Technical Issues

SEO Technical Modules

Contents Table

[Technical Issues](file:///C:\Users\Macfarlanei\Desktop\NEW%20TEMPLATES\ISSUES%20UPDATE.docx#_Toc317519569)

[Contents Table](#_Toc317519570)

[1.1. Parameter Handling](#_Toc317519571)

[1.2. Unnecessary URL Parameter Issue](#_Toc317519572)

[1.3. Parameter Tracking Issue](#_Toc317519573)

[1.4. ID Parameter Issue](#_Toc317519574)

[1.5. Session Issue](#_Toc317519575)

[1.6. Pagination Issue](#_Toc317519576)

[1.7. Optimising Facets](#_Toc317519577)

[1.8. Flash Content Issue](#_Toc317519578)

[1.9. Internal Use Of rel="nofollow" Issue](#_Toc317519579)

[1.10. JavaScript Navigation Issue](#_Toc317519580)

[1.11. Form Field Navigation Issue](#_Toc317519581)

[1.12. Public-Facing Site Map Issue](#_Toc317519582)

[1.13. Server Error Issue](#_Toc317519583)

[1.14. Restricted Spider Access Issue](#_Toc317519584)

[1.15. Frames Issue](#_Toc317519585)

[1.16. Root Index Redirect Issue](#_Toc317519586)

[1.17. Robots Exclusion Protocol Issue](#_Toc317519587)

[1.18. Placeholder Content Issue](#_Toc317519588)

[1.19. 404 Error Handling Issue](#_Toc317519589)

[1.20. Directory Index Error Page Issue](#_Toc317519590)

[1.21. Branded 404 Page Issue](#_Toc317519591)

[1.22. Low Quality Content Issue](#_Toc317519592)

[1.23. Printer-Friendly Duplicate Content Issue](#_Toc317519593)

[1.24. Staging Sub-domain Issue](#_Toc317519594)

[1.25. HTTP Compression Issue](#_Toc317519595)

[1.26. Compound Words & Underscore Issue](#_Toc317519596)

[1.27. Alt Attributes Issue](#_Toc317519597)

[1.28. Trailing Slash Canonicalisation Issue](#_Toc317519598)

[1.29. Search-Generated Content Issue](#_Toc317519599)

[1.30. Non-WWW Domain Canonicalisation Issue](#_Toc317519600)

[1.31. Robots Exclusion Issue](#_Toc317519601)

[1.32. Secure Server Issue](#_Toc317519602)

[1.33. Default Index Page Duplicate Content Issue](#_Toc317519603)

[1.34. Canonical Homepage Issue](#_Toc317519604)

[1.35. Root index page redirects to internal page](#_Toc317519605)

[1.36. Canonical Domain Issue](#_Toc317519606)

[1.37. URL Language Issue](#_Toc317519607)

[1.38. Cookie Based Multiple Language Issue](#_Toc317519608)

[1.39. Multilingual Content Markup](#_Toc317519609)

[1.40. Internal Redirect Issue](#_Toc317519610)

[1.41. URL Case Sensitivity Issue](#_Toc317519611)

[1.42. XML Sitemaps](#_Toc317519612)

[1.43. Image XML Sitemaps](#_Toc317519613)

[1.44. Breadcrumb Trails Issue](#_Toc317519614)

[1.45. Structured Markup](#_Toc317519615)

[1.46. Video Structured Markup](#_Toc317519616)

[1.47. Video Feeds](#_Toc317519617)

[1.48. Non- hierarchical URL structure](#_Toc317519618)

[1.49. URL Duplicate Content and Brand Threat Issue](#_Toc317519619)

[1.50. Broken Links Issue](#_Toc317519620)

[1.51. Pages Missing Analytics Tags](#_Toc317519621)

[1.52. RSS Feeds Issue](#_Toc317519622)

[1.53. Content “Below The Fold” Issue](#_Toc317519623)

Parameter Handling

Parameters are strings of additional information which may be included on a URL after the question mark character. Common reasons for this are session IDs, referral tracking and changes to the way in which page content is displayed (for example, sorting a list of products by price). Most forms of paid search, display advertising etc. use parameters to track clicks and conversions.

#### Example

http://www.example.com/index.php?affiliate\_id=12345

Parameter

Established URL

Some search engines will allow the site owner to set the search engine to ignore specific URL parameters. With the example URL above, you might notify the search engine to ignore the parameter ‘affiliate\_id’.

This can be very useful for:

* Canonicalisation – if parameters create multiple identical or similar versions of an established URL, you can ensure that only one version is crawled, thus avoiding unwanted duplication and freeing up more resource to crawl elsewhere.
* Quality Control – preventing such URLs from being indexed will help to improve the quality of ##clientname##’s URLs returned within the search results, for example, by making them more readable.
* PageRank Distribution – if links are made to different versions of the same URL, this can split the PageRank distribution from those links between the different variants.

### Solution

After crawling the ##clientname## website and analysing search engine webmaster tools, we found the following parameters which could be excluded from search results:

##,##,##

On request, we will notify the search engines how to handle the parameters used on the site. Whilst parameter handling is a useful tool, search engines emphasise that they only use the settings as a hint when crawling.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

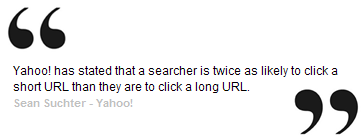
Unnecessary URL Parameter Issue

We found that some URLs on the ##clientname## website include non-essential elements, making URLs longer than may be required.

#### Example

##

In comparison with long URLs, short URLs generally receive higher click-through rates (CTR), are easier remember and are simpler to type into a browser’s address field.



### Solution

Remove unnecessary elements from URL strings. For example, we can remove ##, leaving us with:

##

Once complete, as with all URL changes it is vital to implement a 301 permanent redirect from the old URL to the new one.

### Alternative Solution

If this is not possible, a canonical link element should be placed in the <head> of the old URL referencing the desired URL form. This will help to ensure that existing value is passed to the new version.

#### Example

<link rel="canonical" href="##" />

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Parameter Tracking Issue

It appears that the ##clientname## site appends tracking parameters to URLs in order to track user interaction and, potentially, marketing campaigns.

#### Example

##example##

The tracking parameter “##” is appended to the end of the URL. Such parameters result in content duplication which misuse search engine crawling resources. Additionally, because the original URL isn’t always referenced, link value is spread across more than one location, which negatively impacts page performance.

site:## inurl:##

In Goole’s search results, we were able to determine that there are ## URLs with the term “##” indexed.

### Solution

If possible, replace tracking via URL query parameters (after “?” and separated with “&”) with fragment identifiers (after “#”). For example:

<a href="##" />

Search engines strip out any characters after the anchor (“#”) in the URL, which means that they will see all URLs without the tracking information.

### Alternative Solution

Use plain text links without tracking tags applied, and then use a JavaScript onClick event handler to automatically add the tracking parameters back to links when clicked upon by users. For example:

<a href="##" onClick="##">

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

ID Parameter Issue

The ##clientname## website make use of the ## parameter for management of some of its URLs. The ## parameter, along with SID= is commonly used for managing sessions.

##clienturl##

Some search engines are known to treat websites using either of these parameters as using session IDs, and as a result, the website may not be fully crawled. Reduced crawling will impact the opportunity web pages have to rank in the search results.

Though Google changed its Webmaster Guidelines in October 2006 to state that is it now acceptable to use ID= as a URL parameter, problems associated with ID= parameters in URLs may still persist in other search engines. Consequently, Google’s stance on such parameters should not affect the way in which site structures are designed.

### Solution

In order to resolve this issue, ##clientname## could look to replace the ## parameter with a different parameter, e.g. ##. However, dynamic URLs are less desirable than static URLs, so the ideal solution could be to change the way in which URLs are formed throughout the entire website. Resources required to achieve this will likely be significant.

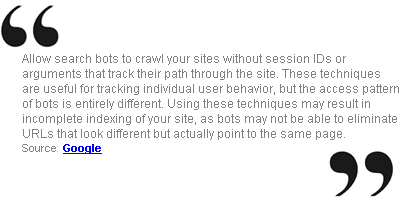
As with any URL change, it is important to ensure that old URLs are 301 permanently redirected to their new location.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Session Issue

Currently, ##clientname## presents the user session within the URL, which is likely negatively impacting search engine crawling and performance.



As a search engine spider crawls the ##clientname## website they create a session. This session is then appended to the URL.

#### Example

##clienturl##?sessid=121fadf324gsdfg435gs

Due to the session being appended to the URL, a search engine spider cannot is not presented with the same URLs on each visit as it changes with each new session that is generated.

History plays an important role in search engine algorithms. URLs which have existed for a long period of time tend to rank better, as a search engine gets to know and trust that the URL will not change.

### Solution

While search engine support for sessions is improving, they will continue to prevent the website from reaching its full potential. For this reason, we strongly suggest that you do not include sessions within URLs.

The integrity of the session data can be maintained by placing the session into a cookie and then tracking the user via this alternative method. Sessions tracked via cookies are generally as effective as sessions in URLs, with the exception that any User-Agent that does not support cookies will not be tracked.

We recommend that User-Agents which do not support cookies (such as search engine spiders) should be allowed to browse the website without tracking tags being used.

We recommend ensuring that this tracking is implemented in such a way as to be compatible with the EU’s e-Privacy Directive.

### Alternative Solution

If there is no ability to drop sessions and use cookies for tracking, we suggest that ##clientname## makes use of spider control methods to try to influence search engine spiders to crawl the URL minus the session.

