελίδες βιβλιοθηκών μπορούν να παρέχουν πιο ακριβή αποτελέσματα, αλλά με τη χρήση πιο σύνθετων εργαλείων αναζήτησης. Το γεγονός αυτό δυσκολεύει το έργο του μέσου χρήστη και προτείνεται από τους συγγραφείς η εκπαίδευση του σε ζητήματα αναζήτησης, εντοπισμού και ανάκτησης της πληροφορίας. Επίσης, προτείνεται η καλύτερη σχεδίαση των ιστοσελίδων προκειμένου να είναι πιο φιλική και να ενισχύει τη ποιότητα της εμπειρίας του χρήστη (user experience) που με έμμεσο τρόπο θα βελτιώσει τη κατάταξη των αποτελεσμάτων.

2.5. Παρόμοιες προσπάθειες

Παρότι η βελτιστοποίηση των ιστοσελίδων στις μηχανές αναζήτησης είναι σημαντική στη διαδικασία ανάκτησης της πληροφορίας, με το συγκεκριμένο φαινόμενο έχουν ασχοληθεί ελάχιστοι στο πλαίσιο των ιστοσελίδων οργανισμών πληροφόρησης. Παρακάτω παρατίθενται οι εργασίες οι οποίες παρουσιάζουν διάφορες μετρικές και μεθόδους βελτιστοποίησης του περιεχομένου για την αύξηση της ευρεσιμότητας.

Σε μία έρευνα από τους συγγραφείς Onaifo και Rasmussen (2013) πραγματοποιείται η αλλαγή της συμπεριφοράς στην αναζήτηση της πληροφορίας λόγω της εμφάνισης των μηχανών αναζήτησης. Οι μηχανές αναζήτησης, ως το νέο μέσο εντοπισμού και ανάκτησης, προσφέρουν άμεση πρόσβαση σε μεγάλο όγκο πληροφοριών χωρίς την διαμεσολάβηση του ανθρώπινου παράγοντα. Ο κεντρικός ρόλος τους στην απόκτηση της πληροφορίας έχει καταστήσει αναγκαία την υψηλή κατάταξη στις σελίδες των αποτελεσμάτων (Search Engine Result Pages – SERPs). Ο σκοπός της έρευνας αυτής είναι να εξετάσει τη διαδικασία της βελτιστοποίησης (SEO) ως ένα μέσο για την αύξηση της χρησιμότητας των ιστοσελίδων βιβλιοθηκών. Η ευρεσιμότητα (findability) του περιεχομένου των ιστοσελίδων εξαρτάται άμεσα από την σειρά εμφάνισης τους στα αποτελέσματα και γι’ αυτό τον λόγο πρέπει να δοθεί ιδιαίτερη προσοχή από τις βιβλιοθήκες. Στη συγκεκριμένη εργασία εξετάζονται μερικά χαρακτηριστικά των ιστοσελίδων βιβλιοθηκών του Καναδά που επηρεάζουν σημαντικά την κατάταξη στα αποτελέσματα, όπως η παγκόσμια κατάταξη (global traffic rank), δηλαδή οι υπερσυνδέσεις (backlinks) από έναν άλλο ιστότοπο, η φήμη (reputation), δηλαδή ο αριθμός των εξερχόμενων ή εισερχόμενων συνδέσμων σε συγκεκριμένη ιστοσελίδα, το σύνολο των ευρετηριασμένων σελίδων από τις μηχανές αναζήτησης

(indexed pages) και οι συνολικές ημερήσιες προβολές (daily page views). Μετά την εφαρμογή στατιστικών μεθόδων, καταλήγουν στο συμπέρασμα ότι υπάρχει άμεση σχέση μεταξύ των χαρακτηριστικών (παγκόσμια κατάταξη – φήμη ιστοσελίδων, ευρετηριασμένες ιστοσελίδες – ημερήσιες προβολές). Τέλος, προτείνουν ορισμένες τεχνικές SEO που θα μπορούσαν να εφαρμοστούν για την προσέλκυση χρηστών στις ιστοσελίδες. Προτείνεται για υψηλή ορατότητα στα αποτελέσματα ο προσεκτικός σχεδιασμός των ιστοσελίδων (website design) και ειδικά το σχετικό περιεχόμενο με τις λέξεις – κλειδιά των αναζητήσεων, οι “ξεκάθαρες” και σημασιολογικές διευθύνσεις (urls), οι ευρετηριασμένες (indexed) και ανιχνευμένες σελίδες (crawled pages), όπως και η χρήση «χαρτών δόμησης» των ιστοσελίδων (sitemaps). Επίσης, σημειώνεται η σημαντικότητα των ποιοτικών εξερχόμενων συνδέσμων (external links) στην κατάταξη των αποτελεσμάτων και η χρήση των μέσων κοινωνικής δικτύωσης (social media optimization).

Σε μία άλλη δημοσίευση ο Scott (2015) ασχολείται με τις τεχνικές βελτιστοποίησης των αποτελεσμάτων στις μηχανές αναζήτησης για ιστοσελίδες βιβλιοθηκών. Συγκεκριμένα, αναλύει τις πιο σωστές και αποδεκτές μεθόδους σύμφωνα με τις πολιτικές των δημοφιλών μηχανών αναζήτησης (“white hat search engine optimization”). Ο κατάλληλος αυτός τρόπος βελτιστοποίησης αναφέρεται στην δημοσίευση χρήσιμου περιεχομένου για τους χρήστες σε συνδυασμό με τη σωστή δόμηση του, προκειμένου να γίνει κατανοητή σημασιολογικά και να τηρεί τους όρους και προδιαγραφές των μηχανών αναζήτησης. Η χρήση «δομημένων δεδομένων» (structured data) στις ιστοσελίδες βιβλιοθηκών μπορεί να διευκολύνει την πρόσβαση στο περιεχόμενο τους με μεγαλύτερη ακρίβεια. Οι σελίδες πρέπει πρώτα να ευρετηριαστούν αποτελεσματικά από τις μηχανές αναζήτησης (Google, Bing, Yahoo, Yandex) με τη χρήση ειδικών προγραμμάτων που ενσωματώνονται στο κώδικα (robots.txt) και των «χαρτών δόμησης», των λεγόμενων “sitemaps”.

Η εκμετάλλευση, επίσης, των μέσων κοινωνικής δικτύωσης με την εισαγωγή μεταδεδομένων για διάφορες σχετικές πληροφορίες αποτελεί μία αποτελεσματική πρακτική που δεν αξιοποιούν οι περισσότερες βιβλιοθήκες και οργανισμοί πληροφόρησης. Η εκτεταμένη χρήση τους ειδικά από τις μικρότερες ηλικιακές ομάδες δίνει την ευκαιρία για περαιτέρω προβολή των υπηρεσιών.

Με τη συνεργασία των μεγαλύτερων μηχανών αναζήτησης παγκοσμίως δημιουργήθηκε το schema.org. Πρόκειται ουσιαστικά για επιπλέον μεταδεδομένα στο κώδικα που δίνουν νόημα στο περιεχόμενο των ιστοσελίδων. Μ’ αυτό τον τρόπο γίνεται κατανοητό από τις μηχανές αναζήτησης, ακολουθώντας τη τάση του σημασιολογικού ιστού (semantic web). Η χρήση των μεταδεδομένων γίνεται για το όνομα, τις ώρες εργασίας, τη τοποθεσία, τα τηλέφωνα, τις υπηρεσίες, τις φωτογραφίες, τις διάφορες εκδηλώσεις που πραγματοποιούνται κ.ά. Το schema.org δεν βελτιστοποιεί τις ιστοσελίδες μόνο στα αποτελέσματα αναζήτησης των μηχανών, αλλά θέτει τα θεμέλια για την χρήση τους σε διαδικτυακές εφαρμογές και εφαρμογές κινητών.

Η επόμενη εργασία από τους Cahill & Chalut (2009) πραγματεύεται τη πρακτική της βελτιστοποίησης των ιστοσελίδων στη μηχανή αναζήτησης της Google. Πιο συγκεκριμένα, αναφέρεται ο αρχικός σκοπός της βελτιστοποίησης στις μηχανές αναζήτησης, ο οποίος είναι η διαφήμιση προϊόντων ή υπηρεσιών μέσω του διαδικτύου (Internet marketing). Με

την εμφάνιση της Google AdWords⁶, οι διαφημίσεις άρχισαν να βασίζονται στις λέξεις – κλειδιά των αναζητήσεων των χρηστών. Φάνηκε λοιπόν πιο αποτελεσματική η προβολή στις σχετικές αναζητήσεις και άρχισε να εμφανίζεται το φαινόμενο του SEO. Στη συνέχεια, αναφέρεται ο αλγόριθμος της Google «PageRank», ο οποίος καθορίζει την κατάταξη των ιστοσελίδων στα αποτελέσματα με βάση τον αριθμό και την ποιότητα των συνδέσμων.

Οι τεχνικές βελτιστοποίησης στις μηχανές αναζήτησης χωρίζονται στις “white hat” και στις “black hat”. Η κύρια διαφορά μεταξύ τους είναι ο απώτερος σκοπός, όπου στην πρώτη περίπτωση έχει να κάνει με την στοχευμένη προώθηση ενός προϊόντος, ενώ στη δεύτερη αποκλειστικά με τον αριθμό των επισκέψεων. Και οι δύο τεχνικές βασίζονται στο περιεχόμενο και στους συνδέσμους των ιστοσελίδων. Οι “black hat” μέθοδοι περιλαμβάνουν τη υπερβολική χρήση λέξεων – κλειδίων (keyword stuffing), και συνδέσμων (“spamdexing”), την αντιγραφή περιεχομένου και λέξεων – κλειδίων (site scraping), αλλά και την εισαγωγή παραπλανητικών σχολίων σε ιστοσελίδες (comment spam). Αντίθετα, οι “white hat” μέθοδοι εμπεριέχουν τη ορθή χρήση των λέξεων – κλειδίων στο πεδίο του τίτλου του κώδικα (<title>) που να έχουν κάποια άμεση ή έμμεση σχέση με τις αναζητήσεις των χρηστών, όπως και την εισαγωγή ποιοτικών συνδέσμων. Τέλος, οι συγγραφείς είναι επιφυλακτικοί για τον βαθμό επίδρασης των τεχνικών βελτιστοποίησης στις σελίδες των αποτελεσμάτων και τη συνάφεια με τις πληροφοριακές ανάγκες των χρηστών. Η Google έχει καθιερωθεί ως η καλύτερη μηχανή αναζήτησης παγκοσμίως, γεγονός που έχει καταστήσει τις επιδιώξεις των χρηστών υψηλές τόσο για τη σχετικότητα, όσο και τη χρησιμότητα των αποτελεσμάτων. Συγκεκριμένα χαρακτηριστικά, όπως η διόρθωση ορθογραφικών λαθών και η επιστροφή σημασιολογικών αποτελεσμάτων βελτιώνουν την εμπειρία των χρηστών, αλλά τους καθιστούν λιγότερο πρόθυμους να ψάξουν πέρα από τη πρώτη σελίδα. Ο ρόλος των βιβλιοθηκών είναι να εκπαιδεύσουν τους χρήστες των μηχανών αναζήτησης να έχουν κριτική στάση στη σειρά και τη ποιότητα των αποτελεσμάτων. Οι ολοένα αυξανόμενες απαιτήσεις για τον εντοπισμό και την ανάκτηση της σωστής πληροφορίας είναι απόρροια της χρήσης μίας και μόνο μηχανής αναζήτησης. Επίσης, στη σύγχρονη κοινωνία της πληροφορίας οι οργανισμοί πληροφόρησης από θεματοφύλακες έχουν γίνει εκπαιδευτές μέσω της παροχής σεμιναρίων, έγκυρων πηγών με εγγραφή, κατηγοριοποιημένων πηγών και βάσεων δεδομένων με πλήρη κείμενα. Η εφαρμογή των τεχνικών βελτιστοποίησης πρέπει να λαμβάνεται υπόψη κατά την αναζήτηση από τους χρήστες εξετάζοντας με κριτική ματιά τη συνάφεια των αποτελεσμάτων.

Το αντικείμενο έρευνας του άρθρου των Lee, Jang κ.α. (2016) είναι η εφαρμογή τεχνικών βελτιστοποίησης (SEO) στις μηχανές αναζήτησης για την ιστοσελίδα της βιβλιοθήκης της εταιρείας LG. Η υψηλή ορατότητα (visibility) στις μηχανές αναζήτησης συνεπάγεται και υψηλή επισκεψιμότητα (visitation) στην ιστοσελίδα της βιβλιοθήκης. Οι οργανισμοί πληροφόρησης χρειάζονται να προβάλλουν στοχευμένα το περιεχόμενο και τις υπηρεσίες τους στον τελικό χρήστη και το SEO αποτελεί το κύριο μέσο για την επίτευξη αυτού του σκοπού. Το SEO (Search Engine Optimization) αναφέρεται γενικότερα στις μεθόδους και τις διαδικασίες με τις οποίες μπορούμε να αυξήσουμε την κινητικότητα σε μία ιστοσελίδα. Είναι γεγονός ότι οι ιστοσελίδες βιβλιοθηκών δεν αποτελούν τους κύριους πληροφοριακούς

⁶ Το Google AdWords είναι η διαδικτυακή υπηρεσία διαφήμισης της Google.

κόμβους, σε αντίθεση με τις μηχανές αναζήτησης. Η ευκολία στην αναζήτηση και η γρήγορη ανάκτηση των αποτελεσμάτων σε συνδυασμό με την υψηλή προσβασιμότητα των μηχανών αναζήτησης οδήγησαν στην καθιέρωση και επικράτηση τους. Οι βιβλιοθήκες ωστόσο μπορούν να επωφεληθούν από τη διάχυση του περιεχομένου τους σ' αυτές. Στη συγκεκριμένη εργασία η ανάλυση 195 σχετικών λέξεων – κλειδιών οδήγησε στην κατάταξη τους χαμηλά στα αποτελέσματα των μηχανών αναζήτησης, όπως και το ποσοστό των επισκεπτών (39%). Οι τεχνικές που επιλέχθηκαν για τη βελτιστοποίηση της ιστοσελίδας, βασιζόνταν σε τρεις βασικούς στόχους: 1) την καλύτερη προσβασιμότητα (accessibility) από τις μηχανές αναζήτησης, 2) την αυξανόμενη σχετικότητα μεταξύ του περιεχομένου και των λέξεων – κλειδιών, και 3) την αυξανόμενη αξιοπιστία της ιστοσελίδας. Για τον πρώτο στόχο οι τεχνικές ήταν οι εξής: η βελτιστοποίηση της δομής των διευθύνσεων (URL structure optimization), η χρήση εσωτερικών ανακατευθύνσεων (Internal redirects), η χρήση XML χαρτών πλοήγησης για τις μηχανές αναζήτησης (XML sitemaps), αρχείων robots.txt κ.ά. Για την σχετικότητα χρησιμοποιήθηκαν οι ετικέτες <meta> και <title> στον κώδικα για την ακριβέστερη περιγραφή των σελίδων, ενώ για την αξιοπιστία της ιστοσελίδας έγιναν ανακατευθύνσεις για πολλά domains (Redirect for multiple domains), χρησιμοποιήθηκαν «Canonical Tags»⁷ για τον ορισμό του πρώτου προτιμώμενου URL και αφαιρέθηκαν οι μη λειτουργικοί ή «νεκροί» σύνδεσμοι (Dead links). Τα αποτελέσματα της εφαρμογής των τεχνικών βελτιστοποίησης οδήγησαν σε σημαντική αύξηση του ποσοστού των επισκεπτών από τις μηχανές αναζήτησης (61%), στην καλύτερη αξιολόγηση των λέξεων – κλειδιών και τη σημαντική αύξηση των ευρετηριασμένων σελίδων (indexed pages) από της μηχανές αναζήτησης από 61.300 σε 221.000.

Ένα από τα πιο κρίσιμα ερωτήματα που προκύπτουν από τη βιβλιογραφία είναι αν αξίζει να εφαρμόζουν οι οργανισμοί πληροφόρησης μεθόδους βελτιστοποίησης για τη προσέλκυση χρηστών. Σε άρθρο τους οι Rushton & Funke (2011) ασχολούνται με το αντικείμενο του SEO και τη στάση που πρέπει να κρατήσουν οι επιστήμονες της πληροφόρησης. Οι μηχανές αναζήτησης έχουν κυριαρχήσει έναντι των βιβλιοθηκών και ο μόνος τρόπος για να συμβαδίσουν είναι μέσω της απευθείας κίνησης (direct traffic) απ' αυτές. Ο στόχος αυτός μπορεί να επιτευχθεί με την εφαρμογή των κατάλληλων μεθόδων, οι οποίοι είναι οι εξής: η χρήση λέξεων – κλειδιών στις ετικέτες τίτλου και τις επικεφαλίδες (H1,H2...), η χρήση λειτουργικών και ποιοτικών συνδέσμων (hyperlinks) και χαρτών πλοήγησης (sitemaps) για τις μηχανές αναζήτησης, η περιγραφή των σελίδων με μεταδεδομένα σχετικά με το περιεχόμενο. Επιπλέον, καθοριστικό ρόλο έχει και η δημοτικότητα (popularity) στην εμφάνιση των αποτελεσμάτων. Εκτός από τις εισερχόμενες συνδέσεις (inbound links), το πλήθος των επισκεπτών και των επιστροφών, καθώς και ο μέσος όρος που μένουν στην ιστοσελίδα ή ο αριθμός των σελίδων ανά περιήγηση (pages per visit) έχουν σημαντική επίδραση στην κατάταξη της ιστοσελίδας. Στη συνέχεια, αναφέρεται ότι οι σελίδες πρέπει να είναι ευρετηριασμένες (indexed pages) από τους «crawlers» των μηχανών αναζήτησης και για να επιτευχθεί αυτό οφείλουν οι επιστήμονες

⁷ Τα Canonical Tags χρησιμοποιούνται για να υποδείξουν στις μηχανές αναζήτησης ότι μία διεύθυνση (URL) είναι η πρωτότυπη. Αυτό βοηθάει στην αποφυγή χρήσης διπλότυπων URL.

