

Τι είναι και πώς ρυθμίζω το robots.txt;

Alexandros Karagiannis - 2021-05-25 - Γενικά

Τι είναι το robots.txt;

Ένα αρχείο robots.txt είναι ένα αρχείο στο root του site σας, το οποίο υποδεικνύει εκείνα τα μέρη του site που δεν θέλετε να υπάρχει πρόσβαση σε ανιχνευτές μηχανών αναζήτησης. Το αρχείο χρησιμοποιεί το πρότυπο αποκλεισμού ρομπότ, το οποίο είναι ένα πρωτόκολλο με ένα μικρό σύνολο εντολών που μπορούν να χρησιμοποιηθούν για να υποδεικνύουν την πρόσβαση στο site σας κατά τομέα και συγκεκριμένα είδη ανιχνευτών ιστού (crawlers), όπως τα προγράμματα ανίχνευσης για κινητά και τα προγράμματα ανίχνευσης επιφάνειας εργασίας (mobile crawlers, desktop crawlers).

Τι χρησιμοποιείται για το robots.txt;

Μη αρχεία εικόνων

Για αρχεία μη εικόνας (δηλ. ιστοσελίδες) το robots.txt θα πρέπει να χρησιμοποιείται μόνο για τον έλεγχο της ανίχνευσης επισκεψιμότητας, συνήθως επειδή δεν επιθυμείτε ο server σας να κατακλύζεται από το πρόγραμμα ανίχνευσης της Google ή να σπαταλάει την ανίχνευση του προϋπολογισμού για ασήμαντες ή παρόμοιες σελίδες στο site σας.

Δεν πρέπει να χρησιμοποιείτε το robots.txt ως μέσο για την απόκρυψη των ιστοσελίδων σας από τα αποτελέσματα αναζήτησης Google. Αυτό οφείλεται στο γεγονός ότι άλλες σελίδες ενδέχεται να δείχνουν στη σελίδα σας και η σελίδα σας θα μπορούσε να έχει ευρετηριαστεί (να γίνει indexed) με αυτό τον τρόπο, αποφεύγοντας το αρχείο robots.txt. Εάν θέλετε να αποκλείσετε τη σελίδα σας από τα αποτελέσματα αναζήτησης, χρησιμοποιήστε μια άλλη μέθοδο, όπως προστασία με κωδικό πρόσβασης ή ετικέτες ή οδηγίες noindex.

Αρχεία εικόνας

Το robots.txt εμποδίζει την εμφάνιση των αρχείων εικόνας στα αποτελέσματα αναζήτησης Google. Ωστόσο, δεν εμποδίζει τη σύνδεση άλλων σελίδων ή χρηστών με την εικόνα σας.

Αρχεία πόρων

Μπορείτε να χρησιμοποιήσετε το αρχείο robots.txt για να αποκλείσετε αρχεία πόρων όπως ασήμαντες εικόνες, scripts ή style files, αν νομίζετε ότι οι σελίδες που φορτώνονται χωρίς αυτούς τους πόρους δεν θα επηρεαστούν σημαντικά από την απώλεια. Ωστόσο, εάν η απουσία αυτών των πόρων καθιστά δυσκολότερη την κατανόηση της σελίδας για το πρόγραμμα ανίχνευσης της Google, δεν πρέπει να τις αποκλείσετε ή αλλιώς η Google δεν θα κάνει καλή δουλειά ανάλυσης των σελίδων σας που εξαρτώνται από αυτούς τους πόρους.

Πώς ρυθμίζω το robots.txt;

Σε κάθε website υπάρχουν directories τα οποία δε θέλουμε να γίνονται indexed από τις μηχανές αναζήτησης, και κάποια άλλα στα οποία επιθυμούμε να επιτραπεί το indexing για λόγους SEO.

Ας δούμε παρακάτω κάποια παραδείγματα:

- Κάθε site περιέχει ευαίσθητα δεδομένα ή δεδομένα τα οποία δε προσφέρουν αξία όσον αφορά στο SEO αν γίνονται indexed. Κάποια από αυτά είναι τα: /cgi-bin/, /wp-admin/, /cart/, /scripts/ , /plugins/ κ.λπ.
- Δεν επιθυμούμε οι μηχανές αναζήτησης να κάνουν index directories με επανάληψη ίδιου περιεχομένου (duplicate content). Για παράδειγμα κάποια site μπορεί να έχουν print friendly version κάποιων σημείων τους (π.χ. σελίδων ή άρθρων) για να διευκολύνουν τους επισκέπτες να τα εκτυπώσουν. Αυτές οι σελίδες δεν θα πρέπει να γίνονται index παραπάνω από μία φορά.
- Περιηγηθείτε στο site σας και ψάξτε αν υπάρχουν συγκεκριμένα αρχεία που θα πρέπει να αποτρέψετε τις μηχανές αναζήτησης από το να τα κάνουν index π.χ: scripts, προσωπικά δεδομένα θόπως email ή τηλέφωνα επικοινωνίας) κ.λπ.

[Εδώ](#) μπορείτε να βρείτε περισσότερες πληροφορίες για το πώς μπορείτε να προσθέσετε κανόνες μέσα στο αρχείο.