* Sitemap

Create a Sitemap which has a list of all ##clientname## URLs, excluding any session data. This can then be submitted to all the major search engines. A Sitemap is typically the first point of call when undertaking the crawling process, so by providing URLs it may enhance how a search engine crawls the website.

* Parameter Handling

Within Google Webmaster Tools, one can assign parameters to be excluded from the crawling cycle. Whilst this approach may work to some extent, it is only a plaster to a larger problem.

* Canonical Link Element

Place a canonical link element on every web page referencing the absolute URL minus the session.

<link rel="canonical" href="http://www.example.com/" />

When supported search engines of the canonical link element crawl the website, they will concatenate all of the canonical versions into one value ready for promotion. Similar to parameter handling, the canonical link element will likely only work as a plaster.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Pagination Issue

Pagination is typically used in cases where users will expect items on the page to move over time, for example: search results; dynamically generated product lists; or news and press release entries.

On the ##clientname## website, it is used ##DESCRIPTION OF USES HERE##.

## EXAMPLES ##

The issue with using pagination in this way is that it produces low quality pages which are then crawled and indexed by search engines. As a result, search engines devote crawling and indexing resource to these pages instead of focusing on more important parts of the ##clientname## website.

Pagination may also result in usability issues. For example visitors may arrive on a page displaying unexpected or out of date content. Typically, the optimal user experience is to arrive at the first page of results.

## IF SITE HAS DUPLICATED PAGE ONE ##

An additional issue is that the first page is duplicated across two URLs, because the link back to “page 1” includes a page number parameter. This results in duplicate content issues.

### Solution

## IF THE SITE FEATURES A “VIEW ALL” PAGE, USE THIS BIT… ##

The “view all” page is typically preferred by end users as it offers the best user experience. Therefore, we recommend using this as the “canonical” version of each page.

Include the canonical link element on all pages in a paginated sequence, with the link pointing to the “view all” page for that series.

## … OTHERWISE USE THIS BIT ##

Use the rel="prev" and rel="next" link elements on pages in the paginated sequence. This allows a page to specify which URL lies before it in the series, and which page follows. This must be set up as follows

* The first page in the sequence should only feature a rel="next" link element
* The last page in the sequence should only feature a rel="prev" link element
* All other pages in a paginated sequence should feature both of these tags
* The rel="prev" link element used on page 2 of any paginated sequence should point to the canonical version of that URL (without a page number in the URL).

Note: rel="previous" can also be used instead of rel="prev".

## IF SITE HAS DUPLICATE PAGE ONE (VALID FOR EITHER CASE ABOVE) ##

For duplicate first pages, we recommend that the link to “page 1” of results in any paginated series points to the default, canonical URL (without a page number in the URL).

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Optimising Facets

“Faceted navigation” or “filter navigation” is method through which content on a web site can be navigated to via a number of different methods. It allows site visitors to drill down into a site’s content by selecting different attributes which match what they are looking for.

It is currently used on the ##clientname## site for ##DESCRIPTION OF USE E.G. PRODUCTS##. For example a visitor could select ##CUSTOMISE FOLLOWING EXAMPLE## a dress which has a type of “maxi dress”, a colour of black, be within a specific price range and made by a specific brand.

The facets on the site are currently not fully optimised for search ##SUMMARY OF CURRENT ISSUES##.

#### Low value facets

The ##EXAMPLE## facets are potentially of low value for search. It is not desirable for these variants to appear within search engine results.

#### Multi-select

The sites feature multi-select facets, i.e. where a site visitor can select more than one ##EXAMPLE FACET E.G. COLOUR## at once. These are typically low value for search, as they are not well targeted to typical search queries.

### Solution

Implement the following on all key facet pages, ensuring that optimal keywords and key phrases are used throughout.

##DELETE AS APPROPRIATE FROM FOLLOWING TWO LISTS##

* Unique, relevant title tag
* Unique, relevant meta description tag
* Unique, relevant <h1> heading tag
* Targeted, relevant copy
* SEO-friendly URL ##EXAMPLE##

Additionally, we recommend the following:

* Use relevant anchor text on links which point to important facet pages
* List key facet URLs within XML Sitemaps

#### Low value facets

Instead of optimising these for search, these facets should be excluded from search engine indexing. The optimal method of achieving this is to include a canonical link element on all low-value facets which points to the closet-matching high-value facet.

For example the ##EXAMPLE## facet should feature a canonical link element which points to the ##EXAMPLE## facet instead.

#### Multi-select

The sites feature multi-select facets, i.e. where a site visitor can select more than one ##EXAMPLE FACET E.G. COLOUR## at once. These should also be excluded from search indexing via the canonical link element, as described above.

#### Pagination

The faceted navigation on the ##clientname## site also utilises pagination. Please see the section on pagination for more information. ##ENSURE THIS SECTION IS INCLUDED##

Flash Content Issue

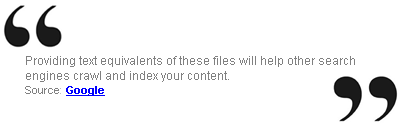
##clientname## makes use of Flash on various areas of the website.

##

Over the years, search engines such as Google have been working hard to improve their support for Flash and have more recently been able to take more value from Flash elements, such as text and hyperlinks. Still, while crawling of Flash content is possible, results are poor in comparison to a standard HTML file, and this will likely hinder the promotion of any web pages which feature a significant amount of Flash content.

In addition, some mobile devices – notably Apple’s iPad and iPhone – do not support Flash.

### Solution

Where Flash is used heavily, we suggest that an accessible alternative is provided. This should be in the form of accessible HTML.

Using CSS, wrap the accessible content in a DIV and position it either behind the Flash element or off the screen. One method of doing so is to use SWFObject.

This ethical form of Flash replacement will allow search engine spiders to crawl and index the content and attribute it to the web page it resides in.

#### Example

The Flash element above contains the following copy:

##

This copy could then be styled in HTML to produce:

<div></div>

Applying CSS to the DIV will then allow ##clientname## to render it behind the Flash element, making it accessible for both site visitors without Flash and search engines. Where accessible alternatives are used, we will audit and provide feedback where necessary, to ensure that ethics are maintained.

### Alternative Solution

Creating accessible alternatives and styling them accordingly may initially be too large a commitment for ##clientname##. If this is the case, we recommend that the noscript tag is used.

<noscript>Copy contained within Flash element</noscript>

The noscript tag can be used in accordance with the W3C guidelines to provide an alternative to any form of scripting, in this case Flash. ##clientname## could extract the contents from the Flash element, create a basic HTML alternative and then place it in a noscript tag next to the Flash element.

#### Cross Digital Impacts

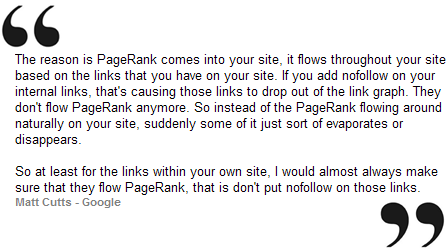
## Are there any cross digital impacts? ##

Internal Use Of rel="nofollow" Issue

Some links on the ##clientname## website use the rel="nofollow" attribute unnecessarily.

##examples##

Using the rel=”nofollow” attribute on less-important internal links was once considered a valid optimisation technique (widely known as PageRank sculpting). In the past, when this attribute was appended to a link the intended PageRank that would have flowed to the link was distributed elsewhere to other links.

However, Google changed the way in which PageRank is calculated with rel=”nofollow” attribute links and the use of this technique can now be detrimental to a campaign.

Instead of distributing the PageRank that would normally have flowed from the rel=”nofollow” link to other links, the PageRank is now simply lost. This can result in less PageRank flowing through the ##clientname## website.

### Solution

Remove all instances of the rel=”nofollow” attribute on internal links.

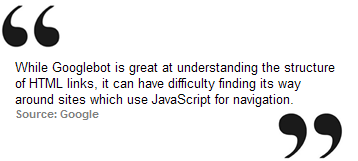
#### Cross Digital Impacts

## Are there any cross digital impacts? ##

JavaScript Navigation Issue

The site makes heavy use of JavaScript within key navigational elements. In particular the main navigation element is entirely composed in JavaScript.

##

Search engines support some aspects of JavaScript, but do so poorly. Search engines frequently struggle to take any value from content that is held within JavaScript, especially when complex JavaScript is used.

The use of JavaScript within the ##clientname## navigation renders the promotion of content held behind such JavaScript inaccessible to search engine spiders.

### Solution

Ideally, we suggest that JavaScript not be used for key functional elements of the website.

In order to open up the website to allow search engine spiders to reach content held behind the JavaScript more effectively, we suggest that you look to redevelop the navigational element by making use of (X)HTML and CSS, with JavaScript solely for non-essential functionality.

Doing so will allow ##clientname## to maintain the current aesthetics and functionality of the current navigation, but achieve the goal of accessibility and search engine compliance.

### Alternative Solution

If there is no immediate scope to redevelop the current JavaScript navigation, we suggest that you make use of the <noscript> tag. This tag can be used in accordance with the W3C guidelines to provide an alternative to any form of scripting, in this case JavaScript.

##clientname## should extract the contents of the JavaScript navigation, create a basic HTML alternative and then place it in a <noscript> tag next to the JavaScript navigation. All user-agents that do not have the ability to render JavaScript (search engine spiders) will then be shown the HTML alternative. Search engine spiders will then be able to follow and attribute PageRank to the destination content.

