Google Tips

URL Structure

A site's URL structure should be as simple as possible. Consider organizing your content so that URLs are constructed logically and in a manner that is most intelligible to humans (when possible, readable words rather than long ID numbers). For example, if you're searching for information about aviation, a URL like http://en.wikipedia.org/wiki/Aviation will help you decide whether to click that link. A URL like http://www.example.com/index.php?id_sezione=360&sid=3a5ebc944f41daa6f849f730f1, is much less appealing to users.

Consider using punctuation in your URLs. The URL http://www.example.com/green-dress.html is much more useful to us thanhttp://www.example.com/greendress.html. We recommend that you use hyphens (-) instead of underscores (_) in your URLs.

Overly complex URLs, especially those containing multiple parameters, can cause a problems for crawlers by creating unnecessarily high numbers of URLs that point to identical or similar content on your site. As a result, Googlebot may consume much more bandwidth than necessary, or may be unable to completely index all the content on your site.

Common causes of this problem

Unnecessarily high numbers of URLs can be caused by a number of issues. These include:

Additive filtering of a set of items Many sites provide different views of the same set of items or search results, often allowing the user to filter this set using defined criteria (for example: show me hotels on the beach). When filters can be combined in a additive manner (for example: hotels on the beach and with a fitness center), the number of URLs (views of data) in the sites explodes. Creating a large number of slightly different lists of hotels is redundant, because Googlebot needs to see only a small number of lists from which it can reach the page for each hotel. For example:

Hotel properties at "value rates":http://www.example.com/hotel-search-results.jsp?Ne=292&N=461

Hotel properties at "value rates" on the beach:http://www.example.com/hotel-search-results.jsp?Ne=292&N=461+4294967240

Hotel properties at "value rates" on the beach and with a fitness center:http://www.example.com/hotel-search-results.jsp?Ne=292&N=461+4294967240+4294967270

Dynamic generation of documents. This can result in small changes because of counters, timestamps, or advertisements.

Problematic parameters in the URL. Session IDs, for example, can create massive amounts of duplication and a greater number of URLs.

Sorting parameters. Some large shopping sites provide multiple ways to sort the same items, resulting in a much greater number of URLs. For example:http://www.example.com/results?search_type=search_videos&search_query=tpb&search_sort=relevance

&search_category=25

Irrelevant parameters in the URL, such as referral parameters. For example:http://www.example.com/search/noheaders?click=6EE2BF1AF6A3D705D5561B7C3564D9C2&clickPage=

OPD+Product+Page&cat=79
http://www.example.com/discuss/showthread.php?referrerid=249406&threadid=535913
http://www.example.com/products/products.asp?N=200063&Ne=500955&ref=foo%2Cbar&Cn=Accessories.

Calendar issues. A dynamically generated calendar might generate links to future and previous dates with no restrictions on start of end dates. For example:http://www.example.com/calendar.php?d=13&m=8&y=2011
http://www.example.com/calendar/cgi?2008&month=jan

Broken relative links. Broken relative links can often cause infinite spaces. Frequently, this problem arises because of repeated path elements. For example:http://www.example.com/index.shtml/discuss/category/school/061121/html/interview/

  category/health/070223/html/category/business/070302/html/category/community/070413/html/FAQ.htm

Steps to resolve this problem

To avoid potential problems with URL structure, we recommend the following:

Consider using a robots.txt file to block Googlebot's access to problematic URLs. Typically, you should consider blocking dynamic URLs, such as URLs that generate search results, or URLs that can create infinite spaces, such as calendars. Using regular expressions in your robots.txt file can allow you to easily block large numbers of URLs.

Wherever possible, avoid the use of session IDs in URLs. Consider using cookies instead. Check our Webmaster Guidelines for additional information.

Whenever possible, shorten URLs by trimming unnecessary parameters.

If your site has an infinite calendar, add a nofollow attribute to links to dynamically created future calendar pages.

Check your site for broken relative links.