

Duke Office of Information Technology

Google Search

The Office of Information Technology (OIT) has licensed the Google Search Appliance as an enterprise search solution for all of Duke. The appliance is managed by OIT's Internet Framework Services group, and it currently crawls across a wide number of domains and websites. Read on for more information about the using the appliance, as well as helpful hints and instructions on getting the most out of the appliance and your websites.

- ? [Basic information about the appliance](#)
 - ? [The Duke appliance "universe"](#)
 - ? [How the Google Search Appliance crawls](#)
 - ? [Where can I find out more?](#)
-

Basic information about the appliance

The Google Search Appliance uses the same technology found on Google.com, so using the Duke appliance to search Duke websites feels similar to using the Google website. Users also should expect the same quality and search accuracy found at Google. The primary difference between the two is that the Duke appliance only "crawls" and indexes websites that fall inside the duke.edu domain (more or less, see below).

[Return to Top](#)

The Duke appliance "universe"

The Duke appliance has three important parameters that define what websites it will and won't crawl:

- ? a list of URLs from which to start crawling
- ? a list of domains to which it must restrict its crawling
- ? a list of URL patterns that it shouldn't crawl

The most salient of these lists is the list of domains to be crawled, all of which are considered to be strongly related to Duke University:

- ? duke.edu
- ? goduke.com
- ? dukehealth.org
- ? dukece.com
- ? dukealumni.com

If the domain name of a website contains one of the entries from this list (e.g. www.aas.duke.edu), then it is very likely it is getting indexed by the appliance. If the domain name of a website does not match one of the entries from the list above (e.g. www.dukenursing.org), then it won't be included in the index.

This "universe" of Web pages is crawled continuously by the Duke appliance, which means a new site often will be searchable within a day of its go-live date (assuming it's been linked to, see below).

[Return to Top](#)

How the Google Search Appliance crawls

How does the Google appliance find and crawl your site? While Google keeps the magic of its search accuracy a secret, they have made no secret about how the engine crawls. As the folks at Google so eloquently phrase it:

The appliance can only find URLs by following links The crawler can follow normal HTML links and links embedded in Flash content, MS Word documents and PDF files. It cannot follow links embedded in Javascript code and it cannot submit HTML forms. In other words, if you want your site to be crawled and indexed by the appliance, make sure it's getting linked to by a page or site that has already been indexed, and make sure your site provides adequate HTML-based links for the appliance to follow. If you are finding that your site isn't getting crawled, follow these guidelines:

- ? Make sure the site is hosted in a Web domain that is included in the list of domains above. If you suspect that your site is being excluded, please contact the OIT Service Desk (see contact information at right).
- ? Make sure you haven't added a robots.txt file or meta tags that prevent the Google robot from indexing your page. Make sure your site is linked from somewhere in the domains listed above.

The Duke Google Search Appliance automatically indexes all pages in the valid Duke domains. There is no need to submit your pages to be indexed as long as your site is within the duke.edu domain.

Please note that if no pages in the list of valid Duke domains link to your site, your pages will not be indexed.

The Google crawler includes changed, newly created, or newly removed pages automatically during its crawls.

For additional information about making your site Google-friendly, please consult Google's "publishing best practices."

[Return to Top](#)

Where can I find out more?

[Creating a Google appliance search form](#)

[How to get your site indexed \(or Not\)](#)

[How to manage and customize your search results \(using frontends\)](#)

[Tips from Google on making your site search-engine friendly](#)

[Return to Top](#)

Duke Office of Information Technology - www.oit.duke.edu - (919) 684-2200 - help@oit.duke.edu