<noscript><a href=”##clienturl##/#content/”>Content></a></noscript>

If there is the scope to generate a Sitemap this should also be considered. Listing all URLs within a Sitemap will increase the likeliness that the URLs are crawled and included (not promoted) within a search engine's index.

Any alternative ways of providing links to such content should also be incorporated. Contextual links and sub-navigations are the most effective ways to achieve this successfully.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Form Field Navigation Issue

Certain internal navigation elements are developed using form field navigation:

##

Search engines support for interpreting and crawling content behind form field navigation has improved over the years but is still limited. Consequently, we recommend that alternate means of navigation are used, either in place of or as well as form field navigation, to ensure optimal performance.

### Solution

In order for ##clientname## to make content held behind the form field navigation more accessible to search engine spiders, we suggest that HTML and CSS are used to mimic the aesthetics of form field navigation while achieving improved accessibility.

Alternatively, the navigation should be completely redeveloped.

### Alternative Solution

If the above solution is not desirable, we suggest that ##clientname## provides alternative paths to content that is held behind forms.

This can be achieved by, for example, incorporating a public facing site map containing a top level list of links, a sub navigation element that contains a list of links to such content.

##

Both alternatives should use contextual links. Contextual links can be placed within copy to offer further routes through which search engines may find the content.

Additionally, we suggest that ##clientname## lists all URLs within an XML Sitemap and submits this to all major search engines. This will assist in allowing URLs to become indexed, although it is important to note that pages which are only listed within XML Sitemaps typically rank less well than pages which can be accessed via spiderable links.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Public-Facing Site Map Issue

When crawling the ##clientname## website, a public facing site map wasn’t found.

A public-facing site map is a top-level list of links arranged in a structured hierarchy, designed to outline a website's structure to the user and to search engines.

Typically found in the footer of a website’s navigation, a public facing site map is beneficial to search engines and users in a number of ways:

* They provide a search engine spider with an indication as to how the directory and file structure of the site is arranged, which is useful when indexing a website.
* They provide one location from which every top-level web page can be accessed. This prevents web pages from being missed and stops the search engine spider from having to travel down many links to access website content.

Site maps also provide support to specialist browsers and users with special accessibility requirements.

### Solution

In order to provide a consistent route for search engine spiders and users to find deeper content within the ##clientname## website, we recommend developing a public facing site map.

Convention states that the site map is laid out according to the design of a website in a tree-like structure. Although it is not a requirement, we recommend following this structure as it has become an unwritten standard.

##

Ideally, a site map should be accessible from every page and link to all the top-level content that a website contains.As new pages are added to the ##clientname## website, it may also be necessary to update the contents of the public facing site map to include such links.

SpeechMarkOpen.png

Offer a site map to your users with links that point to the important parts of your site. If the site map has an extremely large number of links, you may want to break the site map into multiple pages.

Google

SpeechMarkClose.png

### Alternative Solution

If there is no scope to develop a public facing site map, we suggest altering existing web pages to include more contextual links or lists of links that should enable search engine spiders to crawl and assign PageRank value more effectively throughout the website.

#### Example

##

Adapting existing pages to introduce lists of links or contextual links will be an effective approach, but ideally a public facing site map should be developed.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Server Error Issue

There are a number of web pages returning a server error on the ##clientname## website.

#### Examples

## EXAMPLES ##

These pages return a “200 OK” status code, which incorrectly indicates that a valid page is being served, which is not the case.

Actively linking to and such pages could be damaging to the ##clientname## brand, as it provides a poor user-experience. It could also create duplicate content issues, as the web pages are nearly identical.

### Solution

Where web pages are no longer required, implement a 301 permanent redirect from the problematic web pages to a similar, active counterpart.

In order to prevent this issue from occurring in the future, consider setting the server to return a 500 or 503 error code when errors occur so that search engine spiders will detect that an error has occurred and cease to crawl the web page. This is suitable only for temporary issues, and web pages should not continue to serve 500/503 status codes indefinitely.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Restricted Spider Access Issue

The robots.txt file on the ##clientname## website only allows specified user-agents to crawl the website and excludes all others. It does this by excluding all user-agents with the following directive:

##CHECK THAT THE FOLLOWING IS CORRECT##

User-agent: \*

Disallow: /

By disallowing some search engine spiders, crawling and indexing by these search engines will be very limited at best, heavily reducing (or perhaps even completely preventing) their traffic driving potential.

### Solution

Remove the user-agent restrictions in place.

##CHECK THAT THE FOLLOWING IS CORRECT##

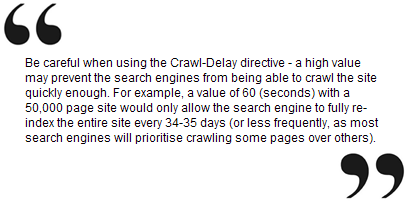
User-agent: \*

Disallow: /

If these restrictions were implemented due to bandwidth limitations, consider implementing the robots Crawl-Delay protocol.

Crawl-Delay: 0.2

The above directive would limit crawl requests to every 0.2 seconds. Crawl delay is supported by search engines such as Ask, Bing and Yahoo!, although not by Google.



#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Frames Issue

Certain areas of the ##clientname## website make use of frames and iframes – essentially pages within pages.

##EXAMPLE##

Although search engines can index frames, not all search engines always index them effectively. For example, some search engines may only index the empty frameset <noframes> content, whilst others may index content from external files as stand-alone pages without the support of the on-page navigation and page structure.

Search engines use URLs as the starting point of the algorithmic analysis process, and as the content within the frame is by definition a separate entity, it makes it harder for a search engine to attribute the value of the frame to the page hosting the frame.

### Solution

Ideally, the ## should be moved into to the main ##clientname## website. By restructuring the website so that it does not use frames, the content would be accessible to search engines.

##

The once framed content would then be served as stand-alone web pages with desired navigation and structure in place, increasing brand awareness, crawl-through potential and search exposure.

### Alternative Solution

If redevelopment of the supporting web page containing the frame is not possible, we suggest that ##clientname## creates additional content around the frame so that search engines can take value from the frame web page. Content should be in the form of copy, a paragraph or two containing at least 100 words about the web page, including any keyword targets.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Root Index Redirect Issue

The root index page of the site redirects to an internal page on the website.

####clienturl##/example/##

Users tend to link to the root of the domain as opposed to the destination of the redirect, thus splitting the link equity and PageRank value over two landing pages. While the redirect will notify search engine to concatenate the two values, not all search engines are equally effective at dealing with redirects. Additionally, sending users and search engines through a redirect slows down the response time of the ##clientname## website.

### Solution

We suggest that ##clientname## moves the contents of the homepage to the root index web page of the website, rather than redirecting the homepage elsewhere. For example:

##clienturl##

Should be the designated homepage as opposed to:

##clienturl##/##old homepage##

Once the move has been completed, implement a 301 permanent redirect from the old homepage URL to the new root “/” URL. This will migrate the authority of the old homepage URL to the new location.

Additionally, all internal links should be updated to point to the new homepage’s location.

#### Cross Digital Impacts

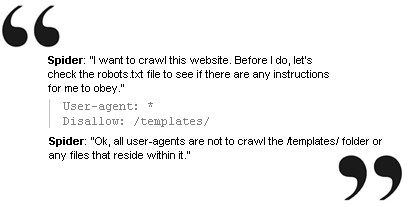
## Are there any cross digital impacts? ##

Robots Exclusion Protocol Issue

##clientname## does not feature a robots.txt file. Also known as the robots exclusion protocol, this file allows websites to give information about their website to search engine spiders.

Before a search engine spider crawls a website, it requests a file from the root of the domain named called robots.txt

##clienturl##robots.txt

Within the robots.txt file, websites can specify instructions that state which robots are excluded from which files or folders on the server.

The robots exclusion protocol can also be used to list the location of XML Sitemap files and define the time delay between crawling web pages.

### Solution

Every website should make use of the robots exclusion protocol. The robots exclusion protocol is considered a positive quality signal to sophisticated search engines such as Google. We recommend that you create a robots.txt file in a similar format to the following:

User-agent: \*

Disallow: /cgi-bin/

Sitemap: ##clienturl##/sitemap.xml

The reasoning behind the above robots.txt commands is:

* ##
* ##

Upon verification within Bing and Google webmaster consoles, there are tools available that allow you to generate and test the content of your robots.txt file.

* Sitemap Protocol

The “Sitemap:” directive within a robots.txt file directs search engine spiders to the location of an XML Sitemap or Sitemap Index file.

### Alternative Solution

If there is no ability to host a robots.txt on the ##clientname## server, we suggest using alternative spider control methods such as Meta robots, X-Robots-Tag and rel=”nofollow”. Such spider control methods should be used on a case by case basis.

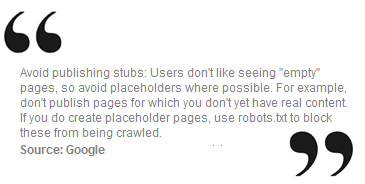
#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Placeholder Content Issue

There are a number of empty ‘placeholder’ web pages on the ##clientname## website, where there is currently no content but which may hold content in the future.

##clienturl##

Search engines dislike placeholders (also known as doorway pages) appearing in their search results, as they provide a bad user experience.

### Solution

Ideally, do not create placeholders in the future.

