

# CORE

---

**White paper:  
Faceted search  
and SEO**

We at Core work with a range of companies who come to us to advise them and manage their search and social requirements.

Dr Jess Spate, our Head of Search and Social, has written this white paper to share some of her own superlative knowledge of search engine optimisation with you, and to show how a carefully thought out and structured approach to designing your website can optimise your business and sales, and keep customers coming back. Examples below have been taken from our SEO work with Graham & Brown whose new site will be launching mid-November 2013.

We hope you find this white paper useful - get in touch with us if we can help you with any aspect of search optimisation.

Our full contact details are available at the end of this document.

## Faceted search and SEO

You have a website. You want to sell things on this website. You do some keyword research, make sure the titles and tags are optimised, add some helpful text content. So far, so good. You also want the page to be user friendly - the right product has to be easy to find. This is really important if you've got a lot of different options.

Forget plain old pagination - sometimes people will wade through pages and pages of products they have no interest in but often they won't.

## What is this faceted search thing anyway?

The thoughtful web designer will introduce faceted search. This lets the user select a subset of products meeting particular requirements. In the example to the right, a shopper can narrow down Graham & Brown's wallpaper collection (1065 different patterns at the time of writing) by colour, by design style, or by designer.

Some of those categories (Kelly Hoppen Wallpaper, Green Wallpaper, etc) have static landing pages of their own. But when we step down another level and select for **designer = Kelly Hoppen AND colour = Green** it's a different story. Then there's **designer = Kelly Hoppen AND (colour = Green OR colour = Blue)** and many, many other possible facet combinations.

Allowing the user to see only the products that meet their personal needs is a sensible thing to do. They might also want to re-order the products, for example by price or popularity. But this does mean that you'll have a lot of different page possibilities.

And you're going to have duplicate content. **Colour = Green AND designer = Kelly Hoppen** might be effectively the same as **designer = Kelly Hoppen AND colour = Green**, but unless you're careful there are still going to be two different URLs involved.

Colour:



---

Design Style:

[Floral Wallpaper \(286\)](#)

[Geometric Wallpaper \(230\)](#)

[Plain/Textured Wallpaper \(180\)](#)

[Motif Wallpaper \(121\)](#)

[more...](#)

---

Designer:

[Steve Leung \(35\)](#)

[Kelly Hoppen \(34\)](#)

[Amy Butler \(33\)](#)

[Julien MacDonald \(26\)](#)

[more...](#)

## Fun with combinatorics

In this example we have 20 colours, and 4 designers. We'll ignore the wallpaper styles for the moment. That means that there are 24 facets the searcher could select. In pairs, these might be **colour = Green AND colour = Blue**, or **designer = Kelly Hoppen AND colour = Green**. Let's also assume that we're smart and our system enforces an order, so if a user searches on **colour = Green AND colour = Blue** they get the results for **colour = Blue AND colour = Green** instead of a whole new URL.

We're allowing **colour = Green AND colour = Blue** because data tells us that colour combinations are surprisingly popular search terms.

The formula for counting possible combinations without repetition (no **colour = Green AND colour = Green**) is  $P = \frac{n!}{r!(n-r)!}$  where  $n$  is the number of variables and  $r$  the number of facets specified. When  $n = 24$  and  $r = 2$ , this comes out at 276 potential faceted search result pages. If the user selects 3 attributes,  $r = 3$  and  $P = 2024$ .

Now let's say we also allow those 4 wallpaper styles to be included in the process. Now  $n = 28$ . If  $r = 2$ ,  $P = 378$ . If we bump  $r$  up to 3,  $P = 3276$ . That's a lot of extra pages.

If you want to experiment with your own numbers and don't feel like calculating factorials by hand - a niche hobby at best - there are plenty of free [combinatorics calculators](#) online.

## Faceted search and crawlers - the bad part

Yes, those 3276 pages would have dynamic URLs. This does not mean that search engines will ignore their existence. Like the intrepid little fellows that they are, they explore. And when they get caught up in an abundant supply of complicated faceted results, this means less attention and less love for your more important pages.

Crawl bandwidth is not in unlimited supply. Tagging these pages 'noindex, nofollow' doesn't actually stop the critters in their tracks either. They still wander about, peering aimlessly at your pages even if they don't write home about it.

When you really want to control where bots do and do not spend their energy, the next step up from noindex, nofollow is robots.txt. You could explicitly exclude faceted search results.

User-agent: \*

Disallow: \*/fsearch/\*

would suggest in rather stronger terms that no bot sticks its nose into your faceted search business (ie, URLs that include /fsearch/).

