Google Publishes New Robots.txt Explainer via @sejournal, @martinibuster

Google published a new Robots.txt refresher explaining how Robots.txt enables publishers and SEOs to control search engine crawlers and other bots (that obey Robots.txt). The documentation includes examples of blocking specific pages (like shopping carts), restricting certain bots, and managing crawling behavior with simple rules.

From Basics To Advanced

The new documentation offers a quick introduction to what Robots.txt is and gradually progresses to increasingly advanced coverage of what publishers and SEOs can do with Robots.txt and how it benefits them.

The main point of the first part of the document is to introduce robots.txt as a stable web protocol with a 30 year history that's widely supported by search engines and other crawlers.

Google Search Console will report a 404 error message if the Robots.txt is missing. It's okay for that to happen but if it bugs you to see that in the GSC you can wait 30 days and the warning will drop off. An alterative is to create a blank Robots.txt file which is also acceptable by Google.

Google's new documentation explains:

> *"You can leave your robots.txt file empty (or not have one at all) if your whole site may be crawled, or you can add rules*

to manage crawling.”

From there it covers the basics like custom rules for restricting specific pages or sections.

The advanced uses of Robots.txt covers these capabilities:

- Can target specific crawlers with different rules.
- Enables blocking URL patterns like PDFs or search pages.
- Enables granular control over specific bots.
- Supports comments for internal documentation.

The new documentation finishes by describing how simple it is to edit the Robots.txt file (it’s a text file with simple rules), so all you need is a simple text editor. Many content management systems have a way to edit it and there are tools available for testing if the Robots.txt file is using the correct syntax.

Read the new documentation here:

Robots Refresher: robots.txt — a flexible way to control how machines explore your website

Featured Image by Shutterstock/bluestork

https://www.searchenginejournal.com/google-publishes-new-robots-txt-explainer/541921/