Where placeholders are not preventable, the optimal solution is to insert the following Meta robots tag into the <head> of the web page:

<meta name="robots" content="noindex" />

This Meta robots tag will prevent the web page from being indexed by the search engines, but will allow the search engines to crawl any links.

When the web page ceases being a placeholder, simply remove this meta tag. When a search engine re-crawls the web page, it will follow the meta robots directive and index the web page, allowing it to appear in the search results for promotion.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

404 Error Handling Issue

Search engines are required to communicate and to read the HTTP response of every web page. The HTTP response of a web page can be used to advise a search engine to undertake a particular action, e.g. crawl the page, follow a redirect, etc.

Search engines use the 404 Not Found HTTP header response as a form of error handling. If a web page echoes a 404 Not Found response, a search engine will cease crawling that URL.

Currently, ##clientname## does not make use of the 404 Not Found header response.

#### Example

##clienturl##/i-do-not-exist

Any requests for a URL that does not exist on the ##clientname## website returns a ## header response which ##.

### Solution

Any requests for a URL that does not exist on the website should return a 404 HTTP header response.

Upon implementation of 404 Not Found error handling, we can report any issues that search engines are encountering from search engine webmaster tools.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Directory Index Error Page Issue

When some URLs are accessed without the filename an unfriendly error message is shown. For example:

* ##example##

Screenshot of the above URL:

##screenshot##

This is a suboptimal user experience, as the site is missing the opportunity to expose the brand to a web visitor, and any visitors to the page are unable to proceed to other pages on the site (if visiting from another site, their likely action would be to press the back button and leave the site).

Visitors can commonly arrive at error pages for a number of reasons, such as links being written incorrectly or email clients breaking long URLs across several lines. It is important to ensure that visitors to a non-existent page are not turned away by the bad user experience.

### Solution

Add index pages to these subdirectories where suitable, such that appropriate content is shown when these URLs are accessed. Where this is not possible, implement 301 redirects to redirect users to a suitable alternative page.

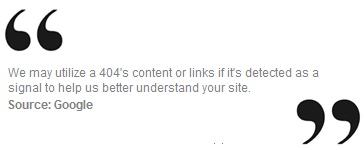
Branded 404 Page Issue

The site’s 404 page lacks branding and navigation.

#### Example

## EXAMPLE ##

##While the 404 Not Found technically issues the correct server header, a 404,## we would like to highlight this issue as a potentially suboptimal user experience, as the website is missing the opportunity to expose the brand to users, and some visitors to the page may be unable to proceed to other sections of the website.

Users commonly arrive at 404 not Found pages for a number of reasons, such as links being written incorrectly or email clients breaking long URLs across several lines. It is important to ensure that users to a non-existent web page are not turned away by a poor user experience.

This situation is also suboptimal for search. Although search engines will not index or rank a web page which serves a 404 status code, they may use elements of the 404 page within their algorithms in order to help them to understand your site.

#### Analytics ##IF APPLICABLE##

Additionally, the current 404 page lacks any analytics tracking tags for the analytics service currently being used (##Google Analytics##). This means that visitors to 404 pages are not being tracked within your analytics, and an opportunity to identify and track errors is therefore missed.

#### Usability

If a user happens to land on an erroneous URL, a custom 404 Not Found page may make the user more inclined to stay on the ##clientname## website by making it as easy as possible for the user to then find their desired content.

### Solution

Implement a custom 404 page with:

* Full branding and navigation
* Useful links to key sections, for example a cut-down version of the site map
* Site search functionality presented prominently
* Analytics tracking code

Another useful technique is to tag 404 Not Found pages with a unique tracking code. You could then set-up a goal to identify common and recurrent errors and, where found, identify solutions which could be put in place to resolve the issues.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Low Quality Content Issue

From our analysis, it appears that search engines have indexed a number of web pages which offer little commercial value to the ##clientname## campaign.

* ## EXAMPLES ##

In total, ## low quality web pages are currently indexed. Low quality content makes poor use of the finite crawling resources allocated to the website. It may also result in a poor user experience if users arrive at these pages via search engines.

### Solution

Exclude the low quality content using the following meta robots tag:

<meta name="robots" content="noindex" />

This tag instructs search engines not to index the web page, but allows it to follow any links. Careful consideration should be given to ensure that this meta tag is only included on the correct pages, as it is critical to ensure that valuable web pages are not de-indexed (removed from search engine listings).

### Alternative Solution

Add the following disallow rules within the robots.txt file.

Disallow: ##add rules##

Using the robots.txt file, we can prevent search engines crawling the low quality content, and in return, we expect search engines to spend more of their finite resources crawling more important web pages such as ##.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Printer-Friendly Duplicate Content Issue

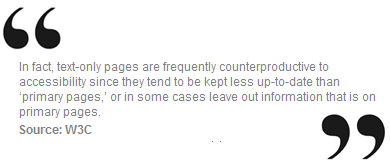
The ##clientname## site features accessible printer-friendly web pages which are identical to versions intended for desktop viewing. Search engines have indexed ##some/many## printer-friendly pages:

##screenshot##

#### Example

##clienturl##

As a result, there is a duplicate version of every web page with this feature, which may divide crawling resources and PageRank across multiple pages, and thus negatively impact search performance.



Contrary to popular belief, the W3C recommends against printer-friendly web pages. As such, our proposed actions won’t impact the accessibility rating of the ##clientname## website.

### Solution

Rather than creating a separate URL for printer-friendly content, serve a printer-friendly version of the desktop web pages via the CSS print media type. This is well supported by all of today’s major browsers.

<link rel="stylesheet" TYPE="text/css" media="print, handheld" href="print.css">

Once adopted, all existing printer-friendly web page should be 301 permanently redirected to their desktop counterparts. If any of the printer-friendly web pages have valuable links, the reputation of these links will be contributed to the desktop versions.

### Alternative Solutions

If the above solution is not possible, one of the following solutions could be utilised, although please note that these are less optimal than the approach listed above.

#### Canonical Link Element

Add the canonical link element to all printer-friendly web pages, with the URL pointing to the standard version of each page instead.

<link rel="canonical" href="http://www.example.com/" />

Please note that the canonical link element is treated as a hint by the search engines, and therefore may not work for all web pages. Even with the canonical link element in place, these web pages are still crawled, which means the spidering issues will still exist.

#### Meta Robots

Yet another approach is to exclude the printer-friendly web pages from search engines using the meta robots tags. The following meta tag would be included within the <head> element of each printer-friendly web page:

<meta name="robots" content="noindex" />

The disadvantage of this solution is that any printer-friendly web pages with valuable inbound links may not contribute their reputation to the desktop versions.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Staging Sub-domain Issue

The ##clientname## website features subdomains which appear/ to be used for testing the main website’s content pre-launch.

##clienturl##

However, if such sub-domains are not blocked by search engines, they may get crawled and indexed and potentially cause a duplicate content issue.

It is recommended to block search spider access to all staging sub-domains.

### Solution

The optimal approach is to set a password for the staging site, e.g., using access authentication. This will prevent search engines from every spidering the staging content and help to keep the work-in-progress material locked away.

### Alternative Solution

Likely a faster approach is to create and upload a robots.txt file that disallows crawling access to all search engines.

User-Agent: \*

disallow: /

User-Agent: Googlebot

noindex: /

Whilst the robots.txt file prevents access, it won’t result in the web pages being removed from search engine results immediately. Search engines will continue to show a cached version for a short period of time until they are confident that the web page is not to be indexed. Even then, search engines may continue to list the URL in their search results (typically minus a title or snippet).

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

HTTP Compression Issue

The ##clientname## website does not appear to be making use of HTTP Compression.

HTTP compression is a protocol that is a part of the HTTP 1.1 specification standard known as "content-encoding" which can be used to improve the bandwidth consumption of a website, improve page load times, and so improve user experience. HTTP compression enables the crawling of a website to be more efficient – the faster a search engine can crawl a website, the more pages on the website it may spider in a set period of time.

The protocol can be used by search engine crawlers to define whether or not a page or file is 'compression enabled' before retrieving the file from the web server; this is essentially the HTTP version of the commonly used ZIP data compression format.

### Solution

It is recommended that HTTP compression be used on all plain text files, e.g. HTML, CSS, JS, etc. If used in conjunction with a GET request, with the Last-Modified or If-Modified-Since, ##clientname## can make the crawling and accessing of the website more efficient for search engines.

#### Apache

If running Apache, you will be required to make use of the “mod\_deflate” tool, which will add a filter to compress any selected content as a gzip file. Apache allows you to apply filters to determine which particular MIME type you would like to compress. This can be determined automatically or by developing a custom script to do so. To enable compression for all MIME types, set the SetOutputFilter directive to a website, directory or sub-domain:

<Directory "/web/mysite/php/">

SetOutputFilter Deflate

</Directory>

Every website is different; you will require in-depth knowledge of servers in order to effectively implement and test this mod\_deflate requirement.

#### Windows IIS

If running Windows IIS 7, HTTP compression for static files is enabled by default. If you would like to compress all files, you will be required to make use of dynamic compression. This can be enabled by going to the 'IIS Services' panel and then selecting 'Compression'. Within this option, you can select your preferred compression type of either static or dynamic.

#### Cross Digital Impacts

HTTP compression will improve page load times, thus positively impacting user experience. Additionally, page load times are a factor in Google’s AdWords Quality Score.

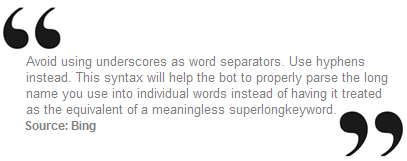
Compound Words & Underscore Issue

URLs for the ##clientname## website currently compound multiple words into a single word, e.g.:

* ##EXAMPLE##

Search engines usually cannot identify that these are multiple words and see a single, joined-up, nonsensical word. In the above URL, they would see the ‘word’ “##compound\_words\_in\_url##” as opposed to the separate words ( “##word##” and “##word##”).

Some search engine spiders, notably Googlebot, rely on specific characters such as a dash (-) to indicate a word space in a URL string. Whilst some search engines treat both a hyphen (-) and an underscore (\_) as a word separator, other search engines (notably Google) do not.



### Solution

We recommend using hyphens (-) rather than underscores (\_) to separate words in URLs (some other characters are also acceptable, e.g. “/” and “.”).

Once a URL has changed format, it is vital that 301 permanent redirects are set up to migrate users and search engines from the old URLs to the new ones. This is required in order to preserve the search performance of these pages.

### Alternative Solution

Where updating word separators is not possible due to technical commitment or potential loss of keyword performance, we recommend that this principle is adopted for all new URLs.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Alt Attributes Issue

Search engines cannot see images. In order to take any value from an image, a search engine will examine the copy around an image in order to decipher what the image might depict. In addition, an alt attribute can be associated with an image, which provides an alternative textual description of the image. In areas of the ##clientname## website, images are missing alt attributes.

#### Example

##

The alt attribute is a requirement to meet basic accessibility guidelines, and plays an important role in optimising pages for image search engines such as Google Image Search.

### Solution

All images on the ##clientname## website should have meaningful alt attributes. If there are any images that are used for design only, these should contain an empty alt attribute or be applied with CSS.

When assigning alt attributes, it is best to describe exactly what the image conveys, as being deceitful will not only impact users using specialist browsers for accessibility purposes, but may impact your search performance.

#### Links

Ideally, no images should be used for conveying hyperlinks. This is because search engines tend to use the anchor text of a link to pre-determine the content of destination web pages.

We have found using copy to be more effective for search engine performance than using alt attributes as the anchor value. This suggestion comes with exceptions: the logo of a website is typically expressed via an image and is almost certainly used to link the user to the website homepage.

### Alternative Solution

In some instances it may be suitable to use image replacement techniques.

#### Cross Digital Impacts

Not using alt attributes may influence how poorly images appear in Universal Search results.

Trailing Slash Canonicalisation Issue

URLs on the ##clientname## website are accessible both with and without the trailing slash character (/). Since the two versions are regarded by search engines as distinct, separate URLs, this results in duplicate content issues, as the same content is being made available on different URLs.

For example, this URL:

##clienturl##/##

Serves the same content as this URL:

##clienturl##/##

These URLs are different to a search engine crawler due to the omitted trailing slash (/) in the ##second/first## version.

Multiple URLs serving the same content results in duplicate content, spreading PageRank over more than one location. This impacts the crawling efficiency of the ##clientname## website and potentially impacts the promotion of the site.

Based on our analysis, there are up to ## instances of this issue. ##Additionally, there is a mix of external links referencing both versions reported via Google Webmaster Tools.##

### Solution

All internal links should point to the “canonical” version chosen.

We suggest that you look to make use of server-side technology to apply a blanket rule that will automatically 301 permanently redirect any URLs omitting the trailing slash (/) to their preferred counterpart.

Whilst the version chosen as canonical will have little impact on search engine rankings, we suggest 301 redirecting directories without a trailing slash to the same page with a trailing slash appended to the end of the URL, as this is the most commonly used approach. This should be applied consistently across the site.

#### Apache Server

Within an Apache Server, you can add rewrite rules to your existing .htaccess file that will apply this solution site-wide.

RewriteEngine On

RewriteBase /

RewriteCond %{REQUEST\_FILENAME} !-f ### Don’t add a / to filenames.

RewriteCond %{REQUEST\_URI} !example.php ### Don’t rewrite me.

RewriteCond %{REQUEST\_URI} !(.\*)/$ ### I rewrite your URLs.

RewriteRule ^(.\*)$ http://example.com/$1/ [L,R=301] ### Final redirect

Using expertise within the IT or web development area of ##clientname##, this rule should be customised and applied accordingly.

**Note:** Rule may differ per server/website.

### Alternative Solution

If using server-side technology is not immediately possible, update all internal links to reference the canonical URLs.

If there are any external links which reference a non-canonical version, consider adding the canonical link element to the destination landing page, for example:

<link rel="canonical" href="http://##" />

When a search engine encounters the canonical link element, it will update its index and promote the preferred version.

#### Cross Digital Impacts

This issue may impact affiliate, display or paid search landing pages if a problematic URL is used. Ideally, all destination landing pages should be updated to reference the new definitive URL string.

Search-Generated Content Issue

Search engines do not like to show other sites’ search results within their search results. Currently, the ##clientname## internal search function generates a unique URL for each query entered, which can then be accessed and indexed by search engine spiders.

Example of search results indexed:

##screenshot##

Not only are such pages a cause of poor quality search results, they will also be allocated valuable crawl time which may deprive more significant areas within the website.

### Solution

Exclude low quality search generated web pages from the index by using spider control methods.

#### Meta Robots

We recommend that the meta robots tag is used to disallow the indexing of search generated content, while allowing search engines to follow links within these pages. This can be achieved by placing the meta robots tag in the <head> of the search generated web page's source code.

<meta name="robots" content="noindex" />

Users frequently link to search generated content, so by using the meta robots tag in the suggested manner, if this does occur, the site will benefit from the external link.

#### Links to Search Generated Content

Due to the low value assigned by search engines to search generated content, we suggest that ##clientname## does not link to the search generated content internally. This will prevent any anchor text or PageRank value from being assigned.

### Alternative Solution

If using the meta robots tag is not possible, an alternative (albeit less optimal) approach is to use the robots exclusion protocol to prevent the crawling and indexing of the search generated content.

This can be achieved by adding the following line to the ## robots.txt file:

User-agent: \*

Disallow: /##

Over time, the disallow rule will result in any search generated pages falling out of the index.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Non-WWW Domain Canonicalisation Issue

The ##clientname## website resolves both with and without the 'non-www' prefix at the start. This allows search engine spiders to crawl the 'non-www' version of the website, resulting in duplicate content issues.

Allowing such duplication to exist takes valuable crawling resource away from the ‘canonical’ version of the domain. Additionally, if links are made to this other version, the benefit is split across multiple locations.

Currently, there are ## 'non-www' URLs indexed in ##search engine(s)##. In addition, there are ## external URLs linking to the 'non-www' version.

### Solution

Select either the 'www' or 'non-www' version of the site and then 301 permanently redirect any requests for the canonical version to the preferred domain prefix. This redirect should maintain the URL path. Based on existing search engine exposure, we would favour the '##' version.

After the 301 permanent redirects are in place, external backlinks, rankings and history that the non-canonical hostname once had will be passed over to the canonical version.

We also recommend setting the “Preferred Domain” within Google Webmaster Tools. This feature allows website owners to specify whether the “www” or “non-www” version of the site should be indexed within Google.

### Alternative Solution

If 301 redirects are not possible, we recommend setting the “Preferred Domain” within Google Webmaster Tools. This will only resolve the issue for Google.

A number of other techniques can also assist with canonicalisation:

* Specifying the correct version for URLs within XML Sitemaps
* Using absolute links for template navigation elements, pointing to the preferred version.
* Using the canonical link element, and specifying the preferred version for URLs listed

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Robots Exclusion Issue

The robots.txt file on the ##clientname## website has a disallow rule that instructs search engine spiders not to crawl ##all/some important## web pages, likely negatively impacting rankings and traffic.

User-agent: \*

Disallow: \*

Search engines will try to index web pages excluded by robots.txt if the pages have inbound links. However, the disallow rule prevents search engines crawling the specified pages, thus they will not be able to see any of the pages’ content or follow any links within them. This reduces the opportunity for such web pages to rank for desired keywords.

#### Example

##

From our example, we can see how the disallow rule has prevented the search engine creating a full snippet for the web page (minus title and/or description).

### Solution

In order to enable the effective crawling and indexing of excluded content, we recommend removing the disallow rule, allowing the web pages to be crawled by all search engines.

In time, the previously disallowed web pages will be crawled and indexed, enabling them to be returned in the search engines for desired keywords.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Secure Server Issue

Search engines are cautious of privacy, preferring to crawl web pages hosted on the Hypertext Transfer Protocol (HTTP) as this is the domain in which websites tend to host public available data. The Hypertext Transfer Protocol Secure (HTTPS) is a combination of the Hypertext Transfer Protocol and a network security protocol which is frequently used to pass and to host secure information such as user specific account details. Ideally, the HTTPS protocol should only be used for hosting secure content.

In addition, where identical web pages are available both on HTTP and HTTPS protocols, a duplicate content issue may arise.

We have found that the ##clientname## site makes unnecessary use of HTTPS on a number of web pages.

##example##

Search engines have indexed some of these pages:

##screenshot##

This results in duplicate content issues.

### Solution

We recommend re-thinking how existing content resides on the ##clientname## website and moving core content that resides on the HTTPS protocol to the HTTP.

##clienturl##

Should be moved to:

##clienturl##

In doing so, all internal links within the ##clientname## website should be updated to point to the new destination of the contents. This will ensure that the new destination URL begins to receive anchor text and PageRank value.

#### Redirects

It is essential that 301 permanent redirects are put in place to redirect users and search engines to the new destination of the contents. When a search engine finds the redirect in place, it will attribute previous history, link and PageRank value to the new destination, making it imperative that redirects are put into place for search engine promotion.

#### Use Absolute Links

Where pages link between the HTTP and the HTTPS versions we recommend using absolute URLs within these links, rather than relying entirely on the 301 redirects to resolve the issue. For example, site navigation links within your page template(s) may use relative links. If a page using such a template is hosted securely – for example a customer login page – links followed from that page will acquire the HTTPS protocol.

### Alternative Solutions

#### New Content on HTTP

If there is no immediate resource available to move content, we suggest that if ##clientname## adds any new content, it is placed on the HTTP protocol as opposed to the HTTPS protocol. This will allow you to build on existing exposure in a more suitable platform for promotion.

#### Canonical Link Element

Add the canonical link element to all HTTPS content, with the URL referencing the HTTP rather than the HTTPS protocol.

The canonical link element present on the HTTPS web page would be:

<link rel="canonical" href="http://##EXAMPLE##" />

If there are web pages which should only be accessed via the HTTPS protocol, the canonical link element should point to the HTTPS protocol instead.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Default Index Page Duplicate Content Issue

The ##clientname## site has canonicalisation issues due to the use of default index pages.

For example, the following are present and have identical content:

* ##URL WITHOUT INDEX PAGE##
* ##URL WITH INDEX PAGE##

This is present on ##the homepage and on each subdirectory on the site.##

Search engines treat each different URL as a separate page, and this issue therefore results in duplicate content issues. For example, the benefit of any links made to both page types will be split between them, rather than being concentrated on one page.

### Solution

Update all links on the site to point to the ‘/’ URLs instead of the ##‘/index.html’## URLs.

Once all links have been updated, implement 301 redirects from the non-canonical URL forms to the ‘/’ forms.

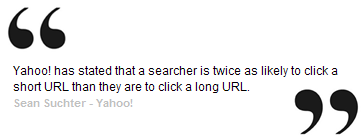
For example, this URL:

* ##URL WITH INDEX PAGE##

Should be 301 redirected to this URL:

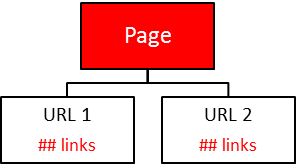
* ##URL WITHOUT INDEX PAGE##

The choice of ‘/’ instead of ##‘/index.html’## for URLs is mostly for presentation and usability purposes – the latter is a ‘cleaner’ URL which is simpler to read and is less likely to be truncated by search engines when listing the URL in the results. This typically results in better click-through rates.



Canonical Homepage Issue

Search engines treat each different URL as a separate entity. Where identical content is served on multiple, technically distinct URLs, a search engine may perceive duplicate content. For this reason, it is important to attempt to ensure that each page can only be accessed via one URL.



Combined: ## links

The ##clientname## website has canonicalisation issues with the homepage.

##clienturl##

is the designated ##clientname## homepage. This is where the largest percentage of internal and external links refer.

However, there is also a duplicate version of the homepage at the following location:

##clienturl## ##URL OF DUPLICATE PAGE##

Splitting anchor text and PageRank value between two pages rather than consolidating it in one location may be impacting the performance of the ##clientname## homepage. Typically, the homepage of a website drives the majority of website traffic, hence the need to ensure optimal performance.

### Solution

Update all internal homepage links to point to the designated canonical homepage URL. This will ensure that all internal anchor text and PageRank values are being sent to one location, which will focus the benefit of any internal links to one location.

#### Redirection

Redirect any duplicate homepages to the designated homepage URL using a 301 permanent redirect. This will advise search engine spiders to attribute all links to the designated homepage.

When implementing this, ensure that an endless loop of redirections is not created on the homepage.

### Alternative Solution

If time and development constraints mean it is not possible to update all internal links to the designated homepage location, and the redirection is not possible, we suggest that ##clientname## makes use of the canonical link element.

This can be used on each page to specify the preferred version of its URL to search engine spiders. This will help resolve the ##clientname## canonical homepage issue, as search engines will then begin to attribute the value of the duplicate page to the destination URL listed within the tag.

<link rel="canonical" href=”##clienturl##” />

Note: Search engines treat the canonical tag as a strong indicator that the value of a web page value should be attributed to another location, although this is not guaranteed.

Root index page redirects to internal page

The root index page of ##clienturl## redirects to an internal page on the site using ##REDIRECT TYPE##.

As a significant number of links to a site link to its root index page, and, as this is the main page of a site (which most site users will see first and from which they will then traverse to all parts of the site), it is usually the most important page of any site.

Redirecting to an internal page is an unexpected action from a user’s point of view and also goes against standard hierarchical site structure design.

Redirecting with a temporary redirect can split the benefits of any incoming links between the two pages, as some websites may link to the root URL and others may link to the page to which the root redirects. If a permanent redirect is used, it may be bad for branding, as only the destination URL may appear in the search results and the domain alone may not be listed.

### Solution

Move the contents of the landing page to the root index page of the site, rather than redirecting the homepage elsewhere. For example, when visiting ##HOMEPAGE URL## the contents of the page that it currently redirects to should be shown.

Once this has been done, 301 redirect the old destination URL to the new “/” URL, i.e. redirect this URL:

##URL OF PAGE THE HOMEPAGE REDIRECTS TO##

To this URL:

##HOMEPAGE URL##

Canonical Domain Issue

Duplicate content is bad for a site’s performance in search engines. There are a number of reasons for this, but principally, the problem is that multiple instances of the same (or very similar) content can produce low quality search results, distribute a page’s authority across multiple URLs, and unnecessarily increase the load on search engine spiders.

##clientname## appears to own a number of domains which have identical, or close to identical content to the main ##clienturl## website.

#### Examples

* ##Example site##
* ##Example site##

Having multiple domains serving the same content may be negatively impacting the search performance of the main site.

#### Additionally

Some websites have placed links to the duplicate domains, ## to be exact. Such links are inevitably being wasted by linking to the incorrect website.

### Solution

Every duplicate domain and its contents must be 301 permanently redirected to an appropriate location within the ##clientname## website. Where no relevant location is found, a redirect to the ##clienturl## homepage is typically the next best solution.

#### Example

##duplicate-url##

Should be 301 permanently redirected to:

##client-url##

By 301 permanently redirecting all duplicate pages to the main ##clientname## website, the offending pages are removed from a search engine index and search engine spiders are advised to pass any link value, PageRank and legacy over to the destination provided.

Additionally, the ## external links that once pointed to the external domains will be passed over to the ##clienturl##, thus positively improving ##clientname## 's link profile.

### Alternative Solution

People started buying domain name variations to account for user error in the late 1990s, but with the majority of users now using search engines as their starting point, the risk of being unable to find a website is nearly eliminated. If there are no issues surrounding the redirection of the duplicate domains, we suggest that you look to verify each such domain within Google Webmaster Tools, which includes a setting that allows users to notify Google of a change of address. This setting could be applied to notify Google that these domains should be associated with the main ##clientname## website.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

URL Language Issue

The ##clientname## site uses ##language## within its URLs for ##different language## content.

### Example

* ##EXAMPLE, e.g. English words used in the URLs of a French site##

This will reduce click-through rates for the site, as it does not target the local language. This will typically result in reduced traffic.

### Solution

Implement localised URLs for the ##clientname## site.

For example, for this URL:

* ##EXAMPLE, e.g. English words used in the URLs of a French site##

Could be written as follows:

* ##EXAMPLE of it being fixed e.g. French translation of the English words##

It is vital that 301 redirects are applied to all affected URLs.

Cookie Based Multiple Language Issue

Pages on the ##clientname## site are available in alternate language versions depending on a cookie set in the user’s web browser, which causes issues for search engine spiders as they discard cookies along with the information they contain.

Alternate language content will therefore have ##limited/no## visibility in the search engine results pages, causing a big negative impact on traffic from native search queries. In this case, it is the ###LANG### language version of the site that is inaccessible to search engine spiders.

An additional issue is that links to these pages may not send users to the intended language version, and in some cases links will be lost as alternative link targets will be sought.

### Solution

Pages containing multiple language content should be split into separate pages, one for each language, and language should be consistent throughout a pages content and metadata.

Many smaller language or country-specific search engines have respectable market shares and may use this language-specification information. It is therefore recommended to use multiple mechanisms in order to ensure that maximum exposure in local search engines (as long as care is taken to ensure that all of these mechanisms remain up-to-date).

The correct language should be specified within the http-equiv "Content-Language" meta tag and, ideally, the lang (and xml:lang for XHTML) attribute of the <html> element. You can use other language-identification mechanisms as well, such as the Content-Language HTTP header and the hreflang attribute of the HTML anchor tag (<a>).

SpeechMarkOpen.png

*If you are directing your website’s contents toward a specific language-speaking audience, you can specify the language of your content using the <meta> tag’s content-language attribute. For example, for a target audience of American English speakers, you would add the following tag to the <head> section of all your pages:*

*<meta http-equiv="content-language" content="en-us" />*

Bing

SpeechMarkClose.png

Where a language is shared across separate markets it is important to include locale information to avoid duplicate content issues. For example; a site serving content to both the US and UK markets could be seen as duplicate content unless the locale is appropriately set.

##if multiple languages are used within a single country##

For countries with more than one language such as the ##clientname## site, we recommend using a subdomain with the appropriate ISO language code. For ##clientname## ##COUNTRY## an appropriate configuration would be:

##SUBDOMAIN##.##clienturl##.XXXX# for the XXX version of the site.

##SUBDOMAIN##.##clienturl##.XXXX# for the XXX version of the site.

##SUBDOMAIN##.##clienturl##.XXXX# for the XXX version of the site.

In this case the default domain may still be present as a language selection page, linking to each of the different-language sub-sections. This page could also check the visitor’s language settings (if any are set), and redirect users with a supported language to the page for that language. In this particular case, a 302 temporary redirect is more appropriate than a 301 permanent redirect.

Each language sub-section of this page (for example, the section of the page stating in French that the user should click on the specified link in order to visit the French-language version) could include the HTML “lang” attribute, in order to specify the language of that block of text on this page.

Language settings should be configured in Google Webmaster Tools to delineate the location and language of content in each language.

Consult your Technical Account Manager for further assistance optimising for multiple languages.

### Alternative Solution

#for multiple languages on the same URL#

For multiple languages on the same URL, subdirectories can be used in place of subdomains. For ##clientname## this would be:

##clienturl##.XXXX#/##subfolder##/ for the XXX version of the site.

##clienturl##.XXXX#/##subfolder##/ for the XXX version of the site.

##clienturl##.XXXX#/##subfolder##/ for the XXX version of the site.

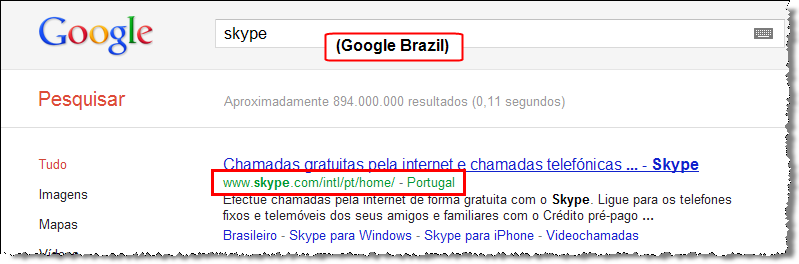
#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Multilingual Content Markup

The ##clientname## site serves similar content over various languages. It can be a major problem for search engines to decide which page to show in which results.

##Replace screenshot below with example image of site page appearing in wrong country results##



##In the example above, a Portuguese page appears in place of a Brazilian page.##

An additional problem is that by publishing the same content over multiple pages, ranking signals such as PageRank (from inbound links) and Social Signals (Facebook Likes, Google +1s etc.) are spread over numerous pages, weakening all of those pages, resulting in lower rankings.

#### Solution

Google’s multilingual content markup allows site owners to specify the language and location target of alternative pages where content has been provided for multiple geographic regions.

For example, the page ##change example.com URLs to actual examples##

## http://wwww.example.com/en-GB/example.html

would utilise the following lines of HTML in the <head> section:

## <link rel="alternate" hreflang="en-US" href=" http://wwww.example.com/en-US/example.html" />

## <link rel="alternate" hreflang="en" href=" http://wwww.example.com/en/example.html" />

This allows Google to consolidate all the ranking signals for a set of equivalent pages whilst being able to present the correct one for each language and region.

##The next part is optional – if you are going to use it, make sure that you have thought about the potential consequences in smaller search engines, especially if Google doesn’t have dominant market share in the specific country.##

#### ## Optional Canonical Link Element

When using the above markup, Google allows the use of canonical link elements on pages containing substantially similar (same language) content. This allows Google to know which page to show when no specific page exists for a country, whilst still showing the relevant versions in other countries due to the multilingual markup.

For example, where no specific Australian page exists, it would be best to show an international English page in Australian results.

In that case, canonical link elements would be used in the <head> section of all the other English pages, pointing to the international English version:

## <link name="canonical" href="http://www.example.com/en/example.html">

Note: Only Google recognises the multilingual markup and uses it to override the canonical link element; other search engines do not recognise it. Implementing the canonical link element in this case may mean that the wrong page is displayed in other search engines such as Bing, Yahoo, Yandex etc.

Internal Redirect Issue

The ##clientname## site features a number of pages that redirect internally.

For example:

##URL## redirects to ##URL## ##302/meta/JavaScript##

##URL## redirects to ##URL## ##302/meta/JavaScript##

##URL## redirects to ##URL## ##302/meta/JavaScript##

These redirects are ##DELETE AS APPROPRIATE## 302 temporary redirects/meta refresh redirects/JavaScript redirects.

##The problem with 302 redirects is that they are treated (as the name implies) as a temporary measure by the search engines. This can lead to both the redirecting and destination page being indexed, possibly causing duplicate content issues. This is a result of a search engine “expecting” that redirection is temporary, and continuing to crawl and index both pages. Another issue with 302 redirects is that they can be susceptible to “302 page jacking” which can lead to ranking and authority being ‘leached’ from the target page.

SpeechMarkOpen.png

302s are only temporary redirects, and unlike with 301s, no link juice credit is passed to the redirected page. Using a 302 redirect is not a coding error per se, but much of the time it is a strategic error from the perspective of SEO. Unless you have a genuinely temporary need to redirect a page, stick with 301s as an SEO best practice.

Bing

SpeechMarkClose.png

##Meta refresh are problematical as they can be treated as either permanent or non-permanent. This depends on a variety of factors including the specific search engine, the timeout value and the source and destination URLs.

##JavaScript redirects are worst of all redirects to use as search engine spiders have trouble understanding them. A search engine may index the redirecting page as a valid page as well as indexing the destination page (if it is found at all). This can lead to at best, duplicate content issues, or in a worst case scenario, a search engine may regard the JavaScript redirect as an attempt at “JavaScript cloaking”.

### Solution

Unless truly temporary, all redirects should be changed to be 301 (permanent) redirects to expedite the transition of visitors and link equity to the destination URL. The destination URL will then take the place of the old URL in the search results when it is discovered by the search engines.

It is also important to ensure that internal links are updated to point to the intended URL.

### Alternative Solution

If it is not possible to alter existing client side redirects, we suggest altering any internal links within the website to ensure that they point to the redirected page.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

URL Case Sensitivity Issue

## NOTE – THIS CAN ALSO HAPPEN WITH APACHE RUNNING ON A WINDOWS SERVER – CHECK THE HTTP HEADERS AND CHANGE SERVER NAME BELOW AS APPROPRIATE ##

The HTTP specification states that URL paths are case-sensitive, but the web server used (##Microsoft IIS server##) is case insensitive. This leads to a situation where pages can be returned irrespective of the case used. Since URLs which differ only in case are technically distinct, this leads to multiple URLs with the same content.

Example of case sensitive variants on the ##clientname## website:

##URL in one case##

##URL in another case##

Although search engines do not automatically look for alternate case version of URLs, links which point to the page can feature the path in the ‘wrong’ case, which causes the duplication issue.

Example:

##SCREENSHOT##

### Solution

The simple solution is to permanently (301) redirect URLs that contain letters in upper case to the same URL in lower case. This can usually be achieved with a single URL rewrite which is preferred from both an implementation point of view and a server performance point of view, as large numbers of rules checked before each page is served can slow down page load times.

LBi can provide you with technical support to help you achieve this.

We also recommend ensuring that all links point to the canonical version of each URL.

### Alternative Solution

Ensure that all internal links point to one URL version of each page in the same case. Consider writing rules for any CMS to ensure that incorrect case URLs are not manually entered.

Use the canonical link element on all pages with the URL specified in lower case.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

XML Sitemaps

The ##clientname## site does not feature an XML Sitemap.

An XML Sitemap is a useful tool that helps search engines to find pages on a website. As the XML Sitemap is updated the search engines can be ‘pinged’ to make them aware of the changes to the site.

This provides search engines with a complete and up-to-date view of the entire site, and can also give them extra information on how often pages on a site change, the date each page was last modified and the relative importance of different pages on the site. Additionally, when search engines discover duplicate URLs on a site, they use the URLs listed within XML Sitemaps to determine which URLs to show within their search results.

### Solution

Create an XML Sitemap, in order to maximise the information available to the search engines and minimise the time new pages take to be indexed.

Ideally XML Sitemaps should be generated dynamically so that changes to the sites structure are automatically included. We recommend using the <lastmod> and <priority> tags for all URLs within the Sitemap.

Each XML Sitemap can list up a maximum of 50,000 URLs and be a maximum of 10MB in size. Larger Sitemaps need to be broken up into multiple smaller files, and referenced within a Sitemap Index file. Note that this can also make use of the <lastmod> tag, to indicate when each Sitemap file listed was last updated.

The XML Sitemap should be referenced with the site’s robots.txt file and submitted to the major search engines.

### Alternative Solution

Create a static XML Sitemap. This will not have all of the benefits of a dynamically generated XML Sitemap, as search engines will not be able to discover changes to pages on the site via this Sitemap, but it will provide some benefits with respect to search engine crawling.