της πληροφόρησης να χρησιμοποιούν στοχευμένα τα αρχεία robots.txt (για την ένδειξη των ανοιχτών σελίδων), τα περιγραφικά μεταδεδομένα και το “alt attribute” για την περαιτέρω περιγραφή των εικόνων. Τονίζουν επίσης τη σημαντικότητα της εμφάνισης αποτελεσμάτων σύμφωνα με τις περιοχές στις οποίες βρίσκονται οι χρήστες κατά την αναζήτηση (localization), και της χρήσης των μέσων κοινωνικής δικτύωσης (social media) για την προβολή των υπηρεσιών και την προσέλκυση νέων χρηστών. Επιπλέον, γίνεται αναφορά στα εργαλεία “Web Analytics”, τα οποία μέσω των αναφορών μπορούν να δώσουν χρήσιμες πληροφορίες σχετικά με τις πηγές των επισκεπτών, τον μέσο όρο που μένουν στις σελίδες, τις λέξεις – κλειδιά που χρησιμοποιούν και τα δημογραφικά δεδομένα τους. Οι πληροφορίες αυτές μπορούν να οδηγήσουν στην αναγνώριση των δημοφιλών σελίδων και αυτών με χαμηλότερη απόδοση, προκειμένου να γίνουν οι απαραίτητες αλλαγές. Τέλος, καταλήγουν στο συμπέρασμα ότι η εφαρμογή των τεχνικών βελτιστοποίησης (SEO) είναι μία καλή και προσιτή πρακτική και οι επαγγελματίες της πληροφόρησης οφείλουν να ασχοληθούν ενεργά με την ενίσχυση της διαδικτυακής παρουσίας των βιβλιοθηκών.

2.6. Συμπεράσματα

Η παρούσα πτυχιακή εργασία εστιάζει στην εξέταση ορισμένων μεθόδων για τη βελτιστοποίηση ιστοσελίδων οργανισμών πληροφόρησης. Η αύξηση της ανακαλυψιμότητας του περιεχομένου αποτελεί αντικείμενο συζήτησης και προβληματισμού στη διεθνή βιβλιογραφία. Όπως έχει παρατηρηθεί σε αρκετές έρευνες, ολοένα και περισσότερα άτομα βασίζονται στις μηχανές αναζήτησης ως τη κύρια πηγή πληροφόρησης, λειτουργία η οποία πραγματοποιούνταν παλαιότερα από τους εκάστοτε πληροφοριακούς οργανισμούς. Όπως παρουσιάστηκε στο κεφάλαιο της βιβλιογραφικής ανασκόπησης, δεν υπάρχει κάποιος προκαθορισμένος τρόπος με τον οποίο πρέπει να ενεργήσουν οι οργανισμοί αυτοί. Βέβαια, είναι σχεδόν υποχρεωτικό γι’ αυτούς να διασφαλίσουν τη πρόσβαση στο περιεχόμενο τους μέσω των ποικίλων μηχανών αναζήτησης, προσφέροντας ποιοτική εμπειρία περιήγησης για τη μέγιστη βελτιστοποίηση στη σειρά των αποτελεσμάτων. Ο κύριος ρόλος τους, εκτός του τεχνικού μέρους, είναι η εκπαίδευση των χρηστών σε θέματα αναζήτησης και ανάκτησης πληροφοριών, καθώς και στο φαινόμενο SEO και την επίδραση του. Σίγουρα δεν έχουν να χάσουν απολύτως τίποτα από την εφαρμογή κάποιων μεθόδων και κανόνων για τη σωστότερη διάρθρωση των ιστοσελίδων τους.

Κεφάλαιο 3. Μεθοδολογία

Στη παρούσα πτυχιακή εργασία εξετάζεται το ζήτημα της βελτιστοποίησης στις μηχανές αναζήτησης (SEO) ως ένας μηχανισμός για την ευκολότερη εύρεση του περιεχομένου των ιστοσελίδων οργανισμών πληροφόρησης. Διερευνώνται τα στοιχεία – χαρακτηριστικά των ιστοσελίδων αυτής της κατηγορίας οργανισμών με απώτερο σκοπό την εύκολη ανάδειξη του περιεχομένου τους. Η ενίσχυση της δυνατότητας εύρεσης του περιεχομένου (content findability) και η εμφάνιση του στις σελίδες αποτελεσμάτων αποτελεί ένα σχετικά νέο ζήτημα για το σύνολο των οργανισμών αυτού του είδους. Λίγες είναι οι μελέτες, οι οποίες εξετάζουν τα προβλήματα που συναντούν οι επιστήμονες της πληροφόρησης.

Στη συγκεκριμένη εργασία εντοπίστηκαν και εξετάστηκαν τα κύρια χαρακτηριστικά που οδηγούν σε υψηλή κατάταξη, ταχύτητα και ασφάλεια, καθώς και τα επικρατέστερα συστήματα διαχείρισης περιεχομένου. Η εφαρμογή των εργαλείων για την καταμέτρηση των παραμέτρων έχει στόχο να εξετάσει τη βελτιστοποίηση του περιεχομένου για πιο εύκολη, ακριβή και γρήγορη ανάκτηση.

3.1. Συνοπτική περιγραφή της εφαρμογής της μεθοδολογίας

Για τη διεξαγωγή συμπερασμάτων προηγήθηκαν ορισμένα βήματα για τη συλλογή των στοιχείων και την επιλογή των κατάλληλων ιστοσελίδων και εργαλείων. Στο πρώτο βήμα προσδιορίστηκαν τα δεδομένα που θα χρησιμοποιηθούν για την ανάλυση και την εξαγωγή συμπερασμάτων. Οι μετρικές που επιλέχθηκαν είναι οι εξής: 1) οι τίτλοι των σελίδων, 2) η ποιότητα του περιεχομένου των σελίδων, 3) η χρήση παραμέτρων στις διευθύνσεις (URL), 4) η χρήση των αρχείων robots.txt, 5) η εισαγωγή χαρτών πλοήγησης (sitemaps), 6) οι ευρετηριασμένες διευθύνσεις 7) η χρήση του HTTPS, 8) ο αριθμός σελίδων περιήγησης ανά επίσκεψη, 9) το συνολικό ποσοστό SEO, 10) το συνολικό ποσοστό ταχύτητας, 11) το συνολικό ποσοστό ασφάλειας. Η επιλογή των συγκεκριμένων στοιχείων έγινε με κύριο γνώμονα τη συμπλήρωση της ήδη υπάρχουσας γνώσης στη βιβλιογραφία.

Σε δεύτερο βήμα πραγματοποιήθηκε η επιλογή των κατάλληλων εργαλείων για την άντληση δεδομένων σχετικά με την απόδοση των τεχνικών χαρακτηριστικών SEO και των δεδομένων συμπεριφοράς (Behavioral Analytics). Χρησιμοποιήθηκαν εργαλεία για την αξιολόγηση των ιστοσελίδων και αποτιμήθηκε η κατάσταση τους σε σχέση με τις απαιτήσεις SEO και user experience. Τα συγκεκριμένα στοιχεία συλλέχθηκαν με τα διαδικτυακά εργαλεία Checkbot και SimilarWeb. Τα εργαλεία αυτά επιλέχθηκαν με απώτερο σκοπό την απόκτηση μιας ολοκληρωμένης εικόνας για τη δομή των ιστοτόπων και των δεδομένων συμπεριφοράς των τελικών χρηστών. Για την ανάλυση των δεδομένων αξιοποιήθηκαν το JASP και το Excel. Τέλος, το Whatcms.org χρησιμοποιήθηκε για αναγνώριση του συστήματος διαχείρισης περιεχομένου που χρησιμοποιείται σε κάθε ιστοσελίδα.

Στο τρίτο βήμα πραγματοποιήθηκε η συγκέντρωση των στοιχείων των οργανισμών πληροφόρησης με τη βοήθεια του Checkbot και του SimilarWeb. Η συλλογή 143 διευθύνσεων ελληνικών και ξένων ιστοσελίδων οδήγησε στην ανάλυση των επιμέρους στοιχείων στο ελεύθερο λογισμικό ανοιχτού κώδικα για στατιστική ανάλυση JASP. Αρχικά, έγινε η ανάκτηση των στοιχείων από το Checkbot σε αρχεία μορφής .pdf. Τα στοιχεία αυτά περάστηκαν χειροκίνητα σε ένα αρχείο excel για τη καλύτερη διαχείριση τους. Έπειτα, έγινε

μετατροπή σε αρχείο .jasr διαδικτυακά και αποθηκεύτηκαν τοπικά. Πραγματοποιήθηκε και ο εννοιολογικός διαχωρισμός τους σε στοιχεία απόδοσης των μεταδεδομένων (Metadata Performance), της απόδοσης των στοιχείων «αρχιτεκτονικής» των ιστοσελίδων (Architecture Structure Performance), της απόδοσης της ασφάλειας (Security Performance) και της συνολικής απόδοσης & του μεγέθους (Overall Performance and Size) και των δεδομένων συμπεριφοράς (Behavioral Analytics) στο περιβάλλον του JASP. Η ίδια διαδικασία ακολουθήθηκε και για τη συλλογή από το εργαλείο SimilarWeb. Η σημασιολογική απόδοση των στοιχείων βοήθησε στη περαιτέρω κατηγοριοποίηση και ανάλυση τους. Η επιλογή των στατιστικών στοιχείων και ο διαχωρισμός τους έγινε χειροκίνητα σε συνδυασμό με τη χρήση φίλτρων ανάλογα με τη προέλευση των οργανισμών πληροφόρησης στο περιβάλλον του JASP. Τέλος, δημιουργήθηκαν οι πίνακες σε αρχείο excel με βάση τις παραπάνω έννοιες και διαχωρίστηκαν σε ξένους και ελληνικούς οργανισμούς για την εξαγωγή αποτελεσμάτων συγκριτικά με τα ποσοστά τους. Τα υπόλοιπα στοιχεία που δεν αξιοποιήθηκαν στην στατιστική ανάλυση παρατίθενται για τη περαιτέρω διεύρυνση των δεδομένων.

Η επεξεργασία των πρωτογενών δεδομένων και η εφαρμογή μεθόδων στατιστικής ανάλυσης έγινε με κύριο γνώμονα την απεικόνιση της κατάστασης των ιστοσελίδων και την εξαγωγή αντικειμενικών συμπερασμάτων. Το κάθε στοιχείο αναλύθηκε με βάση την προέλευση των οργανισμών σε ελληνικές και ξένες ιστοσελίδες. Η εικόνα των επιμέρους αυτών τεχνικών στοιχείων ξεχωριστά συγκρίθηκε με τα συνολικά ποσοστά (SEO, ταχύτητα και ασφάλεια) και οδήγησε στα γενικά συμπεράσματα ανάλογα πάλι τη προέλευση.

Η ίδια ακριβώς διαδικασία πραγματοποιήθηκε και με βάση τα χρησιμοποιούμενα συστήματα διαχείρισης περιεχομένου (CMS) και σε σχέση με κριτήρια όπως το συνολικό σκορ SEO (SEO), την ταχύτητα (Speed) και την ασφάλεια (Security). Η αναγνώριση τους έγινε με τη βοήθεια του εργαλείου Whatcms.org. Η διαδικασία αναγνώρισης ήταν απλή αφού με την εισαγωγή της κάθε διεύθυνσης και το πάτημα ενός κουμπιού εμφανιζόταν το σύστημα διαχείρισης περιεχομένου που χρησιμοποιεί κάθε ιστοσελίδα. Ειδικότερα, σε ότι αφορά τη δημιουργία πινάκων, πραγματοποιήθηκε σε αρχείο excel και έγινε με το διαχωρισμό σε Wordpress, Drupal και Custom. Σχετικά με τα συστήματα διαχείρισης περιεχομένου έγινε σύγκριση μεταξύ των δύο επικρατέστερων Wordpress και Drupal και των ιστοτόπων που δημιουργήθηκαν με ειδικό κώδικα (source code), χωρίς τη χρήση καποιου συστήματος CMS (custom). Η σύγκριση των αποτελεσμάτων των κριτηρίων οδήγησε στο συμπέρασμα ότι το κάθε σύστημα υπερτερεί έναντι των υπολοίπων σε ορισμένα χαρακτηριστικά και ότι η επιλογή του κάθε οργανισμού μπορεί να προσαρμοστεί με βάση τις ανάγκες του.

Στο τέταρτο βήμα πραγματοποιήθηκε η αξιολόγηση των στοιχείων με βάση τα εργαλεία που προαναφέρθηκαν. Τα δεδομένα και οι λεπτομέρειες σε ότι αφορά τα στοιχεία που συγκεντρώθηκαν ανά εργαλείο παρουσιάζονται στις παραγράφους που ακολουθούν. Στο ίδιο βήμα με τη χρήση του Jasp και του Excel πραγματοποιήθηκε η επεξεργασία των αποτελεσμάτων. Έγινε ανατροφοδότηση του τρίτου βήματος για τη καλύτερη διαχείριση των δεδομένων. Τέλος, έγινε η ανάλυση των αποτελεσμάτων με τη δημιουργία των πινάκων και τη βοήθεια του αρχείου excel. Η εξαγωγή συμπερασμάτων προέκυψε από τη

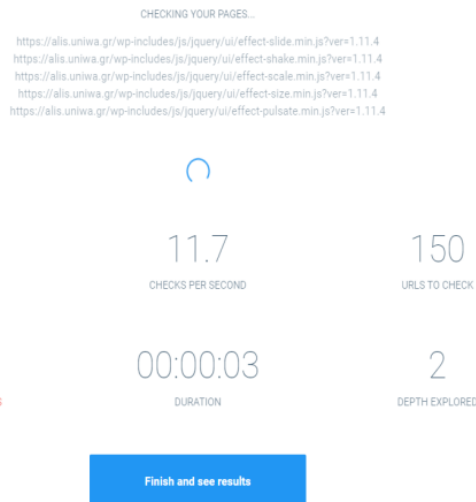
σύγκριση των μετρικών και αξιολογήθηκε η επιτυχία της ερευνητικής διαδικασίας. Είναι σημαντικό να σημειωθεί ότι έγινε μετάβαση από το ένα βήμα στο άλλο πολλές φορές για τη διόρθωση λαθών και την αξιολόγηση της πορείας της έρευνας.

3.2. Εργαλεία SEO

Το Checkbot είναι ένα πρόγραμμα, το οποίο παρέχεται δωρεάν για τον έλεγχο ενός μικρού αριθμού διευθύνσεων. Παρέχεται επίσης και με συνδρομή για 10000 σελίδες ανά διεύθυνση, αποκτώντας κανείς πιο ολοκληρωμένη εικόνα για τις ιστοσελίδες. Η χρήση του εργαλείου μπορεί να βοηθήσει κάποιον να εντοπίσει τις αδυναμίες των ιστοσελίδων που διαχειρίζεται και να δώσει έμφαση στα δυνατά τους σημεία. Κάνοντας κάποιες μικρές αλλαγές στα τεχνικά χαρακτηριστικά των ιστοσελίδων μπορούν οι σελίδες να καταταγούν σε υψηλότερες θέσεις κάνοντας το περιεχόμενο πιο εύκολα προσβάσιμο. Το εργαλείο βοηθά στην εσωτερική βελτιστοποίηση μιας ιστοσελίδας (On-page SEO) περιλαμβάνοντας τον σχεδιασμό της, τα μεταδεδομένα, τις λέξεις-κλειδιά, τα ονόματα των σελίδων, τις εικόνες, τους συνδέσμους, το περιεχόμενο της κάθε σελίδας, τον πίνακα πλοήγησης (sitemap) κ.ά. Με τη διόρθωση ορισμένων μετρικών οι περισσότεροι χρήστες θα είναι πιο πιθανόν να έχουν πρόσβαση στους ιστοτόπους για να καλύψουν τη πληροφοριακή τους ανάγκη. Η επιλογή του συγκεκριμένου εργαλείου έγινε λόγω της πληθώρας των μετρικών που παρέχει, καθώς και του εννοιολογικού τους διαχωρισμού σε μετρικές SEO, ταχύτητας και ασφάλειας. Η ακρίβεια στην εμφάνιση των αποτελεσμάτων είναι ένα από τα προτερήματα του έναντι άλλων προγραμμάτων. Η δυνατότητα επίσης εξαγωγής των αποτελεσμάτων σε μορφή pdf ή csv προσθέτει στη διαλειτουργικότητα των δεδομένων.