---

## White paper: Faceted search and SEO

Another solution to this knotty problem would be to stick all your filters within a piece of JavaScript, keeping the URL the same, or use AJAX to handle this kind of user activity. Crawlers won't see anything that will cause difficulties and there are no duplicate content issues. This is a good solution for re-ordering in particular.

## The middle path

So we want some of the stronger faceted searches included in the index. We can do this by adding a URL parameter that appears when too many attributes are combined, or when the user filters by price, popularity, size, or some other attribute that isn't going to be useful in search terms, and exclude just those pages through the robots.txt file.

User-agent: \*

Disallow: /\*toomanyfilterstag/\*

Now the bots will ignore a lot of the faceted search junk but still crawl the combinations that are likely to be larger-volume and more common. Cool. This method is widely recommended.

## What about authority/linkjuice/PR and all that?

Disallowing crawlers comes with one disadvantage. It can be quite a biggie.

Pulling back a moment, we all know that authority (or linkjuice, or PageRank, or whatever you want to call it) flows through internal and external links. Let's say a user links to a faceted search page, for example to our old friend Kelly Hoppen green wallpapers. The page might have a cumbersome URL (more about that later) but that's often a limited concern in the world of cut and paste.

If the page has been crowbarred out of the index, then the little burst of goodness brought by the link won't get anywhere. So value could be lost.

The solution is the rel=canonical tag, usually used to control duplicated content. Placed in the head section of that designer = Kelly Hoppen AND colour = Green page, it would look like

```
<link rel="canonical" href="http://www.grahambrown.com/us/designer-wallpapers/Kelly-Hoppen"/>
```

and tell crawlers that the page is a close variant of the base Kelly Hoppen page (that of the first specified category) or even the base Wallpaper category.

It can be slightly more difficult to implement than just removing most faceted search result pages from the index, and certainly more complex than running the whole lot through AJAX under a single URL, but pages tagged as non-canonical duplicates do pass PageRank.

## Getting organic value from faceted search results

Next problem. The pages in question are usually SEO-unfriendly to the point of being downright antisocial. The URLs are usually long and not easily interpretable. There is no unique content and no optimised titles or tags. Our pages might rank but they probably won't where there is much competition.

Options now are: A) to provide dedicated landing pages for popular facet combinations, perform on-page optimisation in the usual way, and create careful navigation to lead viewers down the right paths, or B), to do the minimum via automated means.

Packages are available that will transform faceted search URLs and title tags into something friendlier. Manadev's [Layered Navigation Plus](#) is one ready-made example. No doubt clever code ninjas can do similar things on their own.

## Use the lovely data

Building a new site from the ground up gives you the flexibility to do it right the first time, but on the other hand (and assuming you've got halfway decent analytics in place) an existing site means a rich and tasty bucket of lovely data. Mmmm!

Check both internal and external search traffic to find out what the key combinations are before deciding how deep into facet territory you want spiders to go, and what terms really deserve landing pages of their own.

New sites can do a similar thing, albeit less deeply in most long-tail cases, with keyword research. Test and check volumes on possible key attribution combinations before building a faceted search system and editing robots.txt.

# Getting organic value from faceted search results

The simplest approach, suitable for sites where a smaller number of pages might be valuable or those where every ounce of authority must count, is as follows:

- 1.1 Identify potentially useful attribute combinations through keyword research and analysis.
- 1.2 Build optimised landing pages for matching products and insert into navigation.
- 1.3A Handle further faceted search with AJAX (for example) without generating additional pages or internal links.

Or

- 1.3B Use rel=canonical to tag all faceted search pages.

A more complex approach, suitable for those occasions when you still want to index quite a few faceted search result pages and cannot practically create static landing pages, or for those lucky few with a big bucket of authority to play with, might look like this:

- 2.1 Set rel=canonical tags on faceted search pages that:
  - A. Involve more than a set number of variables
  - B. Involve attributes like price, popularity, or size, which are unlikely to add value at the organic search level.
  - C. Are duplicated (eg if designer = Kelly Hoppen AND colour = Green is to be indexed, colour = Green AND designer = Kelly Hoppen does not need to be).
  - D. Involve two or more references to certain attributes (eg we might want to include colour = Green OR colour = Blue but not searches that include two different designers).
- 2.2 Ensure URLs and titles on indexable pages produced by faceted search meet at least the most basic on-page SEO standards.

Web. Apps. Mobile. Social networking.  
Digital media is changing the world.  
We can get it working for you too.

Can we help promote your business using search engine optimisation and digital marketing?

If so, get in touch here:

## Contact us

---

Tel: +44 (0)29 2070 4700

Email: [hello@wearecore.co.uk](mailto:hello@wearecore.co.uk)

Web: <http://wearecore.co.uk>

Twitter: [@weare\\_core](https://twitter.com/weare_core)

LinkedIn: Core - Digital Everything