

Accorgimenti tecnici per impedire l'indicizzazione nei motori di ricerca

(Per i responsabili tecnici dei siti nei quali vengono pubblicati i curricula)

Uno dei metodi per impedire l'indicizzazione dei curricula nel sito da parte degli spider dei comuni motori di ricerca è quello di utilizzare il file di testo robots.txt.

Il file robots.txt consente di non indicizzare una sola pagina o un'intera directory.

Occorre creare un file di testo e nominarlo robots.txt.

Questo file deve essere posizionato nella directory radice del sito web nel quale vengono pubblicati i file relativi ai curricula.

All'interno di questo file occorrerà inserire delle stringhe di testo con una opportuna sintassi.

Esempio del contenuto di un file robots.txt:

User-agent: *

Disallow: /file_da_non_indicizzare.htm

Disallow: /cartella_file_da_non_indicizzare/

Spiegazione dell'esempio

Il valore User-Agent contiene il nome del robot che si deve attenere alle restrizioni. Con il carattere * la regola viene applicata a qualsiasi spider dei più comuni motori di ricerca.

Il valore del campo Disallow contiene la pagina del sito o la directory che non devono essere indicizzate dagli spider dei motori di ricerca.