#### Cross Digital Impacts

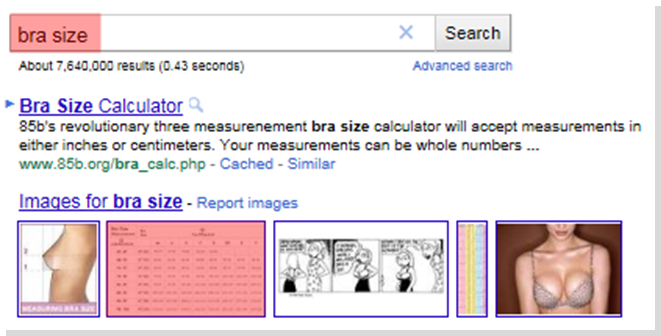
## Are there any cross digital impacts? ##

Image XML Sitemaps

Image Sitemaps are a Google-supported extension to the existing Sitemap protocol which allows site owners to submit image information to Google.

Example of images within Google’s search results:

## REPLACE WITH CLIENT-SPECIFIC IMAGE ##



The Image Sitemaps format works by embedding image data within a regular XML Sitemap, and images are then associated with each web page included within the Sitemap.

By creating an Image XML Sitemap, images on the ##clientname## site will be discovered more quickly. There is also an added benefit that images which are behind “spider-traps” such as JavaScript forms can also be discovered. Meta data associated with each image will also help Google identify those images as relevant to search queries.

### Solution

Include image meta tags within all XML Sitemaps. Each page should feature an image tag specifying the information about all of the important images which are relevant to that page (excluding irrelevant and boilerplate images). A maximum of 1,000 images can be associated with each web page.

We recommend making use of all optional meta data which is relevant to each image. Optional meta data includes: the image title; a caption for the image; the geographic location depicted within the image; and licensing information.

The image XML Sitemap should be referenced with the site’s robots.txt file and submitted to the major search engines (for example via Webmaster Tools).

### Alternative Solution

Add Image Sitemap elements only to a limited number of pages, within a static Sitemap file.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Breadcrumb Trails Issue

Breadcrumbs provide a useful secondary navigational feature to users, allowing them to see where they are within a site and enabling them to easily navigate to higher points within a website’s hierarchy. On large hierarchical sites they can be invaluable, with benefits such as improved interaction and reduced bounce rates.

Additionally, Google often uses breadcrumb trails as a basis for a feature called “site hierarchies”.

Example snippet without site hierarchies (source: Google)



Example snippet with site hierarchies (source: Google)



These result snippets have the benefit of presenting the user with additional sections of the site to navigate to directly.

### Solution

##Implement a breadcrumb trail.## Use the following markup to delineate breadcrumb trails on the site.

##SUGGEST USING AN EXAMPLE ON THE CLIENT’S WEBSITE INSTEAD##

Here's an example of a short HTML block showing a basic breadcrumb trail:

<a href="http://www.example.com/dresses">Dresses</a> ›

<a href="http://www.example.com/dresses/real">Real Dresses</a> ›

Real Green Dresses

Here is the same HTML marked up with structured data:

##use the microdata example if the site is HTML5, otherwise use the RDFa example##

##RDFa example##

<div xmlns:v="http://rdf.data-vocabulary.org/#">

<span typeof="v:Breadcrumb">

<a href="http://www.example.com/dresses" rel="v:url" property="v:title">

Dresses

</a> ›

</span>

<span typeof="v:Breadcrumb">

<a href="http://www.example.com/dresses/real" rel="v:url" property="v:title">

Real Dresses

</a> ›

</span>

<span>

<strong>Real Green Dresses</strong>

</span>

</div>

##HTML5 microdata example – needs to switch to Schema.org example ideally##

<div itemscope itemtype="http://data-vocabulary.org/Breadcrumb">

<a href="http://www.example.com/dresses" itemprop="url">

<span itemprop="title">Dresses</span>

</a> ›

</div>

<div itemscope itemtype="http://data-vocabulary.org/Breadcrumb">

<a href="http://www.example.com/dresses/real" itemprop="url">

<span itemprop="title">Real Dresses</span>

</a> ›

</div>

<div>

<strong>Real Green Dresses</strong>

</div>

The key elements in the above code are:

* Utilises Google’s structured data markup for breadcrumb trails
* Hierarchical links which match the structure of the site
* A link to the homepage
* A reference to the page that the site is on, rather than a link. An inline <strong> tag has also been added in order to emphasise the keyword in the breadcrumb relating to the current page.

It should be noted that whilst arranging a website’s breadcrumb trails in a suitable format is a required step in order for a site to obtain site hierarchies, this does not guarantee that the site’s listings in Google will automatically feature site hierarchies, as they are only provided at Google’s discretion

Following this advice will, however, provide the best chance of Google picking up this information to better understand the structure of the ##clientname## site and also to achieve site hierarchy links.

#### Cross Digital Impacts

## Are there any cross digital impacts? ##

Google’s help documentation for breadcrumb trails and site hierarchies:

[www.google.com/support/webmasters/bin/answer.py?&answer=185417](http://www.google.com/support/webmasters/bin/answer.py?&answer=185417)

Structured Markup

## THERE ARE 3 DIFFERENT TYPES OF MARKUP - YOU NEED TO CHECK THAT THEY ARE NOT USING ANY OF THEM – ALSO CHECK THAT THEY DON’T HAVE RICH SNIPPETS ALREADY ##

## NOTE – MARKUP FOR VIDEOS IS COVERED IN ANOTHER SECTION ##

Some search engines support enhancements of their search results snippets based on the use of structured markup on web pages.

Example of a Rich Snippet *(source: Google)*



Enhanced snippets such as this catch users’ attention better than conventional snippets, which typically results in increased click-through rates. A study by Yahoo! showed a 15% increase in CTR for its enhanced snippets.

### Solution

We recommend adding structured HTML to websites in formats supported by Google.

Examples where structured HTML can be used:

## INCLUDE AN EXAMPLE OF EACH DATA TYPE THAT THEY CAN USE – LIST OF SUPPORTED TYPES IS HERE <http://goo.gl/3Sjtd> ##

There are currently three competing formats which can be used to mark up structured data: microformats; RDFa; and HTML5 microdata. We recommend using ##microdata if their website is HTML5, otherwise RDFa/microformats##.

It is important to note that, whilst adding structured HTML is a required step in order for a site to obtain Rich Snippets, it does not guarantee that a site’s listings in Google will automatically feature Rich Snippets, as they are currently only provided to sites at Google’s discretion.

Google’s documentation for Rich Snippets:

[www.google.com/support/webmasters/bin/topic.py?topic=21997](http://www.google.com/support/webmasters/bin/topic.py?topic=21997)

Video Structured Markup

## THERE ARE 3 DIFFERENT TYPES OF MARKUP, SCHEMA.ORG, SEARCHMONKEY RDFa AND FACEBOOK SHARE - YOU NEED TO CHECK THAT THEY ARE NOT USING EITHER OF THEM – SEE [www.google.com/support/webmasters/bin/answer.py?answer=162163](http://www.google.com/support/webmasters/bin/answer.py?answer=162163) AND CHECK IF THEY HAVE THE ENTRIES FLAGGED AS ‘REQUIRED’ FOR EITHER FORMAT ##

Video markup formats are structured formats which help video search engines to identify your video content more easily. Videos on the #clientname# site do not use these formats:

## EXAMPLE VIDEOS ##

There are several benefits to using video markup formats.

Firstly, some search engines (notably Google) use structured markup formats to help them to find videos for inclusion within their indices. Marking up your videos using these formats can help these search engines to index your videos. For example, with Google, these videos can appear within both its dedicated video search engine and its main web search results via Universal Search.

Secondly, proper optimisation of the meta data within this video markup makes it easier for search engines to identify the subject matter of the video and therefore increases the chances of it being found for relevant searches.

There are currently several competing on-page formats which can be used to mark up videos: Schema.org, Facebook Share and SearchMonkey RDFa.

## IF VIDEOS DON’T HAVE THEIR OWN PAGES ##

An additional issue on the site is that individual videos do not have their own pages, and are instead grouped together. This is suboptimal – just as each product on a website should have its own main target page, the same is also true for videos. Grouping multiple videos on a single page makes it harder to optimise each video properly.

### Solution

Implement a video markup format on all videos. Include as much meta data as possible for each video, including relevant target keywords, in order to increase the chances of videos being found for relevant searches.

Ensure that each video file has its own web page, rather than including multiple videos on a single page. In addition to including video markup formats, this page should be properly optimised, as a normal web page would be.

Google’s documentation for Schema.org video markup is here:

<http://support.google.com/webmasters/bin/answer.py?hl=en&answer=2413309>

Google’s documentation for other video markup formats is here:

[www.google.com/support/webmasters/bin/answer.py?hl=en&answer=162163](http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=162163)

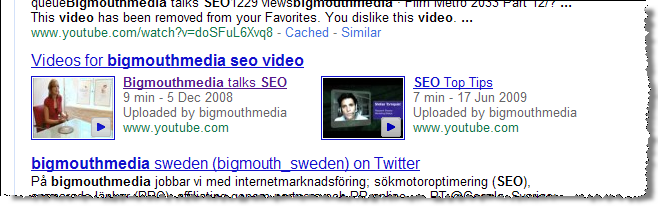
We also recommend using an XML feed-based format for submitting videos, in addition to the HTML-based format. Meta data should be consistent between these two formats.

Video Feeds

Video feeds are a very effective mechanism for syndicating videos to search engines and other third-party services. Video feeds enable search engines to discover videos on websites more easily and to index new videos more quickly. The meta data which can be embedded within video feeds also provides search engines with valuable data about the content of each video.

##consider replacing example below with an example which is relevant to the client##

Example of videos within Google’s search results:



### