interweb

Generated on 15/11/2026 00:11

KNOWLEDGE BASE · PUBLIC DOCUMENT

robots.txt and sitemap.xml: what they do and how to check them

Google Search & Indexing

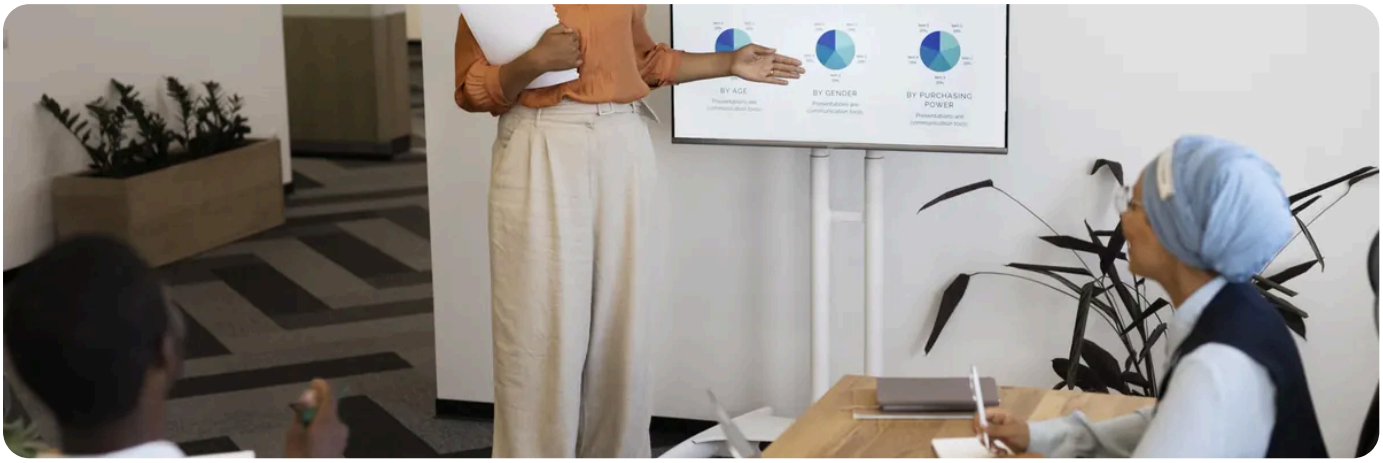
CODE	VERSION	COUNTRY · LANGUAGE
RLX-AO-KB-824	v2026.10.10	Angola · English

Interweb Angola
RLX Angola, Lda. · Group RLX Holdings
Tax ID (NIF) 500 260 5724 · Commercial registration no. 26464-25/250724

https://www.meu.interweb.ao/knowledgebase/824

robots.txt and sitemap.xml: what they do and how to check them

Google Search & Indexing Created on **October 10, 2026** 3 min read



They are two small files that speak to search engines, and it is very common for one of them to be wrong without anyone noticing. **robots.txt** tells robots **where they may go**. **sitemap.xml** hands them **the list of pages** you want them to know. Neither guarantees a page appears on Google, but a mistake in either can hide it.

One thing at a time

	robots.txt	sitemap.xml
What it is for	Asking a robot not to visit certain folders.	Listing the pages you want found.
Where it lives	At the root of the domain: `https://asuaempresa.ao/robots.txt`	Usually at the root too. robots.txt itself or Search Console tells robots where.
On WordPress	If there is no file, WordPress shows a virtual one.	WordPress ships its own, at `/wp-sitemap.xml`. An SEO plugin may create another.
What it does not do	It hides nothing: it is a public request, and only those who want to obey do.	It does not force Google to index what is listed.

How to check both

1. **Open robots.txt in the browser:** your domain followed by `/robots.txt`. It must open as plain text. A 404 means there is no file, so WordPress's virtual one applies, or nothing.

2. **Look for the line that ruins everything:** a lone `Disallow: /` after `User-agent: *` asks every robot to visit nothing at all. It is usually a leftover from when the site was being built.

3 **Check WordPress's own setting.** In Settings, Reading, there is an option asking search engines not to index the site. If it is ticked, the site is invisible. Untick it.

4 **Open the sitemap.** Try `/wp-sitemap.xml` or the address your SEO plugin gives. It must show a list of addresses with the right domain and `https`.

5 **Hand the sitemap to Google.** In Search Console, under Sitemaps, submit the address and see whether it is read without errors. The walkthrough is in **getting your site into Search Console**. Search Console is a Google product, used by you.

A reasonable robots.txt for a WordPress site

To create the file, open File Manager, go to the folder holding the site and create a file called `robots.txt` with this content:

```
User-agent: *
Disallow: /wp-admin/
Allow: /wp-admin/admin-ajax.php
Sitemap: https://asuaempresa.ao/wp-sitemap.xml
```

Swap the address for your own domain. A real file outranks WordPress's virtual one.

! **robots.txt is not for hiding pages.**
It is public: anyone can read it, and a list of "forbidden" folders is a map for people hunting secrets. And asking a robot not to visit a page does not remove it from Google if other pages link to it. To keep a page out of search results use the `noindex` tag, and let the robot visit the page so it can see the tag. For anything truly private, use a password.

Each subdomain has its own robots.txt.
The one at `https://asuaempresa.ao/robots.txt` does not cover a subdomain of yours. And if you do not want AI robots reading your site, this is also where you ask: see **AI bots crawling your site**.

NEED A HAND?

Does robots.txt or the sitemap fail on the server side (404, 403 or 500)? Tell us the address and what appears.

Open a support ticket →

SEE ALSO

- **WordPress hosting**
- **SSL certificates**
- **Support Policy**

interweb

Interweb Angola

RLX Angola, Lda. · Group RLX Holdings

TAX ID (NIF)	500 260 5724
COMMERCIAL REGISTRATION NO.	26464-25/250724
REGISTERED OFFICE	Rua Direita do Patriota, Bairro Dangereux Belas, Luanda, Angola
CONTACT	info@interweb.ao · +244 942 011 191 · www.interweb.ao
OPENING HOURS	Monday to Friday - 08 AM to 04 PM Saturday & Sunday - Closed

Online version of this document

RLX-AO-KB-824 · v2026.10.10 · Generated on 15/11/2026 00:11