3.2.1. Checkbot

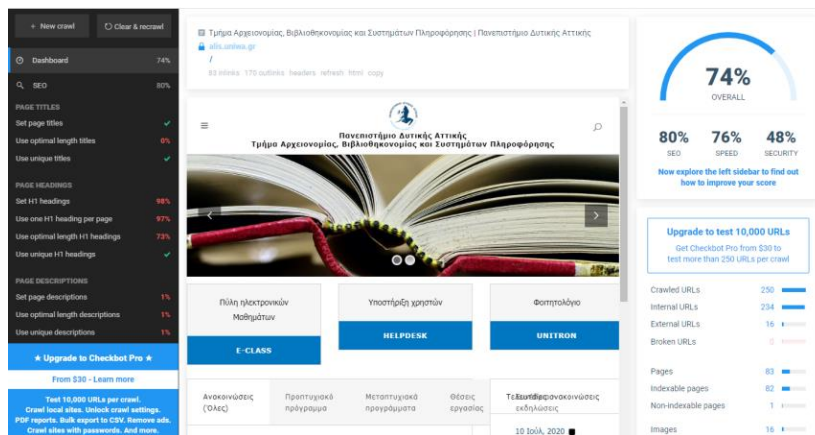
Το Checkbot (*Checkbot*) είναι ένα διαδικτυακό εργαλείο, το οποίο ελέγχει τις σελίδες διευθύνσεων (domains) για τη βελτιστοποίηση τους στις μηχανές αναζήτησης (SEO), για προβλήματα ταχύτητας και ασφάλειας. Το πρόγραμμα με την έναρξη του αρχίζει και ψάχνει (crawling) τις σελίδες και καταγράφει τις αντίστοιχες μετρικές που εντοπίζει. Τα αποτελέσματα εμφανίζονται σε ποσοστό ή με βάση τη ποσότητα τους, ενώ δίνεται και μία σχετική επεξήγηση των μετρικών. Στις παραγράφους που ακολουθούν παρουσιάζονται οι μετρικές που αφορούν χαρακτηριστικά SEO, συμπεριλαμβανομένων και των συνολικών σκορ που επιτυγχάνει ο ιστότοπος υπό ανάλυση καθώς και μετρικές σχετικά με την ασφάλεια και την ταχύτητα.



Εικόνα 1. Ενδεικτική οθόνη λειτουργίας του Checkbot κατά την διάρκεια ελέγχου ιστοσελίδας

3.2.1.1. Μετρικές SEO

Η πρώτη κατηγορία σχετίζεται με το SEO και περιλαμβάνει τις εξής ενότητες: 1) τον τίτλο της σελίδας (Page Titles), 2) τις επικεφαλίδες της σελίδας (Page Headings), 3) τη περιγραφή της σελίδας (Page Descriptions), 4) το διπλότυπο περιεχόμενο (Duplicate Content), 5) το περιεχόμενο της σελίδας (Page Content), 6) τα ονόματα των διευθύνσεων (Url Names), 7) την εγκυρότητα του κώδικα (Code Validation), 8) τους συνδέσμους (Links), 9) τα αρχεία robots.txt (robots.txt), 10) τις ανακατευθύνσεις (Redirects).



Εικόνα 2. Ενδεικτική οθόνη εμφάνισης των αποτελεσμάτων - Checkbot

Η κάθε μία από τις παραπάνω ενότητες περιλαμβάνει επιμέρους στοιχεία – μετρικές. Στον τίτλο σελίδας λοιπόν εμπεριέχεται η προσθήκη του τίτλου (Set page titles). Σε κάθε σελίδα πρέπει να δίνεται ένας τίτλος που να περιγράφει περιεκτικά το περιεχόμενο της. Η ακρίβεια στην απόδοση του τίτλου μπορεί να βελτιώσει σημαντικά την κατάταξη της στα αποτελέσματα των μηχανών αναζήτησης λόγω της σχετικότητας του με τις ανάγκες των χρηστών (queries). Επίσης, σημαντική μετρική είναι η χρήση του σωστού μήκους του τίτλου (Use optimal length titles), ο οποίος θα πρέπει να είναι μεταξύ 10 – 60 χαρακτήρων. Οι πληροφορίες που δίνονται σ’ αυτόν πρέπει να είναι περιεκτικές και να παρέχουν την κατάλληλη πληροφορία. Η χρήση μοναδικού τίτλου (Use unique titles) είναι απαραίτητη για τη βελτιστοποίηση. Η εμφάνιση του ίδιου τίτλου (duplicate title) σε δύο διαφορετικές

σελίδες δυσκολεύει τόσο τον χρήστη στην επιλογή, όσο και τις μηχανές αναζήτησης στην εμφάνιση. Η προσθήκη τίτλου γίνεται με τη χρήση HTML κώδικα προσθέτοντας το στοιχείο <title>...</title> μέσα στο <head>.

Όσον αφορά τις επικεφαλίδες της σελίδας, θα πρέπει να περιέχεται σε κάθε σελίδα μία που να περιγράφει το περιεχόμενο της (Set H1 headings). Η προσθήκη της θα βοηθήσει τους χρήστες να κατανοήσουν το περιεχόμενο. Οι λέξεις κλειδιά που περιλαμβάνονται αποτελούν ένα σημαντικό δείκτη κατάταξης. Η προσθήκη τους γίνεται με το στοιχείο <h1>...</h1>. Η χρήση μίας επικεφαλίδας H1 (Use one H1 heading per page) ενδείκνυται για τη καθιέρωση ενός συγκεκριμένου θέματος. Προτείνεται η χρήση των επικεφαλίδων ιεραρχικά στη δομή του περιεχομένου, επιλέγοντας μία ως κύρια H1 και τις υπόλοιπες στη συνέχεια (<h2>, <h3>, <h4>). Το μήκος των επικεφαλίδων επιπλέον δεν θα πρέπει να είναι μεγάλο και να ξεπερνάει τους 70 χαρακτήρες (Use optimal length H1 headings). Όπως και στους τίτλους, η ακρίβεια στην απόδοση τους είναι απαραίτητη για τη βελτιστοποίηση της σελίδας και επομένως η χρήση μοναδικών επικεφαλίδων επιβάλλεται (Use unique H1 headings).

Σχετικά με τη περιγραφή της σελίδας, πρέπει να υπάρχει μία για να συνοψίζει το περιεχόμενο (Set page descriptions). Αν και η περιγραφή εμφανίζεται στα αποτελέσματα, δεν αποτελεί άμεσα παράγοντα κατάταξης. Μπορεί όμως να βοηθήσει σημαντικά το ποσοστό των εισόδων στην σελίδα (click-through rates) και να ενισχύσει τη θέση έμμεσα. Για την εισαγωγή περιγραφής χρησιμοποιείται το στοιχείο <meta name="description" content="..."> μέσα στο <head>. Η έκταση της περιγραφής πρέπει να κυμαίνεται μεταξύ 100-320 χαρακτήρων (Use optimal length descriptions). Ο τίτλος της σελίδας, η επικεφαλίδα και η περιγραφή οφείλουν να έχουν μέτρια έκταση, να είναι μοναδικές και να περιγράφουν με το βέλτιστο τρόπο τη σελίδα. Επιπρόσθετα, κάθε σελίδα πρέπει να έχει μία μοναδική περιγραφή (Use unique descriptions) για να μη προκαλεί σύγχυση στο χρήστη και τις μηχανές αναζήτησης.

Το διπλότυπο περιεχόμενο μπορεί να αποφευχθεί με τη καθιέρωση κανονικών διευθύνσεων (Set canonical URLs). Σε κάθε σελίδα ξεχωριστά θα πρέπει να έχει οριστεί ένα URL προκειμένου να περιοριστούν οι πιθανότητες χρήσης του ίδιου για παραπάνω από μία σελίδα. Με αυτόν τον τρόπο οι σελίδες θα να είναι πιο εύκολο να ευρετηριαστούν και να καταταγούν στη καλύτερη δυνατή θέση. Η προσθήκη κανονικών URLs γίνεται με το στοιχείο <link rel="canonical" href="..."> μέσα στο <head>. Θα πρέπει επίσης να αποφεύγεται η αντιγραφή περιεχομένου και η επανάληψη του. Κάθε σελίδα πρέπει να έχει το δικό της μοναδικό περιεχόμενο και να μην εμφανίζεται πουθενά αλλού στην ιστοσελίδα (Avoid duplicate page content). Οι μηχανές αναζήτησης δεν εμφανίζουν διπλότυπο περιεχόμενο, καθώς αυτό δεν εξυπηρετεί τους χρήστες.

Το περιεχόμενο της ιστοσελίδας μπορεί να βελτιστοποιηθεί με την παροχή χρήσιμης πληροφορίας σχετική με την αναζήτηση των χρηστών. Τα αποτελέσματα υψηλής ποιότητας κατατάσσονται στις πρώτες θέσεις βοηθώντας περισσότερο τους χρήστες. Οπότε, μία καλή πρακτική είναι η αποφυγή σελίδων με "φτωχό" περιεχόμενο (Avoid thin content pages). Η ακριβής περιγραφή των εικόνων βοηθάει επίσης τις μηχανές αναζήτησης στην καλύτερη κατανόηση του περιεχομένου της σελίδας ελέγχοντας τις σχετικές λέξεις –

κλειδιά (Set image ALT text). Η προσθήκη του στοιχείου `` στον κώδικα είναι σημαντική και για τις βοηθητικές τεχνολογίες, γεγονός που ενισχύει την κατάταξη της σελίδας. Η παρουσία στοιχείων που κάνουν πιο προσίτες τις σελίδες στις κινητές συσκευές, ενισχύει τη θέση και τη προσβασιμότητα (Set mobile scaling). Το στοιχείο `viewport` για παράδειγμα μπορεί να προσαρμόσει το μέγεθος της σελίδας στη κάθε κινητή συσκευή, προκειμένου το περιεχόμενο να είναι ευανάγνωστο και φιλικό. Αυτό επιτυγχάνεται με την εισαγωγή του στοιχείου `<meta name="viewport" content="width=device-width, initial-scale=1">`. Η χρήση ειδικών plugins στους περιηγητές (browsers) για την αναπαραγωγή περιεχομένου πρέπει να μην χρησιμοποιείται, καθώς πολλοί χρήστες μπορεί να μην διαθέτουν το κατάλληλο (Avoid plugins). Αντίθετα, προτείνεται η χρήση των στοιχείων `<object>`, `<embed>`, `<applet>` στον κώδικα για δυναμικό περιεχόμενο.

Έχουν δοθεί κάποιες οδηγίες από τις μεγαλύτερες μηχανές αναζήτησης για τα ονόματα διευθύνσεων (URLs), ώστε να είναι πιο ακριβή και σχετικά με τις αναζητήσεις. Αρχικά, δεν πρέπει να είναι εκτενή και να ξεπερνούν τους 100 χαρακτήρες, αλλά να είναι περιεκτικά και χρηστικά (Use short URLs). Η χρήση επίσης επεκτάσεων (.htm .html, .php κ.ά.) δεν προτείνεται γιατί απλώς δίνει πληροφορίες για τη τεχνολογία που χρησιμοποιήθηκε για τη δημιουργία της σελίδας, κάνοντας το url πιο εκτενές (Avoid URL extensions). Καλή πρακτική θεωρείται η χρήση των καθέτων για τη κατηγοριοποίηση του περιεχομένου (example.com/articles/news) αντί για τη χρήση των παραμέτρων (example.com/articles?topic=news) (Avoid URL parameters) . Οι παράμετροι κάνουν το όνομα να φαίνεται λιγότερο φιλικό και πιο περίπλοκο για τις μηχανές. Επιπλέον, δεν προτείνεται η χρήση συμβόλων (Avoid symbols in URLs), αλλά αναγνώσιμων εναλλακτικών, όπως οι παύλες ή οι κάθετοι ανάμεσα στις λέξεις. Ενδείκνυται η εισαγωγή πεζών γραμμάτων (Use lowercase URLs) στα ονόματα διευθύνσεων για την αποφυγή διπλότυπων σελίδων και η χρήση παυλών ανάμεσα στις λέξεις αντί για κάτω παύλες (Avoid underscores in URLs). Τέλος, προτείνονται οι πιο απλές διευθύνσεις χωρίς πολλούς υποφακέλους (Avoid deeply nested URLs) για να είναι μικρότερες σε έκταση, λιγότερο περίπλοκες και ευανάγνωστες.

Σχετικά με την εγκυρότητα του κώδικα, προτείνεται η δημιουργία σωστού κώδικα HTML, CSS και JavaScript (Use valid HTML, CSS, JavaScript). Τυχόν προβλήματα μπορούν να παρουσιάσουν με διαφορετικό τρόπο το περιεχόμενο απ' ότι προοριζόταν και να καταταγούν σε χαμηλότερη θέση στις μηχανές αναζήτησης μειώνοντας τη ποιότητα της εμπειρίας του χρήστη (user experience).

Η παρουσία των συνδέσμων παίζει σημαντικό ρόλο στη βελτιστοποίηση μίας ιστοσελίδας. Οι μηχανές αναζήτησης προτείνουν τη χρήση του στοιχείου 404 στο κώδικα για συνδέσμους που δεν λειτουργούν (Use 404 code for broken URLs). Όταν μία διεύθυνση ζητείται, η οποία δεν υπάρχει τότε πρέπει να εμφανίζεται η θέση 404 HTTP προκειμένου τα bots να το γνωρίζουν. Οι σελίδες που περιέχονται στους συνδέσμους δεν περιλαμβάνονται στα αποτελέσματα βοηθώντας έμμεσα στη βελτιστοποίηση. Επιπλέον, προτείνεται να μην χρησιμοποιούνται μη λειτουργικοί σύνδεσμοι εσωτερικοί ή εξωτερικοί (Avoid broken internal / external links), οι οποίοι παρεμποδίζουν το έργο των μηχανών αναζήτησης στην ευρετηρίαση (indexing) και μειώνουν την ποιότητα της εμπειρίας των χρηστών. Μία σελίδα

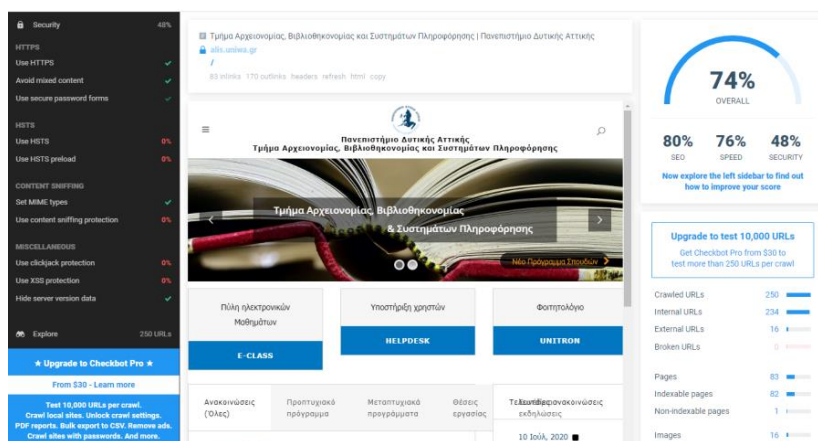
μπορεί να μη φορτώνει ή να παρουσιάζεται εσφαλμένα από τη χρήση μη λειτουργικών πηγών, όπως εικόνες, αρχεία CSS και JavaScript (Avoid broken page resources).

Τα αρχεία robots.txt προσθέτονται σε κάθε υποδιεύθυνση προκειμένου να δημιουργηθούν τοποθεσίες στον χάρτη πλοήγησης (sitemap) και να καθοριστούν οι σελίδες στις οποίες θα γίνεται ευρετηρίαση ή όχι (Use robots.txt files). Σ' αυτά τα αρχεία θα πρέπει να οριστούν οι συγκεκριμένες διευθύνσεις των σελίδων που θα διαβαστούν από τα bots των μηχανών αναζήτησης προσθέτοντας μεταδεδομένα περιγραφής για τη κάθε σελίδα και κάνοντας πιο εύκολο το έργο τους (Set sitemap locations).

Σχετικά με τις ανακατευθύνσεις προτείνεται η χρήση των server – side redirects έναντι των meta redirects (Avoid meta redirects). Αυτό συμβαίνει γιατί οι ανακατευθύνσεις στην πλευρά του χρήστη (user – side) μπορεί να εμφανίζονται στην οθόνη του προβάλλοντας μία σελίδα στην αρχή ακολουθώντας με άλλη μία, γεγονός που προκαλεί σύγχυση σ' αυτόν και τις μηχανές αναζήτησης. Κάποιες μηχανές αναζήτησης μπορεί επίσης να παραβλέψουν την ύπαρξη της. Συνιστάται και η χρήση μόνιμων ανακατευθύνσεων (status 301) έναντι των προσωρινών (status 302–307) (Avoid temporary redirects). Οι προσωρινές ανακατευθύνσεις υποδεικνύουν ότι μία σελίδα έχει μετακινηθεί προσωρινά και ότι οι μηχανές αναζήτησης δεν πρέπει να ενημερώσουν τους συνδέσμους, επηρεάζοντας αρνητικά την κατάταξη της σελίδας.

3.2.1.2. Μετρικές σχετικές με την ασφάλεια

Η δεύτερη κατηγορία σχετίζεται με την ασφάλεια (security) των ιστοσελίδων και περιλαμβάνει τις εξής κατηγορίες: 1) HTTPS, 2) HSTS, 3) Content sniffing και 4) Miscellaneous.



Εικόνα 3. Ενδεικτική οθόνη εμφάνισης των αποτελεσμάτων για τις μετρικές ασφαλείας

Η κάθε μία από τις παραπάνω ενότητες περιλαμβάνει επιμέρους στοιχεία – μετρικές. Στο HTTPS εμπεριέχεται η χρήση του (Use HTTPS) για την αποφυγή επιθέσεων και την προστασία της ιδιωτικότητας των χρηστών. Οι κωδικοί πρόσβασης και τα προσωπικά στοιχεία δεν είναι εκτεθειμένα σε τρίτους όταν προστίθεται το s στο http:// διότι υπάρχει κρυπτογράφηση των δεδομένων που ανταλλάσσονται μεταξύ των ιστοσελίδων και των περιηγητών. Η μη χρήση επίσης μεικτού περιεχομένου (Avoid mixed content), δηλαδή η στρατηγική αποφυγής των http:// πηγών σε όλα τα urls μίας ιστοσελίδας προσθέτει επιπλέον ασφάλεια. Οι ασφαλείς φόρμες κωδικών πρόσβασης (Use secure password forms)

με την εισαγωγή του POST αιτήματος μέσω `https://` ενισχύει περαιτέρω στην εμπιστευτικότητα του ιστοτόπου.

Το HSTS (HTTP Strict Transport Security) στέλνει αίτημα στους περιηγητές να χρησιμοποιούν πάντα HTTPS αντί για HTTP ιστοσελίδες (Use HSTS), ενώ το HSTS preload το χρησιμοποιεί προκαθορισμένα (Use HSTS preload).

Στο content sniffing η χρήση επικεφαλίδων «Content – Type» καθορίζει το τύπο της απάντησης που δίνεται από τους περιηγητές στις ιστοσελίδες. Ο ακριβής τύπος MIME βοηθά στην αποφυγή λαθών και τη δημιουργία ασφαλών ιστοσελίδων (Set MIME types). Η απενεργοποίηση της δυνατότητας εισχώρησης περιεχομένου είναι άλλη μία μέθοδος για τη προστασία ιστοσελίδων (Use content sniffing protection). Όταν οι περιηγητές αγνοήσουν τις επικεφαλίδες Content – Type, συνήθως υποθέτουν τον τύπο ανάλογα με την απάντηση. Η εκμετάλλευση αυτής της ανακρίβειας μπορεί να περιοριστεί με την εισαγωγή του `nosniff` στην επικεφαλίδα. Με αυτόν τον τρόπο οι περιηγητές (browsers) βασίζονται αποκλειστικά στο Content – Type.

Η προστασία από επιθέσεις τύπου clickjack γίνεται με τη χρήση της επικεφαλίδας X – Frame – Options, προκειμένου να περιοριστεί η δυνατότητα ένθεσης στις σελίδες (Use clickjack protection). Το clickjacking χρησιμοποιείται για την εκμετάλλευση των χρηστών μέσω συνδέσμων σε μέσα κοινωνικής δικτύωσης, σε διαφημίσεις και την κλοπή κωδικών πρόσβασης. Οι επιθέσεις XSS γίνονται με τη προσθήκη κώδικα σε μία σελίδα που στέλνεται στους χρήστες. Η χρήση του X-XSS-Protection: 1; mode=block στην επικεφαλίδα απάντησης κάθε σελίδας περιορίζει αυτό τον κίνδυνο (Use XSS protection), ωστόσο αυτή η μέθοδος πρέπει να είναι η τελευταία στη σειρά έναντι τέτοιου είδους επιθέσεων. Τέλος, προτείνεται η αποφυγή διαρροής πληροφοριών σε εισβολείς με την αφαίρεση των επικεφαλίδων απάντησης που αναφέρουν το λογισμικό που χρησιμοποιεί ο διακομιστής (Hide server version data).

3.2.1.3 Μετρικές σχετικές με την ταχύτητα

Η τρίτη κατηγορία σχετίζεται με τη ταχύτητα (speed) των ιστοσελίδων και περιλαμβάνει τις εξής κατηγορίες: 1) Page size, 2) Javascript, 3) CSS, 4) Redirects, 5) Caching. Η κάθε μία από τις παραπάνω ενότητες περιλαμβάνει επιμέρους στοιχεία. Στο Page size, δηλαδή το μέγεθος της σελίδας, περιέχονται 5 μετρικές. Η αποστολή δεδομένων στον εξυπηρετητή (server) σε συμπιεσμένη μορφή μπορεί να βοηθήσει σημαντικά στη μείωση του χρόνου μεταφοράς τους και να επιτευχθεί με αυτό τον τρόπο μεγαλύτερη ταχύτητα κατά τη φόρτωση τους (Use compression). Η συμπίεση δεδομένων δεν πρέπει να επαναλαμβάνεται σε ήδη συμπιεσμένα αρχεία, όπως αρχεία τύπου jpg, png, mp4 και pdf για την αποφυγή σπατάλης επιπλέον χρόνου (Avoid recompressing data). Η ελαχιστοποίηση επίσης των αρχείων CSS και Javascript βοηθούν στη μείωση του “βάρους” των ιστοσελίδων και του χρόνου φόρτωσης τους. Η μεταφορά ή η περικοπή ενός μέρους του περιεχομένου αυτών χωρίς την απώλεια της συνεισφοράς τους μπορεί να διατηρήσει τη συμπεριφορά, αλλά και να μειώσει το μέγεθος τους (Use minification). Η μη χρήση επιπλέον των ελαχιστοποιημένων αρχείων για τους “χάρτες πηγής” (source maps) είναι μία αποδεκτή μέθοδος για τη λειτουργικότητα αμφότερων. Οι χάρτες πηγής χρησιμοποιούνται για τον εντοπισμό σφαλμάτων στους συμπιεσμένους φακέλους. Τα εργαλεία ελαχιστοποίησης

αποθηκεύουν τους χάρτες πηγής είτε σε εξωτερικό φάκελο είτε στον ίδιο το αρχείο ελαχιστοποίησης, μέθοδος που πρέπει να αποφεύγεται (Avoid inline source maps).

Δεν προτείνεται η χρήση Javascript μέσα στο κώδικα διότι μειώνει τους χρόνους απόδοσης των σελίδων και της προσωρινής αποθήκευσης (caching). Είναι προτιμότερο να χρησιμοποιείται εξωτερικό αρχείο με την εισαγωγή του `<script src='...'>` στο κώδικα (Avoid excessive inline Javascript). Η χρήση Javascript στο κώδικα συνιστάται μόνο για αρχεία μικρού όγκου. Τα εξωτερικά αρχεία Javascript πρέπει επίσης να χρησιμοποιούνται με τέτοιο τρόπο ώστε να μη μειώνουν την απόδοση των ιστοσελίδων. Η χρήση της ετικέτας `<script>` μπορεί να μπλοκάρει τα αρχεία HTML, αλλά με την εισαγωγή της πριν το κλείσιμο της ετικέτας `</body>` αυτό μπορεί να αποφευχθεί (Avoid render-blocking Javascript).

Συνιστάται επίσης η χρήση εξωτερικών αρχείων CSS αντί μέσα στο κώδικα για τη βελτίωση της ταχύτητας (Avoid excessive inline CSS) και η αποφυγή του στοιχείου `@import`, το οποίο αποτρέπει την παράλληλη φόρτωση των αρχείων CSS. Η παράλληλη φόρτωση επηρεάζει σημαντικά τον χρόνο φόρτωσης των σελίδων.

Η χρήση των ανακατευθύνσεων για σελίδες δεν φαίνεται να είναι η καλύτερη δυνατή πρακτική, καθώς αυξάνει τον χρόνο φόρτωσης (Avoid resource redirects). Για παράδειγμα αν φορτώσει ένα αρχείο CSS χρησιμοποιώντας μία διεύθυνση της ιστοσελίδας και ανακατευθύνει τον χρήστη σε μία νέα θα υπάρχει αισθητή καθυστέρηση και μείωση της απόδοσης χρόνου. Επίσης, πρέπει να αποφεύγεται η χρήση αλυσίδας ανακατευθύνσεων για να μη προκαλείται σύγχυση στο χρήστη τη στιγμή της περιήγησης και να υπάρχει αργή ανταπόκριση (Avoid chain redirects).

Η ενεργοποίηση της δυνατότητας προσωρινής αποθήκευσης των πηγών των ιστοσελίδων από τους περιηγητές (browsers) μπορεί να βοηθήσει στη ταχύτητα φόρτωσης των επιμέρους σελίδων (Use caching). Η εισαγωγή της επικεφαλίδας Cache-Control στο κώδικα των ιστοσελίδων έχει άμεση επίδραση στη ταχύτητα. Το ίδιο συμβαίνει και με τη χρήση μεγάλης διάρκειας αποθήκευσης των πηγών (Use long caching times). Η διαμόρφωση των ιστοσελίδων με αυτόν τον τρόπο θα δώσει τη δυνατότητα στους περιηγητές να κρατήσουν περισσότερο τα δεδομένα τους και να είναι αποτελεσματικοί για μεγαλύτερο χρονικό διάστημα. Η εισαγωγή της ρύθμισης Expires θέτει τη διάρκεια της παραμονής των περιηγητών. Τέλος, προτείνεται η χρήση των ίδιων σελίδων για τη βελτιστοποίηση της απόδοσης αποθήκευσης των εκάστοτε διευθύνσεων (Avoid duplicate resources). Οι περιηγητές χρησιμοποιούν μόνο τις αποθηκευμένες σελίδες τις οποίες ζητάνε οι χρήστες από τις αντίστοιχες διευθύνσεις (URLs). Είναι σημαντικό να σημειωθεί ότι στην εργασία συμπεριλήφθηκε η συνολική ταχύτητα των ιστοσελίδων, η οποία όμως προκύπτει από τις παραπάνω μετρικές. Η ανάλυση τους έγινε για τη καλύτερη κατανόηση της μετρικής που χρησιμοποιείται.

3.2.1.4. Τελικά σκορ

Οι ποσοστιαίες μονάδες για το συνολικό SEO, τη συνολική ταχύτητα και τη συνολική ασφάλεια προκύπτουν αθροιστικά ως μέσοι όροι των παραπάνω μετρικών που αναλύθηκαν και ανήκουν στην αντίστοιχη κατηγορία. Το συνολικό SEO εμπεριέχει μετρικές σχετικά με τη δομή των ιστοσελίδων και τα τεχνικά χαρακτηριστικά τους. Η βελτιστοποίηση των ιστοσελίδων μπορεί να επιτευχθεί δίνοντας προσοχή σε στοιχεία, όπως τον τίτλο, τις

επικεφαλίδες, τη περιγραφή, τις διευθύνσεις (urls), το περιεχόμενο και την εγκυρότητα του κώδικα. Επίσης, η χρήση συνδέσμων και χαρτών πλοήγησης (sitemap) περιέχεται σ' αυτή τη κατηγορία.

Η συνολική ασφάλεια υπολογίζεται από μετρικές που αφορούν την ασφάλεια των χρηστών κατά η πρόσβαση στις ιστοσελίδες, αλλά και των ίδιων των ιστοσελίδων από "κυβερνοεπιθέσεις". Η χρήση του https στις διευθύνσεις των ιστοσελίδων, όπως και του HSTS κατά την είσοδο των χρηστών αποτελούν καίρια σημεία αναφοράς. Επιπλέον, το είδος της απάντησης από τους περιηγητές κατά την είσοδο στις ιστοσελίδες και η προστασία από επιθέσεις clickjack από διαφημίσεις συναποτελούν τις μετρικές ασφάλειας. Η συνολική ταχύτητα περιλαμβάνει μετρικές που αφορούν την εγκυρότητα του κώδικα Javascript και CSS, το μέγεθος των ιστοσελίδων, καθώς και τη χρήση ανακατευθύνσεων για μη λειτουργικές σελίδες και της προσωρινής αποθήκευσης (caching) για γρηγορότερη φόρτωση.

3.3. Άντληση δεδομένων χρηστών

Το SimilarWeb είναι μία διαδικτυακή υπηρεσία, η οποία παρέχεται δωρεάν. Συλλέγει και διαμοιράζει στοιχεία σχετικά με τη συμπεριφορά των χρηστών κατά την επίσκεψη τους στις ιστοσελίδες. Το πρόγραμμα με την έναρξη του αρχίζει και καταγράφει τα στοιχεία που συλλέχθηκαν από τις ιστοσελίδες. Τα αποτελέσματα εμφανίζονται με βάση τη ποσότητα τους, ενώ στις παραγράφους που ακολουθούν παρουσιάζονται οι μετρικές που αφορούν χαρακτηριστικά συμπεριφοράς. Η γνώση του τρόπου περιήγησης των χρηστών και των αδύναμων σημείων μπορούν να οδηγήσουν σε μικρές αλλαγές στις ιστοσελίδες και να καταταγούν σε υψηλότερες θέσεις, αυξάνοντας τη προσβασιμότητα. Η επιλογή του συγκεκριμένου εργαλείου έγινε λόγω της ακρίβειας στην εμφάνιση των αποτελεσμάτων έναντι άλλων γνωστών προγραμμάτων.

3.3.1. SimilarWeb

Το SimilarWeb (SimilarWeb) είναι μία ιστοσελίδα που παρέχει πληροφορίες σχετικά με ποικίλα στοιχεία επισκεψιμότητας των ιστοτόπων. Τα στοιχεία που συλλέγονται, είναι η μέση διάρκεια επίσκεψης των χρηστών (Visit Duration), ο μέσος αριθμός των σελίδων ανά επίσκεψη (Pages per visit), το ποσοστό εξόδου από την ιστοσελίδα (Bounce Rate), το ποσοστό των επισκεπτών από τα «οργανικά» αποτελέσματα των μηχανών αναζήτησης (Percent of Organic Traffic). Η εξαγωγή των στοιχείων γίνεται με την αποδοχή των cookies από τους χρήστες, τα οποία καταγράφουν την διαδικτυακή δραστηριότητα, δηλαδή τις ιστοσελίδες που επισκέπτονται και το περιεχόμενο. Πιο συγκεκριμένα, η μέση διάρκεια επίσκεψης των χρηστών δίνει το μέσο χρονικό διάστημα περιήγησης των χρηστών στην ιστοσελίδα. Η μετρική αυτή δείχνει το βαθμό σχετικότητας του περιεχομένου με την αναζήτηση τους, όπως και ο μέσος αριθμός των σελίδων ανά επίσκεψη. Το ποσοστό εξόδου από την ιστοσελίδα υποδεικνύει τη έλλειψη συνάφειας και την αδυναμία εκπλήρωσης των πληροφοριακών αναγκών. Οι βελτιστοποιημένες ιστοσελίδες θα κρατήσουν τους επισκέπτες για μεγαλύτερη διάρκεια, αφήνοντας τους να εξερευνήσουν όλες τις θεματικές κατηγορίες. Το μεγαλύτερο ποσοστό των επισκεπτών έρχεται από τις μηχανές αναζήτησης (Οργανικά αποτελέσματα). Εκατομμύρια χρήστες αναζητούν καθημερινά ποικίλα θέματα που τους αφορούν και κατευθύνονται αμέσως στις ιστοσελίδες με το πιο σχετικό περιεχόμενο.



Εικόνα 4. Ενδεικτική οθόνη εμφάνισης των αποτελεσμάτων - SimilarWeb

Με τη συλλογή των παραπάνω στοιχείων δίνεται η δυνατότητα να ανακαλυφθούν οι καλύτερες τεχνικές βελτιστοποίησης των ιστοσελίδων στις μηχανές αναζήτησης. Η προσεκτική ανάλυση των στοιχείων μπορεί να δώσει σημαντικές πληροφορίες για τα δυνατά σημεία, αλλά και να αναδείξει τις αδυναμίες συγκριτικά πάντοτε με τα υπόλοιπα στοιχεία SEO, ταχύτητας και ασφάλειας.

3.4. Επιλογή δείγματος ιστοτόπων

Στην εργασία συμπεριλήφθηκαν συνολικά 143 οργανισμοί πληροφόρησης. Ο εννοιολογικός διαχωρισμός έγινε βάση το είδος των οργανισμών και της προέλευσης τους. Η αναζήτηση και η συλλογή των διευθύνσεων έγινε χειροκίνητα με τη βοήθεια του καταλόγου ΣΕΑΒ (Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών) για τις ακαδημαϊκές βιβλιοθήκες. Από αυτούς τους οργανισμούς πληροφόρησης οι 39 αφορούν ιστοσελίδες ακαδημαϊκών βιβλιοθηκών, 14 αρχειακών οργανισμών, 18 μουσείων, 13 εθνικών βιβλιοθηκών, 22 ερευνητικών κέντρων και 37 πανεπιστημιακών ιδρυμάτων. Οι 71 διευθύνσεις είναι ελληνικών οργανισμών πληροφόρησης, ενώ οι υπόλοιπες 72 είναι ξένων. Έγινε προσπάθεια για τη συλλογή ισάριθμων διευθύνσεων προκειμένου να είναι αντικειμενική η σύγκριση και η επακόλουθη εξαγωγή των αποτελεσμάτων. Όσον αφορά τα συστήματα διαχείρισης περιεχομένου, υπάρχει πληθώρα αυτών που χρησιμοποιούνται. Ενδεικτικά παρατίθενται όλα τα συστήματα διαχείρισης περιεχομένου που εντοπίστηκαν κατά τη συλλογή τους με το διαδικτυακό εργαλείο whatcms.org σε φθίνουσα σειρά: Custom (31), Drupal (30), Wordpress (23), Joomla (18), TYPO3 (4), Adobe Experience Manager (3), Bloomreach (1), Contentful (1), Open CMS (1), Ruby on Rails (1), Weebly (1), Adobe Dreamweaver (1), Bootply (1), dotCMS (1), Hubspot (1), Laravel (1), PHP – Nuke (1), SharePoint (1), SiteCore (1), Bitrix (1), FrontPage (1), Hugo (1), Liferay (1), Umbraco (1).

3.4.1. Προσδιορισμός τύπου CMS - Whatcms.org

Το [Whatcms.org](http://whatcms.org) (whatcms.org) είναι ένα δωρεάν διαδικτυακό εργαλείο, το οποίο αναγνωρίζει το σύστημα διαχείρισης περιεχομένου (Content Management System – CMS) που χρησιμοποιείται σ' έναν ιστότοπο. Επίσης, παρέχονται περαιτέρω πληροφορίες για τις ιστοσελίδες όπως είναι η έκδοση του CMS, οι γλώσσες προγραμματισμού που χρησιμοποιήθηκαν, οι βάσεις δεδομένων και το είδος server.

The image shows two screenshots from the WhatCMS.org website. The top screenshot is for 'alis.uniwa.gr' and shows the following technical details:

Category	Software	Version
CMS, Blog	WordPress	5.2.2
Programming Language	PHP	
Database	MySQL	
Web Server	Nginx	

The bottom screenshot is for 'uniwa.gr' and shows the following information:

Checked Pages	CMS
11	WordPress

Εικόνα 5. Ενδεικτική οθόνη εμφάνισης των αποτελεσμάτων - Whatcms.org

3.5. Στατιστικοί παράμετροι

Στη πρώτη κατηγορία επιλέχθηκαν συγκεκριμένες μετρικές, οι οποίες φαίνεται ότι έχουν μεγάλη επιρροή στην εμφάνιση των αποτελεσμάτων. Πιο αναλυτικά επιλέχθηκαν στοιχεία απόδοσης των μεταδεδομένων (Metadata Performance), της απόδοσης των στοιχείων «αρχιτεκτονικής» των ιστοσελίδων (Architecture Structure Performance), της απόδοσης της ασφάλειας (Security Performance) και της συνολικής απόδοσης & του μεγέθους (Overall Performance and Size) και των δεδομένων συμπεριφοράς (Behavioral Analytics) τα οποία περιλαμβάνουν τα εξής: 1) Set page titles, 2) Avoid thin content pages, 3) Avoid URL parameters, 4) Use robots.txt files, 5) Set sitemap locations, 6) Crawled URLs 7) Use HTTPS, 8) Pages per visit, 9) SEO, 10) Speed, 11) Security.

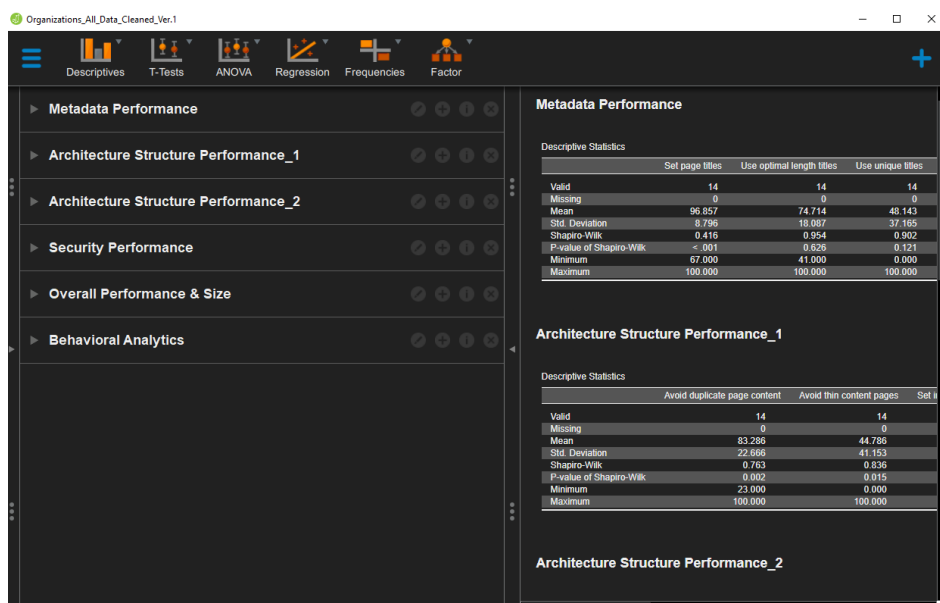
Χρησιμοποιήθηκαν τα εξής στατιστικά στοιχεία για την ανάλυση των μετρικών:

- Το “mean” είναι ο μέσος όρος των μετρικών και συνυπολογίζεται προσθέτοντας το σύνολο των αντικειμένων και διαιρώντας με το πλήθος τους.
- Το “Standard deviation” δείχνει το ποσό της μεταβολής ή της διασποράς ενός συνόλου τιμών δεδομένων.
- Το “Shapiro – Wilk” αποτελεί δείκτη κανονικότητας, δηλαδή χρησιμοποιείται για να εξεταστεί αν σ’ ένα σύνολο δεδομένων (data set) έχει γίνει σωστή κατανομή για αποτελεσματική ανάλυση.
- Η τιμή “P-value of Shapiro-Wilk” υποδεικνύει τη πιθανότητα ενάντια στη μηδενική υπόθεση (null hypothesis). Οι μικρές τιμές παρέχουν ισχυρότερα στοιχεία ενάντια στη μηδενική υπόθεση.
- Οι τιμές “Minimum” και “Maximum” δείχνουν τη μικρότερη και τη μέγιστη δυνατή τιμή για τις μετρικές.

3.6. Χρήση του JASP

Το JASP είναι ένα ελεύθερο λογισμικό ανοιχτού κώδικα για στατιστική ανάλυση, το οποίο διατίθεται δωρεάν από το Πανεπιστήμιο του Άμστερνταμ. Έχει σχεδιαστεί με συγκεκριμένο τρόπο ώστε να είναι εύκολο στη χρήση, προσφέροντας τυποποιημένα αναλυτικά αποτελέσματα. Προωθεί την διάδοση της γνώσης και το διαμοιρασμό των δεδομένων για τη πολύπλευρη ανάλυση και ερμηνεία τους μέσω της ενσωμάτωσης του Open Science Framework. Κατά την περιήγηση στο πρόγραμμα μπορεί κανείς να επιλέξει το είδος ανάλυσης που επιθυμεί, καθώς και τις επιμέρους μετρικές που χρειάζεται. Το φιλικό προς το χρήστη περιβάλλον αναδεικνύεται από τη διάταξη του υπολογιστικού φύλλου και τη δυνατότητα drag-and-drop των στοιχείων. Η σταδιακή παράθεση των δεδομένων σε συνδυασμό με τη δυνατότητα διαχωρισμού (split) και φιλτραρίσματος (filter) τους προβάλλει την ευκολία και τη διαλειτουργικότητα του προγράμματος. Επίσης, η γρήγορη έξοδος των αποτελεσμάτων και η αποθήκευση τους σε αρχείο μορφής .jasp κάνει πιο εύκολο και προσιτό το έργο των φοιτητών και των ερευνητών.

Για την εργασία χρησιμοποιήθηκε η επιλογή “Descriptive Statistics”, δηλαδή τα περιγραφικά στατιστικά για τους οργανισμούς πληροφόρησης. Τα επιμέρους στοιχεία του εργαλείου SEO Checkbot οργανώθηκαν με βάση τους εξής νοηματικούς διαχωρισμούς σε: 1) Metadata Performance (Απόδοση των Μεταδεδομένων), 2) Architecture Structure Performance (Απόδοση της Δομής των ιστοσελίδων), 3) Security Performance (Απόδοση της Ασφάλειας), 4) Overall Performance & Size (Συνολική απόδοση και μέγεθος), 5) Behavioral Analytics (Ανάλυση συμπεριφοράς χρηστών).



Εικόνα 6. Ενδεικτική οθόνη εμφάνισης των αποτελεσμάτων στο περιβάλλον JASP

Η απόδοση των μεταδεδομένων (Metadata Performance) περιλαμβάνει τα εξής στοιχεία: Set page titles, Use optimal length titles, Use unique titles, Set H1 headings, Use one H1 heading per page, Use optimal length H1 headings, Use unique H1 headings, Set page descriptions, Use optimal length descriptions, Use unique descriptions. Η απόδοση της δομής των ιστοσελίδων (Architecture Structure Performance) εμπεριέχει: Avoid duplicate page content, Avoid thin content pages, Set image ALT text, Set mobile scaling, Avoid

plugins, Use short URLs, Avoid URL extensions, Avoid URL parameters, Avoid symbols in URLs, Use lowercase URLs, Avoid underscores in URLs, Avoid deeply nested URLs, Use valid HTML, CSS, JavaScript, Use 404 code for broken URLs, Use robots.txt files, Set sitemap locations, Avoid broken internal / external links, Avoid broken page resources. Η κατηγορία απόδοση της ασφάλειας Security Performance)έχει τα εξής στοιχεία: Use HTTPS, Avoid mixed content, Use HSTS, Use clickjack protection, Hide server version data.

Η συνολική απόδοση και το μέγεθος (Overall Performance & Size) περιλαμβάνει τα εξής στοιχεία: Total Website SEO Score, SEO, Speed, Security, Pages, Internal URLs, External URLs, Images, CSS, Javascript.

Η ανάλυση συμπεριφοράς χρηστών (Behavioral Analytics) περιέχει τα εξής στοιχεία από το πρόγραμμα Similar Web: Visit Duration, Pages per Visit, Bounce Rate, Percent of Organic Traffic.

Κεφάλαιο 4. Αποτελέσματα

Οι ξένες ιστοσελίδες οργανισμών πληροφόρησης που επιλέχθηκαν για τη συγκεκριμένη εργασία υπερिशύουν έναντι των ελληνικών σε όλες τις μετρικές με μικρή ή μεγάλη διαφορά. Παρακάτω παρατίθενται έξι πίνακες που δημιουργήθηκαν κατά τη διάρκεια ανάλυσης των επιλεγμένων μετρικών. Ο πρώτος πίνακας αναφέρεται στα στοιχεία απόδοσης των μεταδεδομένων (Metadata Performance), ο δεύτερος πίνακας στα στοιχεία απόδοσης «αρχιτεκτονικής» των ιστοσελίδων (Architecture Structure Performance), ο τρίτος στα στοιχεία της απόδοσης της ασφάλειας (Security Performance), ο τέταρτος πίνακας στα στοιχεία της συνολικής απόδοσης & του μεγέθους των ιστοσελίδων (Overall Performance and Size) και ο πέμπτος αναφέρεται στα στοιχεία των δεδομένων συμπεριφοράς των χρηστών (Behavioral Analytics). Ο τελευταίος πίνακας περιέχει τα στοιχεία συγκριτικά με τα συστήματα διαχείρισης περιεχομένου. Όλοι οι πίνακες με τα τεχνικά χαρακτηριστικά χωρίζονται σε δύο στήλες με βάση την προέλευση των οργανισμών πληροφόρησης. Ο πίνακας με τα συστήματα διαχείρισης περιεχομένου έχει χωριστεί σε τρεις κατηγορίες με βάση τα τρία πιο δημοφιλή. Οι πίνακες παρουσιάζουν τις τιμές των περιγραφικών στοιχείων, το βαθμό διακύμανσης των τιμών (Std. Deviation), την αποκλίνουσα τιμή (Shapiro-Wilk), την τιμή σημαντικότητας (P-value of Shapiro-Wilk) και τη μικρότερη και μεγαλύτερη δυνατή τιμή (Minimum, Maximum).

4.1. Αποτελέσματα Metadata performance

Ο μέσος όρος για τον τίτλο της σελίδας (Set page titles) βρίσκεται στο 97,8% για τις ξένες ιστοσελίδες με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 8,9. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,235 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Για τις ελληνικές ιστοσελίδες ο μέσος όρος είναι 95,5% με το βαθμό διακύμανσης των τιμών στο 14,5. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,341 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Οι μέσοι όροι βρίσκονται σε υψηλά επίπεδα για τις ξένες ιστοσελίδες και τις ελληνικές. Ο βαθμός διακύμανσης των τιμών όμως είναι μεγαλύτερος κατά 5,6 μονάδες για τις ελληνικές ιστοσελίδες, γεγονός που δείχνει τη μεγαλύτερη κλίμακα τιμών ανά περίπτωση. Η μετρική Shapiro – Wilk είναι επίσης λίγο μεγαλύτερη για τις ελληνικές, δηλαδή η αποκλίνουσα τιμή από τη κανονική. Η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι μικρότερη του 0.05 και απορρίπτεται αμέσως.

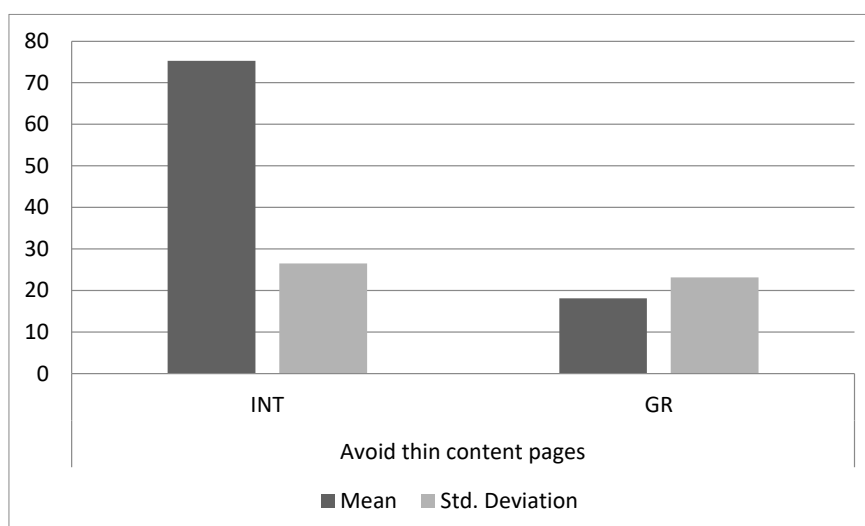
Όπως παρουσιάζεται στο πίνακα 1 οι μέσοι όροι για τον τίτλο της σελίδας (97,8%-95,5%) είναι ικανοποιητικός και για τις δύο. Στις περισσότερες ξένες και ελληνικές ιστοσελίδες έχει τοποθετηθεί ορθά το πεδίο του τίτλου στον κώδικα, με μικρές βέβαια αποκλίσεις. Η τυπική απόκλιση (Std. Deviation) των τιμών είναι 8,9 και 14,5 αντίστοιχα υποδεικνύοντας χαμηλή διασπορά για τις ξένες ιστοσελίδες και υψηλότερη για τις ελληνικές.

Πίνακας 1 Στοιχεία απόδοσης των μεταδεδομένων

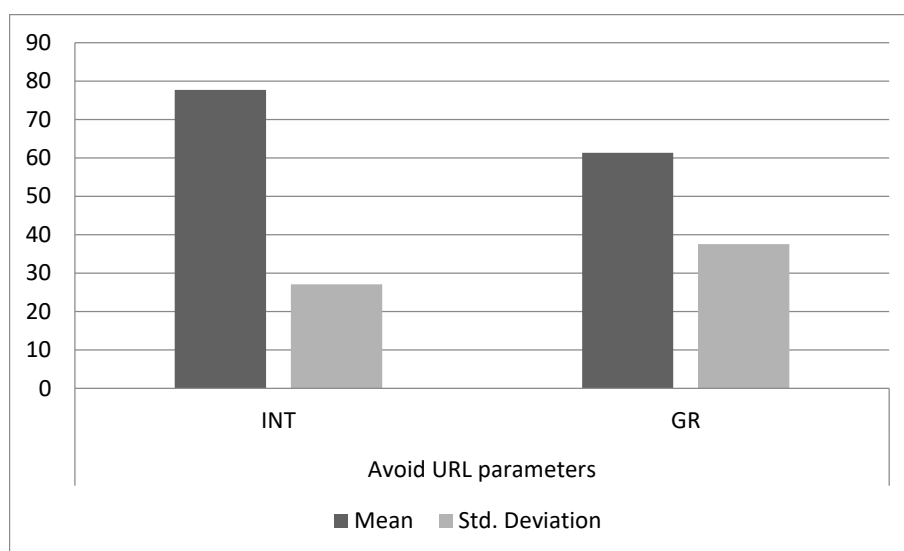
Metadata performance		
	Set page titles	
	INT	GR
Valid	72	71
Missing	0	0
Mean	97,875	95,563
Std. Deviation	8,958	14,548
Shapiro-Wilk	0,235	0,341
P-value of Shapiro-Wilk	< .001	< .001
Minimum	32	0
Maximum	100	100

4.2. Αποτελέσματα Architecture Structure performance

Ο μέσος όρος για το περιεχόμενο της ιστοσελίδας (Avoid thin content pages) βρίσκεται στο 75,2% για τις ξένες ιστοσελίδες με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 26,5. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,847 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Για τις ελληνικές ιστοσελίδες ο μέσος όρος είναι 18% με το βαθμό διακύμανσης των τιμών στο 23,1. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,751 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Ο μέσος όρος για τις ξένες ιστοσελίδες είναι πολύ μεγαλύτερος (57%), και αναδεικνύει το «φτωχό» και μη σχετικό περιεχόμενο των ελληνικών ιστοσελίδων με τις αναζητήσεις των χρηστών στις μηχανές αναζήτησης (βλέπε Γράφημα 1). Ο βαθμός διακύμανσης των τιμών είναι στα ίδια επίπεδα με μέτρια διακύμανση των τιμών και για τις δύο. Η μετρική Shapiro – Wilk είναι επίσης λίγο μεγαλύτερη για τις ξένες. Η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι μικρότερη του 0.05 και απορρίπτεται αμέσως επίσης.

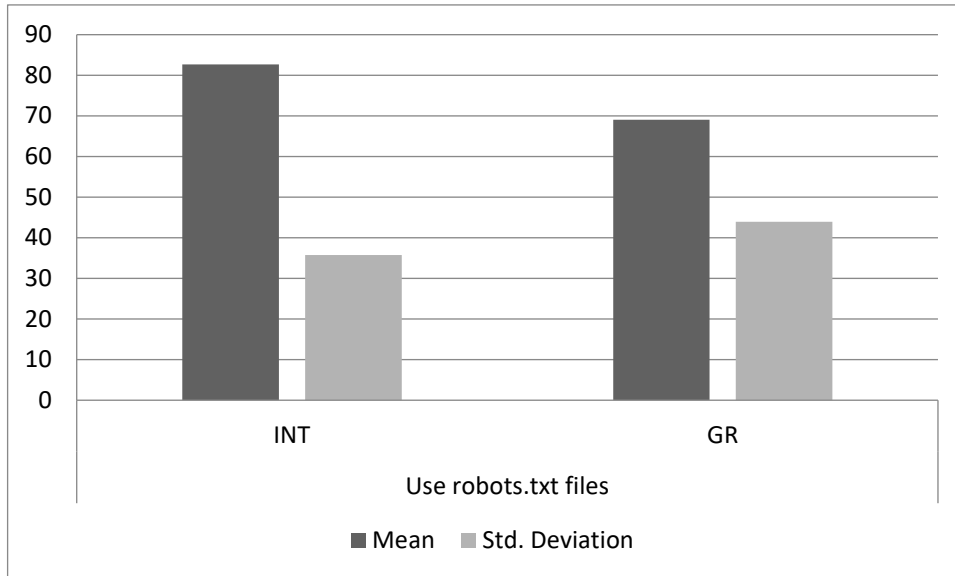


Γράφημα 1. Μέσος όρος και απόκλιση για το περιεχόμενο της ιστοσελίδας (Avoid thin content pages) – Σύγκριση ελληνικών και ξένων ιστοτόπων. Ο μέσος όρος για τη χρήση των παραμέτρων στις διευθύνσεις (Avoid URL parameters) βρίσκεται στο 77,6% για τις ξένες ιστοσελίδες με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 27,1. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,803 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Για τις ελληνικές ιστοσελίδες ο μέσος όρος είναι 61,3% με το βαθμό διακύμανσης των τιμών στο 37,5 (βλέπε Γράφημα 2). Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,839 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Ο μέσος όρος για τις ξένες ιστοσελίδες είναι μεγαλύτερος κατά 16 μονάδες, αλλά παραμένει σε ικανοποιητικό επίπεδο και για τις δύο. Ο βαθμός διακύμανσης των τιμών (Std. Deviation) είναι επίσης μεγαλύτερος κατά 10 μονάδες περίπου για τις ελληνικές, υποδεικνύοντας τη μεγάλη διαφορά ανάμεσα στους εκάστοτε οργανισμούς. Η μετρική Shapiro – Wilk έχει την ίδια τιμή. Η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι μικρότερη του 0.05 και απορρίπτεται.



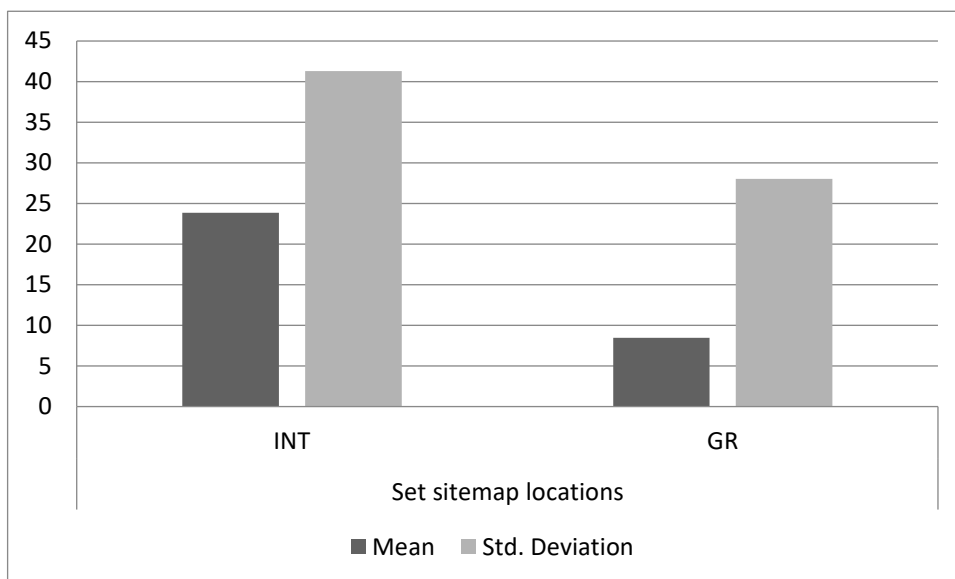
Γράφημα 2. Μέσος όρος και απόκλιση για τη χρήση των παραμέτρων στις διευθύνσεις (Avoid URL parameters) – Σύγκριση ελληνικών και ξένων ιστοτόπων

Ο μέσος όρος για τα αρχεία robots.txt (Use robots.txt files) βρίσκεται στο 82,6% για τις ξένες ιστοσελίδες με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 35,7. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,512 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Για τις ελληνικές ιστοσελίδες ο μέσος όρος είναι 68,9% με το βαθμό διακύμανσης των τιμών στο 43,9 (βλέπε Γράφημα 3). Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,643 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Ο μέσος όρος για τις ξένες ιστοσελίδες είναι μεγαλύτερος κατά 14 μονάδες, αλλά παραμένει σε ικανοποιητικό επίπεδο και για τις δύο. Ο βαθμός διακύμανσης των τιμών είναι στα ίδια επίπεδα με μέτρια προς μεγάλη κλίμακα τιμών αμφότερα. Η μετρική Shapiro – Wilk έχει την ίδια σχεδόν τιμή, ενώ η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001 και απορρίπτεται.



Γράφημα 3. Μέσος όρος και απόκλιση για τα αρχεία robots.txt – Σύγκριση ελληνικών και ξένων ιστοτόπων

Ο μέσος όρος για τους χάρτες πλοήγησης (Set sitemap locations) βρίσκεται στο 23,8% για τις ξένες ιστοσελίδες με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 41,2. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,567 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι $<.001$. Για τις ελληνικές ιστοσελίδες ο μέσος όρος είναι 8,4% με το βαθμό διακύμανσης των τιμών στο 28 (βλέπε Γράφημα 4). Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,311 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι $<.001$. Ο μέσος όρος για τη συγκεκριμένη μετρική έχει χαμηλό ποσοστό, διότι αφορά μία όχι τόσο γνωστή αλλά αποτελεσματική μέθοδο για τη βελτιστοποίηση ιστοσελίδων. Στις ελληνικές ιστοσελίδες το ποσοστό είναι ακόμα μικρότερο αποδεικνύοντας το γεγονός ότι δεν λαμβάνεται υπόψη κατά τη δημιουργία ιστοσελίδων. Το τεστ κανονικότητας Shapiro – Wilk διαφέρει υπέρ των ξένων ιστοσελίδων και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι $<.001$ και απορρίπτεται.



Γράφημα 4. Μέσος όρος και απόκλιση για τους χάρτες πλοήγησης (Set sitemap locations) – Σύγκριση ελληνικών και ξένων ιστοτόπων

Στον πίνακα 2 παρουσιάζονται οι μέσοι όροι για το περιεχόμενο της ιστοσελίδας (75,2%-8%) με τη τυπική απόκλιση (Std. Deviation) των τιμών να είναι 26,5 και 23,1, τη χρήση των παραμέτρων στις διευθύνσεις (77,6%-61,3%) με βαθμό διακύμανσης στο 27,1 και 37,5, τα αρχεία robots.txt (82,6%-68,9%) με βαθμό διακύμανσης στο 35,7 και 43,9, τους χάρτες πλοήγησης (23,8% - 8,4%) με τη τυπική απόκλιση στο 41,2 και 28. Οι μέσοι όροι για το περιεχόμενο της ιστοσελίδας και τους χάρτες πλοήγησης παρουσιάζουν τεράστια διαφορά μεταξύ ξένων και ελληνικών ιστοτόπων. Η έλλειψη ποιοτικού περιεχομένου επηρεάζει αρνητικά την ανακαλυψιμότητα των ιστοσελίδων και οι χάρτες πλοήγησης την εμπειρία πλοήγησης των χρηστών. Οι υπόλοιπες δύο μετρικές για την απόδοση των μεταδεδομένων επικρατούν υπέρ των ξένων ιστοσελίδων, αλλά με αξιοπρεπή ποσοστά και για τις ελληνικές.

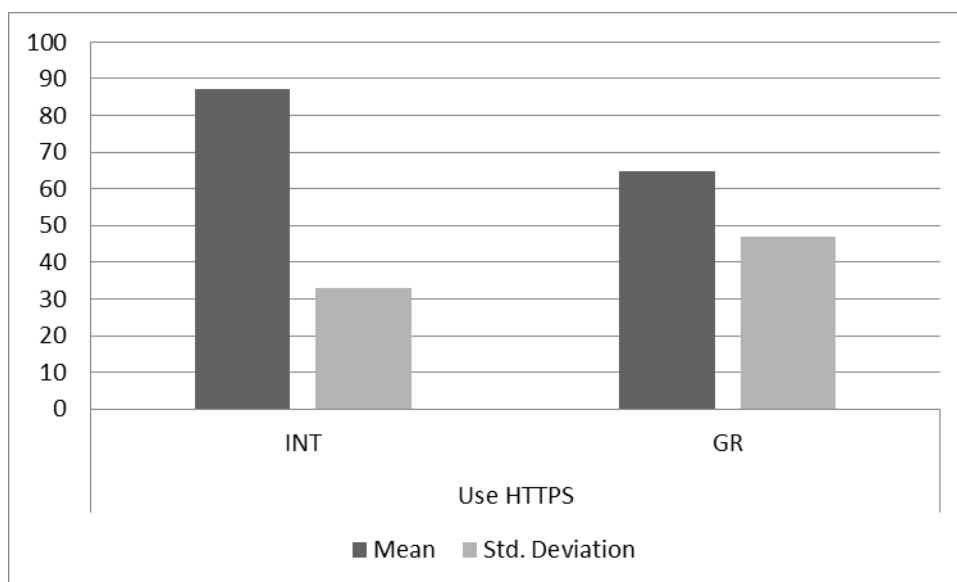
Πίνακας 2 - Στοιχεία απόδοσης "αρχιτεκτονικής" των ιστοσελίδων

Architecture Structure performance								
	Avoid thin content pages		Avoid URL parameters		Use robots.txt files		Set sitemap locations	
	INT	GR	INT	GR	INT	GR	INT	GR
Valid	72	71	72	71	72	71	72	71
Missing	0	0	0	0	0	0	0	0
Mean	75,264	18,099	77,694	61,324	82,639	68,986	23,833	8,451
Std. Deviation	26,535	23,139	27,107	37,512	35,761	43,944	41,275	28,013
Shapiro-Wilk	0,847	0,751	0,803	0,839	0,512	0,643	0,567	0,311
P-value of Shapiro-Wilk	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
Minimum	3	0	0	0	0	0	0	0
Maximum	100	100	100	100	100	100	100	100

4.3. Αποτελέσματα Security performance

Ο μέσος όρος για τη χρήση του https στις διευθύνσεις (Use HTTPS) βρίσκεται στο 87% για τις ξένες ιστοσελίδες με βαθμό διακύμανσης των τιμών (Std. Deviation) στις 32,8. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,416 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Για τις ελληνικές ιστοσελίδες ο μέσος όρος είναι 64,9 με το βαθμό διακύμανσης των τιμών στο 47 (βλέπε Γράφημα 5). Το τεστ κανονικότητας

Shapiro – Wilk βρίσκεται στο 0,629 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Ο μέσος όρος για τους ξένους οργανισμούς είναι μεγαλύτερος, καθώς έχει χαμηλότερο ποσοστό διακύμανσης τιμών. Το ποσοστό για τους ελληνικούς οργανισμούς είναι σε μέτρια προς χαμηλά επίπεδα για τη συγκεκριμένη μετρική, η οποία είναι απολύτως σημαντική και απαραίτητη. Το τεστ κανονικότητας των τιμών διαφέρει υπέρ των ελληνικών ιστοσελίδων και η τιμή σημαντικότητας P-value of Shapiro – Wilk απορρίπτεται.



Γράφημα 5. Μέσος όρος και απόκλιση για τη χρήση του https στις διευθύνσεις (Use HTTPS) – Σύγκριση ελληνικών και ξένων ιστοτόπων

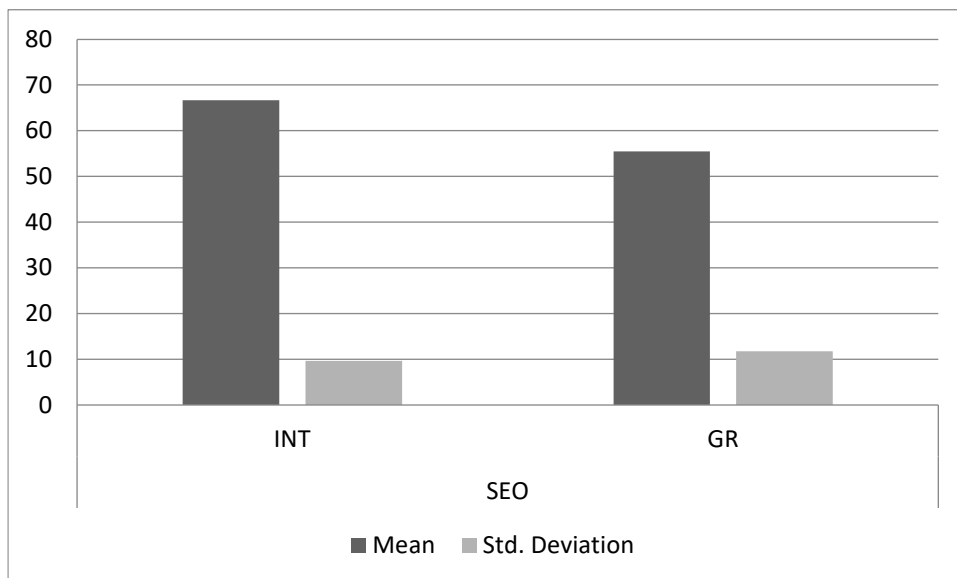
Στον πίνακα 3 παρουσιάζονται οι μέσοι όροι για τη χρήση του στοιχείου https στις διευθύνσεις (87%-64,9%). Τα ποσοστά είναι ικανοποιητικά και για τις δύο. Στις περισσότερες ξένες και ελληνικές ιστοσελίδες έχει τοποθετηθεί ορθά το στοιχείο https στις διευθύνσεις. Σίγουρα πάντως υπάρχουν περιθώρια βελτίωσης στους ελληνικούς ιστότοπους. Η τυπική απόκλιση (Std. Deviation) των τιμών είναι 32,8 και 47 αντίστοιχα και επιδεικνύουν υψηλή διασπορά των τιμών για τις ξένες και για τις ελληνικές ιστοσελίδες.

Πίνακας 3 - Στοιχεία απόδοσης της ασφάλειας

Security performance		
	Use HTTPS	
	INT	GR
Valid	72	71
Missing	0	0
Mean	87,069	64,944
Std. Deviation	32,841	47,007
Shapiro-Wilk	0,416	0,629
P-value of Shapiro-Wilk	< .001	< .001
Minimum	0	0
Maximum	100	100

4.4. Αποτελέσματα Overall performance & size

Ο μέσος όρος για το συνολικό SEO (SEO) βρίσκεται στο 66% για τις ξένες ιστοσελίδες με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 9,6. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 951 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 7. Για τις ελληνικές ιστοσελίδες ο μέσος όρος είναι 55,4% με το βαθμό διακύμανσης των τιμών στο 11,7. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 983 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 437. Ο μέσος όρος των ξένων ιστοσελίδων είναι μεγαλύτερος κατά 10,6% σημειώνοντας σημαντική διαφορά. Η σύγκριση των παραπάνω στοιχείων για τη βελτιστοποίηση διαφόρων παραγόντων δείχνει την υπεροχή των ξένων έναντι των ελληνικών ιστοτόπων. Η διακύμανση των τιμών είναι ελάχιστα μεγαλύτερη στις ελληνικές ιστοσελίδες και η αποκλίνουσα τιμή από τη κανονική είναι μεγάλη και στις δύο περιπτώσεις. Η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι μεγάλη αμφότερα και επιβεβαιώνει την ακραία παρατήρηση.

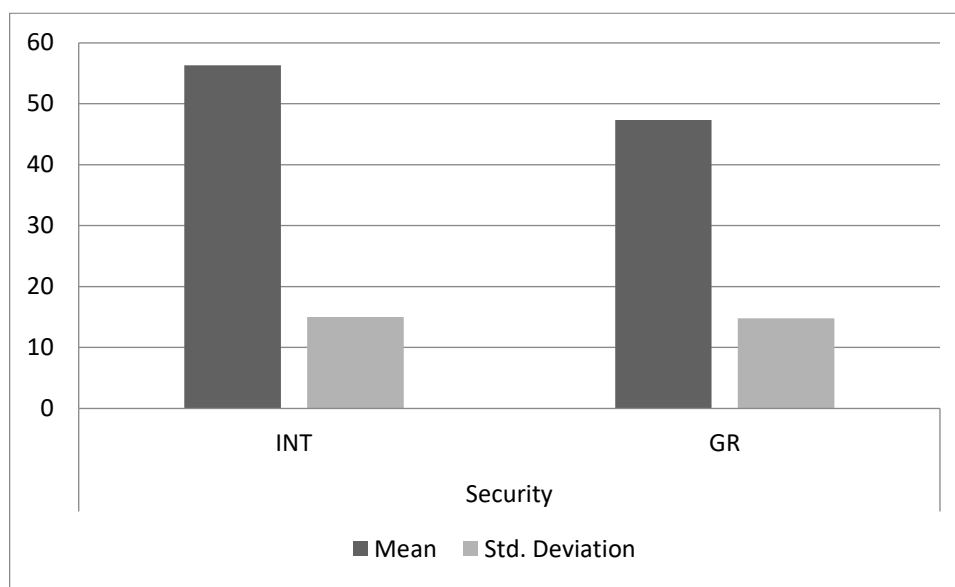


Γράφημα 6. Μέσος όρος και απόκλιση για το συνολικό SEO (SEO) – Σύγκριση ελληνικών και ξένων ιστοτόπων

Ο μέσος όρος για τη συνολική ταχύτητα (Speed) βρίσκεται στο 70% για τις ξένες ιστοσελίδες με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 9,4. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 988 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 746. Για τις ελληνικές ιστοσελίδες ο μέσος όρος είναι 72,8% με το βαθμό διακύμανσης των τιμών στο 9,8. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 974 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 148. Ο μέσος όρος είναι μεγαλύτερος για τις ελληνικές ιστοσελίδες κατά 2,8% γεγονός που προφανώς οφείλεται στις περισσότερες ανιχνευμένες διευθύνσεις ιστοσελίδων (Crawled URLs). Ο βαθμός διακύμανσης των τιμών (Std. Deviation) είναι στα ίδια επίπεδα, όπως και ο δείκτης κανονικότητας Shapiro – Wilk. Η τιμή σημαντικότητας είναι μεγάλη και για τις δύο επιβεβαιώνοντας τη παρατήρηση.

Ο μέσος όρος για τη συνολική ασφάλεια (Security) βρίσκεται στο 56,3% για τις ξένες ιστοσελίδες με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 14,9. Το τεστ

κανονικότητας Shapiro – Wilk βρίσκεται στο 954 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 10. Για τις ελληνικές ιστοσελίδες ο μέσος όρος είναι 47,3% με το βαθμό διακύμανσης των τιμών στο 14,8. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 930 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Ο μέσος όρος των τιμών είναι επίσης μεγαλύτερος για τις ξένες ιστοσελίδες κατά 6%. Η διακύμανση των τιμών και η τιμή κανονικότητας είναι στο ίδιο ακριβώς επίπεδο. Η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι μεγάλη για τις ξένες ιστοσελίδες επιβεβαιώνοντας τη παρατήρηση σε αντίθεση με τις ελληνικές.



Γράφημα 7. Μέσος όρος και απόκλιση για τη συνολική ασφάλεια (Security) – Σύγκριση ελληνικών και ξένων ιστοτόπων

Ο μέσος όρος για το συνολικό SEO είναι στο 66% και 55,4% για τις ξένες και ελληνικές ιστοσελίδες. Ο βαθμός διακύμανσης των τιμών (Std. Deviation) στο 9,6 και 11,7 αντίστοιχα. Ο μέσος όρος για τη συνολική ταχύτητα (Speed) βρίσκεται στο 70% και 72,8%. Ο βαθμός διακύμανσης των τιμών (Std. Deviation) είναι στο 9,4 και 9,8. Η μικρή διαφορά υπέρ των ελληνικών ιστοσελίδων οφείλεται στον αριθμό των ανιχνευμένων διευθύνσεων και στο μικρό πλήθος των πληροφοριών που παρέχουν στο κοινό τους. Ο μέσος όρος για τη συνολική ασφάλεια (Security) βρίσκεται στο 56,3% και 47,3%. Ο βαθμός διακύμανσης των τιμών (Std. Deviation) στο 14,9 και 14,8. Οι ανιχνευμένες διευθύνσεις ιστοσελίδων (Crawled URLs) είναι 6.983 και 6.410. Τα αποτελέσματα για τις τρεις μεγαλύτερες μετρικές επιδεικνύουν την επικράτηση των ξένων ιστοσελίδων στα τεχνικά χαρακτηριστικά SEO και την ασφάλεια, στην οποία μπορούν να πραγματοποιηθούν σημαντικές αλλαγές τόσο για τις ξένες όσο και τις ελληνικές ιστοσελίδες.

Ο μέσος αριθμός για τις ανιχνευμένες διευθύνσεις ιστοσελίδων (Crawled URLs) βρίσκεται στις 6.410 για τις ξένες ιστοσελίδες με βαθμό διακύμανσης των τιμών (Std. Deviation) στις 3.745. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 808 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Για τις ελληνικές ιστοσελίδες ο μέσος όρος είναι 6.983 με το βαθμό διακύμανσης των τιμών στο 3.539. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 771 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι

<.001. Ο μέσος όρος για αυτή τη μετρική είναι μεγαλύτερος για τις ελληνικές ιστοσελίδες, έχοντας ταυτόχρονα χαμηλότερο βαθμό διακύμανσης τιμών απ’ ότι οι ξένες. Το τεστ κανονικότητας των τιμών διαφέρει υπέρ των ξένων ιστοσελίδων και η τιμή σημαντικότητας P-value of Shapiro – Wilk απορρίπτεται.

Πίνακας 4 - Στοιχεία της συνολικής απόδοσης & του μεγέθους των ιστοσελίδων

Overall performance & size								
	SEO		Speed		Security		Crawled URLs	
	INT	GR	INT	GR	INT	GR	INT	GR
Valid	72	71	72	71	72	71	72	71
Missing	0	0	0	0	0	0	0	0
Mean	66,694	55,451	70,056	72,873	56,319	47,338	6.410	6.983
Std. Deviation	9,686	11,776	9,441	9,863	14,984	14,808	3.745	3.539
Shapiro-Wilk	951	983	988	974	954	930	808	771
P-value of Shapiro-Wilk	7	437	746	148	10	< .001	< .001	< .001
Minimum	30	34	50	53	23	23	278	580
Maximum	90	88	97	92	95	83	10.001	10.000

4.4.1 Αποτελέσματα Overall performance & size με βάση τον τύπο CMS

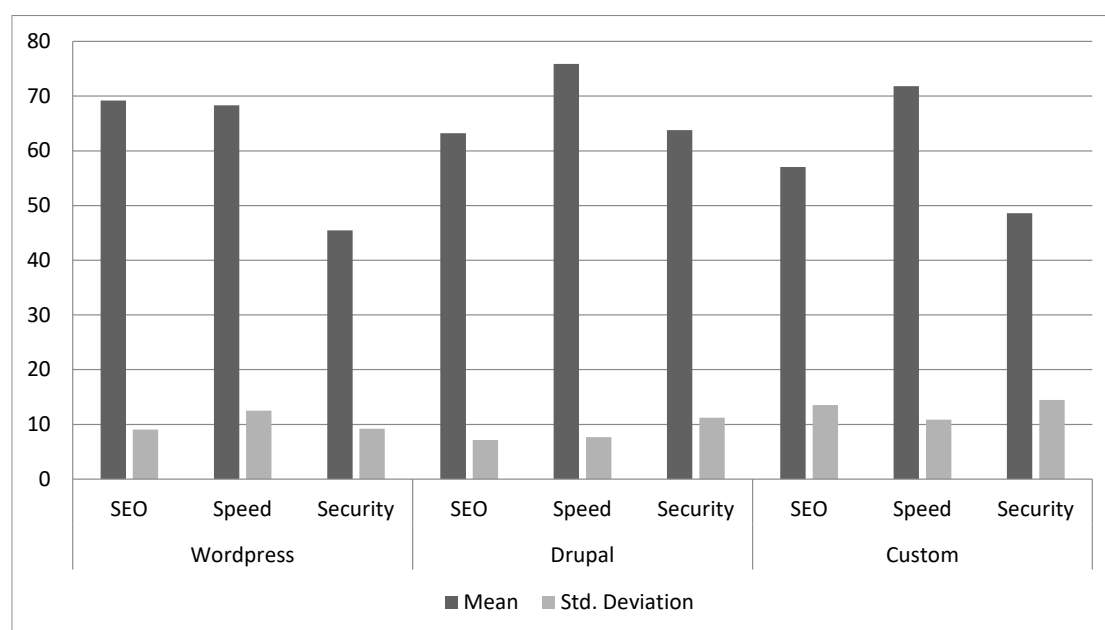
Στη συνέχεια συγκρίνονται τα δύο πιο γνωστά και χρησιμοποιούμενα συστήματα διαχείρισης περιεχομένου (CMS) και οι περιπτώσεις CMS που αναπτύχθηκαν από τους φορείς για ιδιοκτήτη χρήση. Συγκεκριμένα συγκρίνονται σε σχέση με τις μετρικές της κατηγορίας overall performance [το συνολικό SEO, τη συνολική ταχύτητα (Speed) και τη συνολική ασφάλεια (Security)] οι ιστότοποι που χρησιμοποιούν το Wordpress, το Drupal καθώς και ιδιόκτητες λύσεις κώδικα (Custom).

Ο μέσος όρος του συνολικού SEO για το Wordpress βρίσκεται στο 69,1% με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 9. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,97 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 0,684. Για τη συνολική ταχύτητα ο μέσος όρος είναι 68,3% με το βαθμό διακύμανσης των τιμών στο 12,4. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,909 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 0,039. Για τη συνολική ασφάλεια ο μέσος όρος είναι 45,4% με το βαθμό διακύμανσης των τιμών στο 9,1. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,838 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 0,002.

Ο μέσος όρος του συνολικού SEO για το Drupal βρίσκεται στο 63,2% με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 7,1. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,963 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 0,377. Για τη συνολική ταχύτητα ο μέσος όρος είναι 75,9% με το βαθμό διακύμανσης των τιμών στο 7,6. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,972 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 0,587. Για τη συνολική ασφάλεια ο μέσος όρος είναι 63,7% με

το βαθμό διακύμανσης των τιμών στο 11,1. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,979 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 0,797.

Ο μέσος όρος του συνολικού SEO για το Custom βρίσκεται στο 57% με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 13,5. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,969 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 0,483. Για τη συνολική ταχύτητα ο μέσος όρος είναι 71,8% με το βαθμό διακύμανσης των τιμών στο 10,8. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,967 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 0,433. Για τη συνολική ασφάλεια ο μέσος όρος είναι 48,5% με το βαθμό διακύμανσης των τιμών στο 14,4. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,943 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι 0,098.



Γράφημα 8. Μέσος όρος και απόκλιση για το τις παραμέτρους του overall performance – Σύγκριση μεταξύ κατηγοριών CMS

Συγκρίνοντας τα παραπάνω στοιχεία, το μεγαλύτερο ποσοστό SEO έχει το Wordpress. Αυτό σημαίνει ότι η δομή του συγκεκριμένου συστήματος είναι η πιο κατάλληλη για τους οργανισμούς πληροφόρησης που επιθυμούν βελτιστοποίηση στα αποτελέσματα των μηχανών αναζήτησης. Ακολουθεί έπειτα το Drupal και τελευταίο το Custom, γεγονός που υποδεικνύει ότι αρκετοί δημιουργοί ιστοσελίδων δεν δίνουν μεγάλη σημασία στα στοιχεία μεταδεδομένων και τη ιεραρχική δομή τους (Titles, H1 headings, Page descriptions). Σχετικά με τη συνολική ταχύτητα το μεγαλύτερο ποσοστό έχει το Drupal, μετά το Wordpress και τέλος το Custom. Η πιο σωστή επιλογή επομένως για την ιστοσελίδα ενός οργανισμού πληροφόρησης για ταχύτητα είναι με μικρή διαφορά το Drupal. Τέλος, για τη μέγιστη ασφάλεια καλύτερη επιλογή αποτελεί πάλι το Drupal με μεγάλη όμως διαφορά (63,7%). Ακολουθεί το Custom με 48,5% και το Wordpress με μόλις 45,4%, καθιστώντας τη λιγότερο επιθυμητή επιλογή.

Το μεγαλύτερο ποσοστό SEO έχει το Wordpress (69,1%). Αυτό σημαίνει ότι η δομή του συγκεκριμένου συστήματος είναι η πιο κατάλληλη για τους οργανισμούς

πληροφόρησης που επιθυμούν τη μέγιστη βελτιστοποίηση στα αποτελέσματα των μηχανών αναζήτησης. Ακολουθεί έπειτα το Drupal (63,2%) και τελευταίο το Custom (57%), γεγονός που υποδεικνύει ότι αρκετοί δημιουργοί ιστοσελίδων δεν δίνουν μεγάλη σημασία στα στοιχεία μεταδεδομένων και τη ιεραρχική δομή τους (τίτλοι, H1 επικεφαλίδες, πεδίο περιγραφής). Σχετικά με τη συνολική ταχύτητα το μεγαλύτερο ποσοστό έχει το Drupal (75,9%), μετά το Custom (71,8%) και έπειτα το Wordpress (68,3%). Η πιο σωστή επιλογή επομένως για την ιστοσελίδα ενός οργανισμού πληροφόρησης για ταχύτητα είναι με μικρή διαφορά το Drupal. Τέλος, για τη μέγιστη ασφάλεια καλύτερη επιλογή αποτελεί πάλι το Drupal με μεγάλη όμως διαφορά (63,7%). Ακολουθεί το Custom με 48,5% και το Wordpress με μόλις 45,4%, καθιστώντας τη λιγότερο επιθυμητή επιλογή. Πιθανόν να χρειάζεται παραμετροποίηση της επιλογής ορισμένων ιστοσελίδων οργανισμών με βάση τον σκοπό τους και τις ανάγκες του κοινού που εξυπηρετούν.

Πίνακας 5 - Στοιχεία της συνολικής απόδοσης & του μεγέθους των συστημάτων διαχείρισης περιεχομένου

CMS - Overall performance & size									
	Wordpress			Drupal			Custom		
	SEO	Speed	Security	SEO	Speed	Security	SEO	Speed	Security
Valid	23	23	23	30	30	30	31	31	31
Missing	0	0	0	0	0	0	0	0	0
Mean	69,174	68,348	45,478	63,233	75,9	63,767	57,032	71,806	48,581
Std. Deviation	9,049	12,492	9,194	7,162	7,671	11,19	13,524	10,83	14,44
Shapiro-Wilk	0,97	0,909	0,838	0,963	0,972	0,979	0,969	0,967	0,943
P-value of Shapiro-Wilk	0,684	0,039	0,002	0,377	0,587	0,797	0,483	0,433	0,098
Minimum	54	52	23	51	61	36	30	50	23
Maximum	88	89	71	83	97	87	79	92	82

4.5. Αποτελέσματα Behavioral Analytics

Ο μέσος όρος για τις σελίδες ανά επίσκεψη (Pages per visit) βρίσκεται στα 5,1 λεπτά για τις ξένες ιστοσελίδες με βαθμό διακύμανσης των τιμών (Std. Deviation) στο 5,5. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,568 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Για τις ελληνικές ιστοσελίδες ο μέσος όρος είναι 5 λεπτά με το βαθμό διακύμανσης των τιμών στο 2,8. Το τεστ κανονικότητας Shapiro – Wilk βρίσκεται στο 0,859 και η τιμή σημαντικότητας P-value of Shapiro – Wilk είναι <.001. Ο μέσος όρος

βρίσκεται για λίγο υπέρ των ξένων ιστοσελίδων σε ικανοποιητικά επίπεδα και τις δύο. Ο βαθμός διακύμανσης των τιμών είναι μεγαλύτερος για τις ελληνικές, αποδεικνύοντας για άλλη μία φορά τη μεγάλη κλίμακα τιμών και τη παρουσία “καλών” και “κακών” παραδειγμάτων προς μίμηση.

Πίνακας 6 - Στοιχεία της απόδοσης δεδομένων συμπεριφοράς χρηστών

Behavioral Analytics		
	Pages per visit	
	INT	GR
Valid	72	71
Missing	0	0
Mean	5,132	5,091
Std. Deviation	5,545	2,871
Shapiro-Wilk	0,568	0,859
P-value of Shapiro-Wilk	< .001	< .001
Minimum	1,38	1,26
Maximum	36,3	18,47

Κεφάλαιο 5. Ανακεφαλαίωση – Συμπεράσματα – Προτάσεις

Το αντικείμενο ενασχόλησης της εργασίας είναι η εξέταση των μεθόδων βελτιστοποίησης των ιστοσελίδων προκειμένου να αυξηθεί η δυνατότητα εύρεσης του ψηφιακού περιεχομένου μέσω των μηχανών αναζήτησης. Όπως έχει παρατηρηθεί, ολοένα και περισσότεροι χρήστες χρησιμοποιούν τις μηχανές αναζήτησης ως τη κύρια πηγή πληροφόρησης, αποκλείοντας το ενδεχόμενο αναζήτησης των ιστοσελίδων των πληροφοριακών οργανισμών. Σύμφωνα με τις βιβλιογραφικές πηγές που αναφέρθηκαν προηγουμένως, δεν υπάρχει κάποια συγκεκριμένη προσέγγιση του προβλήματος. Κρίνεται όμως απαραίτητο οι οργανισμοί πληροφόρησης να εξασφαλίσουν τη πρόσβαση στο περιεχόμενο τους μέσω της βελτιστοποίησης των ιστοσελίδων. Επίσης, είναι ζωτικής σημασίας η πληροφοριακή παιδεία του κοινού μέσω προγραμμάτων για την αναζήτηση και τον εντοπισμό έγκυρων πηγών στο διαδίκτυο. Η γνώση της διαδικασίας εξαγωγής των αποτελεσμάτων είναι σημαντική για την επιλογή των σωστών πηγών πληροφόρησης.

Έχει αναφερθεί σε προηγούμενη έρευνα ότι η πλειοψηφία των χρηστών των μηχανών αναζήτησης δεν βλέπουν τα αποτελέσματα μετά τις τρεις σελίδες (Neal, Campbell, Williams, Liu, & Nussbaumer, 2011). Γι' αυτόν ακριβώς το λόγο οι οργανισμοί πληροφόρησης θα πρέπει να εμφανίζονται στα πρώτα αποτελέσματα. Αυτός ο στόχος μπορεί να επιτευχθεί μόνο μέσω της σωστής εφαρμογής των μεθόδων βελτιστοποίησης ορισμένων τεχνικών χαρακτηριστικών. Παρακάτω παρατίθενται κάποιες μετρικές που μπορούν να βοηθήσουν, όπως και επιπλέον προτεινόμενες τεχνικές από τη βιβλιογραφία.

Η σχεδίαση των ιστοσελίδων μπορεί να επηρεάσει τη βελτιστοποίηση του περιεχομένου στις μηχανές αναζήτησης. Το μέγεθος και ο αριθμός των ανιχνευμένων σελίδων μπορούν να ενισχύσουν την αποτελεσματικότητα των μεθόδων βελτιστοποίησης (Dannay, 2011). Ο μεγάλος αριθμός σελίδων σε έναν ιστότοπο είναι ένδειξη ότι παρέχει συνήθως περισσότερες πληροφορίες και επομένως μπορεί να ευρετηριαστεί, να ανιχνευθεί και να του αποδοθεί μεγαλύτερη προτεραιότητα από τις μηχανές αναζήτησης. Οι σελίδες με μεγάλες διευθύνσεις (URLs) και παραμέτρους είναι δύσκολο να κατανοηθούν από τις μηχανές αναζήτησης και οι διαδικασίες της ανίχνευσης και ευρετηρίασης γίνονται δύσκολα. Επίσης, είναι δύσκολο να εξάγουν σημασιολογικές έννοιες, οι οποίες θα ενισχύσουν την εμφάνιση των κατάλληλων αποτελεσμάτων. Αυτός είναι ένας σημαντικός λόγος που το περιεχόμενο των ιστοσελίδων βιβλιοθηκών σπάνια κατατάσσεται στις πρώτες θέσεις. Επίσης, ένα μέρος του περιεχομένου βρίσκεται στο “βαθύ” αόρατο διαδίκτυο (Deer Web⁸). Τα αποτελέσματα της έρευνας των Onaifo και Rasmussen δείχνουν ότι υπάρχει σημαντική σχέση ανάμεσα στις ανιχνευμένες σελίδες (crawled pages) και στον αριθμό των προβολών των σελίδων καθημερινά (daily pageviews) (2013). Όσο πιο μεγάλος είναι ο αριθμός των ανιχνευμένων σελίδων από τις μηχανές αναζήτησης, τόσο πιο πιθανό είναι να εντοπίσουν οι χρήστες το επιθυμητό περιεχόμενο. Μία προτεινόμενη πρακτική για την αντιμετώπιση του προβλήματος είναι η σύνδεση των κρυμμένων σελίδων με εκείνες που είναι ορατές στις ιστοσελίδες. Οι χάρτες πλοήγησης επίσης μπορούν να κατευθύνουν τους χρήστες με την ιεραρχική δομή και να βοηθήσουν τις μηχανές στη διαδικασία της ευρετηρίασης. Η σωστή

⁸ Ως “Deer Web” αναφέρεται το περιεχόμενο που δεν ανήκει στον Επιφανειακό Ιστό, το οποίο ευρετηριάζεται από τις δημοφιλείς μηχανές αναζήτησης.

σχεδίαση επιτρέπει τη συνεχή εφαρμογή αποτελεσματικών μεθόδων και την αύξηση της εμπειρίας περιήγησης (user experience).

Ένας από τους πιο σημαντικούς παράγοντες στη βελτιστοποίηση είναι το ίδιο το περιεχόμενο των ιστοσελίδων. Σε αντίθεση με άλλες ιστοσελίδες, αυτές των οργανισμών πληροφόρησης δεν συναντούν κάποιο εμπόδιο στο συγκεκριμένο πεδίο. Η ποιότητα του περιεχομένου είναι ιδιαίτερα σημαντική για την αύξηση της επισκεψιμότητας μέσω των οργανικών αποτελεσμάτων και τη σύνδεση με άλλους ιστοτόπους αντίστοιχης επισκεψιμότητας. Οι ιστότοποι με μεγάλο πλήθος ιστοσελίδων, με συνεχώς ανανεωμένο περιεχόμενο, προσελκύουν περισσότερους χρήστες. Ο Dover στην εργασία του αναφέρει ότι ο αριθμός των επισκεπτών και των σελίδων περιήγησης επηρεάζει την κατάταξη (2011). Όσο πιο ποιοτικό και σχετικό είναι το περιεχόμενο των ιστοσελίδων με τις αναζητήσεις, τόσο πιο πολλές θα επιθυμούν να συνδεθούν μαζί τους. Αυτή η διαδικασία οδηγεί στην ενοποίηση των πληροφοριών του διαδικτύου και τη καλύτερη εξυπηρέτηση του κοινού. Οι οργανισμοί πληροφόρησης έχουν το σημαντικό πλεονέκτημα της κατοχής ποιοτικού περιεχομένου. Ωστόσο, η κατάταξη των ιστοσελίδων τους σε γενικές γραμμές δεν είναι ικανοποιητική συγκριτικά με τις δυνατότητες τους.

Σε αντίθεση με άλλες εργασίες που εξετάζουν το φαινόμενο της βελτιστοποίησης των ιστοσελίδων οργανισμών πληροφόρησης, στη συγκεκριμένη εργασία γίνεται ανάλυση των μετρικών. Μέσα από τη διαδικασία της στατιστικής ανάλυσης, παρουσιάστηκαν οι επιλεγμένες μετρικές σχετικά με τις κατάλληλες μεθόδους για την υψηλότερη δυνατή κατάταξη και τη καλύτερη εμπειρία χρήσης των ιστοσελίδων. Με βάση τα αποτελέσματα που αναφέρθηκαν υπάρχουν ορισμένα ευρήματα και προτάσεις για τη βελτιστοποίηση, τα οποία είναι τα εξής:

- Οι ιστοσελίδες οργανισμών πληροφόρησης της αλλοδαπής υπερیشύουν έναντι των ελληνικών στα ποσοστά των μετρικών με μικρή ή μεγάλη διαφορά
- Η ποιότητα του περιεχομένου των ιστοσελίδων, η αποφυγή χρήσης παραμέτρων στις διευθύνσεις, η χρήση αρχείων robots.txt και χαρτών πλοήγησης (sitemaps) καθορίζουν θετικά το ποσοστό βελτιστοποίησης των ιστοσελίδων στις μηχανές αναζήτησης και κατ' επέκταση τον βαθμό ανίχνευσης τους
- Οι ξένες ιστοσελίδες έχουν πιο πλούσιο περιεχόμενο (Avoid thin content pages) από τις ελληνικές ιστοσελίδες (75% - 18%)
- Στις ξένες ιστοσελίδες χρησιμοποιούνται λιγότερο οι παράμετροι στις διευθύνσεις (Avoid URL parameters) από τις ελληνικές με μικρότερη διασπορά τιμών (77,6% - 61,3%), γεγονός που ενισχύει τη θέση τους στα αποτελέσματα των μηχανών αναζήτησης
- Η χρήση αρχείων robots.txt (Use robots.txt files) είναι μεγαλύτερη για τις ξένες ιστοσελίδες (82,6% - 68,9%), ενώ η διασπορά των τιμών είναι μεγάλη για τις ξένες

και τις ελληνικές ιστοσελίδες (35,7 – 43,9) υποδεικνύοντας τη διαφορά διακύμανσης στους επιμέρους ιστοτόπους

- Η χρήση των χαρτών πλοήγησης (Set sitemap locations) είναι μικρή (23,8%- 8,4%) με μεγάλη διασπορά τιμών (41,2 – 28), γεγονός που δείχνει ότι υπάρχουν αρκετές ξένες και ελληνικές ιστοσελίδες που δεν έχουν εισάγει χάρτη πλοήγησης, μειώνοντας την ευρετηρίαση και κατ' επέκταση το συνολικό ποσοστό SEO
- Η χρήση του https στις διευθύνσεις (Use https) είναι μεγαλύτερη για τις ξένες ιστοσελίδες από τις ελληνικές (87% - 64,9%) και η διασπορά ιδιαίτερα μεγάλη και για τις δύο (32,8 – 47). Το ποσοστό της μεγάλης διακύμανσης των τιμών είναι αποτέλεσμα της χρήσης του στοιχείου μόνο από ορισμένες ιστοσελίδες του δείγματος της εργασίας
- Το συνολικό ποσοστό SEO (SEO) είναι μεγαλύτερο για τις ξένες ιστοσελίδες (66%- 55,4%), επιβεβαιώνοντας τα μεγαλύτερα ποσοστά στις παραπάνω μετρικές, οι οποίες χρησιμεύουν στη διεξαγωγή του συνολικού SEO
- Η δομή του Wordpress (CMS – Overall performance & size) αποτελεί τη πιο κατάλληλη επιλογή για τους οργανισμούς πληροφόρησης που επιθυμούν τη μέγιστη βελτιστοποίηση στα αποτελέσματα των μηχανών αναζήτησης (69%, 63% - Drupal, 57% - Custom)
- Η πιο ορθή επιλογή συστήματος διαχείρισης περιεχομένου για μέγιστη ταχύτητα είναι με μικρή διαφορά το Drupal (75,9%, 68,3% - Wordpress, 71,8% - Custom)
- Η καλύτερη επιλογή συστήματος διαχείρισης περιεχομένου για μέγιστη ασφάλεια αποτελεί πάλι το Drupal με μεγάλη όμως διαφορά συγκριτικά με τα υπόλοιπα συστήματα διαχείρισης περιεχομένου (63,7%, 45,4% - Wordpress, 48,5% - Custom)
- Προτείνεται η παραμετροποίηση της επιλογής συστημάτων διαχείρισης περιεχομένου για τους οργανισμούς πληροφόρησης με βάση τις ανάγκες του κοινού που εξυπηρετούν

Παρόλο που τα συμπεράσματα της εργασίας είναι χρήσιμα για τη κατανόηση της διαδικασίας της βελτιστοποίησης στο πλαίσιο των οργανισμών πληροφόρησης, υπάρχουν ορισμένοι περιορισμοί που πρέπει να ληφθούν υπόψη. Αρχικά, η συλλογή των διευθύνσεων δεν ήταν τυχαία, αλλά έγινε με βάση τη προέλευση των οργανισμών και κατανεμήθηκε ισόποσα. Το μέγεθος του δείγματος είναι μικρό και επομένως η εξαγωγή περαιτέρω συμπερασμάτων πρέπει να γίνει προσεκτικά. Η συλλογή των διευθύνσεων έγινε επίσης με κύριο γνώμονα τη κάλυψη του μεγαλύτερου εύρους οργανισμών, προβάλλοντας σημαντικά δεδομένα για επιπλέον έρευνα με βάση το είδος τους. Επίσης, η εργασία βασίζεται σε στατιστικά εργαλεία και εργαλεία συλλογής δεδομένων και ενδέχεται να μην είναι απόλυτα ακριβή τα στοιχεία, ειδικά λόγω της υψηλής μεταβλητότητας τους. Ένας

αξιοσημείωτος περιορισμός από τη χρήση του εργαλείου συλλογής και ανάλυσης των διευθύνσεων είναι η έλλειψη γνώσης εκείνων που δεν επιτρέπουν την ανίχνευση τους. Η διαδικασία λόγω αυτού περιορισμού έγινε πιο περίπλοκη και οδήγησε στην αναδιαμόρφωση των διευθύνσεων με την αφαίρεση των μη λειτουργικών.

Με βάση τα παραπάνω, η εργασία μπορεί να συνεχιστεί είτε με την εξέταση επιπλέον στοιχείων και την επίδραση τους υπό το πρίσμα του συνολικού ποσοστού του SEO, της ταχύτητας και της ασφάλειας. Οι μελλοντικές επεκτάσεις μπορούν να δώσουν έμφαση στη κατηγορία των οργανισμών πληροφόρησης και στα επιμέρους τεχνικά χαρακτηριστικά. Η χρήση επιπλέον διαφορετικών μεθόδων συλλογής δεδομένων, όπως είναι οι έρευνες ή οι συνεντεύξεις μπορούν να δώσουν μία σφαιρικότερη άποψη επί του θέματος. Αυτή η εργασία επέτρεψε την περαιτέρω κατανόηση της επίδρασης της βελτιστοποίησης στην ευρεσιμότητα των ιστοσελίδων οργανισμών πληροφόρησης με την ανάλυση των μετρικών.

Βιβλιογραφικές Αναφορές

Arlitsch, K., & O'Brien, P. S. (2012). Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google. *Library Hi Tech*, 30(1), 60–81. <https://doi.org/10.1108/07378831211213210>

Bawden, D., & Vilar, P. (2006). Digital libraries: To meet or manage user expectations. *Aslib Proceedings: New Information Perspectives*, 58(4), 346–354. <https://doi.org/10.1108/00012530610687713>

Cahill, K., & Chalut, R. (2009). Optimal results: What libraries need to know about google and search engine optimization. *Reference Librarian*, 50(3), 234–247. <https://doi.org/10.1080/02763870902961969>

Checkbot. (n.d.). Retrieved from <https://www.checkbot.io/>

Danny, D. (2011). *Search Engine Optimization (SEO) Secrets*. Retrieved from <https://dl.acm.org/doi/book/10.5555/2012012>

DeRidder. (2008). *Googlizing a Digital Library*. Retrieved from <https://journal.code4lib.org/articles/43>

Drivas, I. C., Sakas, D. P., Giannakopoulos, G. A., & Kyriaki-Manessi, D. (2020). Big Data Analytics for Search Engine Optimization. *Big Data and Cognitive Computing*, 4(2), 5. <https://doi.org/10.3390/bdcc4020005>

Erin E. Rushton, M. D. K. & M. A. S. (2008). *Searching for a New Way to Reach Patrons: A Search Engine Optimization Pilot Project at Binghamton University Libraries*. Retrieved from https://www.tandfonline.com/doi/pdf/10.1080/19322900802484248?casa_token=WbrYGdMcNGoAAAAA%3AXmaiJSfi2IsEp4JQirOCHI9Hy0vb_GkO93H1Cf1kfKpybDt3-sOEYEq_2qHCxSb2xwmgqdutYoA&

Grimmelmann, J. (2011). Some Skepticism About Search Neutrality. *The Next Digital Decade: Essays on the Future of the Internet*, 435–459. Retrieved from <http://ssrn.com/abstract=1742444>

Jasp. (n.d.). Retrieved from <https://jasp-stats.org/>

Lee, S., Jang, W., Lee, E., & Oh, S. G. (2016). Search engine optimization: A case study using the bibliographies of LG Science Land in Korea. *Library Hi Tech*, 34(2), 197–206. <https://doi.org/10.1108/LHT-02-2016-0014>

Malaga, R. A. (2008). Worst practices in search engine optimization. *Communications of the ACM*, 51(12), 147–150. <https://doi.org/10.1145/1409360.1409388>

McGee, M. (2010). *By the numbers twitter vs facebook vs google buzz*. Retrieved from <https://searchengineland.com/by-the-numbers-twitter-vs-facebook-vs-google-buzz-36709>

Neal, D. M., Campbell, A. J., Williams, L. Y., Liu, Y., & Nussbaumer, D. (2011). “I did not realize so many options are available”: Cognitive authority, emerging adults, and e-mental health. *Library and Information Science Research*, 33(1), 25–33. <https://doi.org/10.1016/j.lisr.2010.07.015>

Onaifo, D., & Rasmussen, D. (2013). Increasing libraries’ content findability on the web with search engine optimization. *Library Hi Tech*, 31(1), 87–108. <https://doi.org/10.1108/07378831311303958>

Purcell, K. (2011). Search and email still top the list of most popular online activities. *Search*, 6, 1–15. Retrieved from http://www.pewinternet.org/~media//Files/Reports/2011/PIP_Search-and-Email.pdf

Rushton, E., & Funke, S. (2011). The Goodness in the Evil of SEO. Why Search Engine Optimization Matters to Information Professionals. *Searcher*, 19(9), 30–35. Retrieved from <http://www.infotoday.com/searcher/nov11/index.shtml>

Scott, D. (2015). White Hat Search Engine Optimization (SEO): Structured Web Data for Libraries. *Partnership: The Canadian Journal of Library and Information Practice and Research*, 10(1), 1–21. <https://doi.org/10.21083/partnership.v10i1.3328>

SimilarWeb. (n.d.). Retrieved from <https://www.similarweb.com/>

whatcms.org. (n.d.). Retrieved from <https://whatcms.org/>

Yalçın, N., & Köse, U. (2010). What is search engine optimization: SEO? *Procedia - Social and Behavioral Sciences*, 9(July 2009), 487–493. <https://doi.org/10.1016/j.sbspro.2010.12.185>

Zhang, J., & Dimitroff, A. (2005). The impact of metadata implementation on webpage visibility in search engine results (part II). *Information Processing and Management*, 41(3), 691–715. <https://doi.org/10.1016/j.ipm.2003.12.002>

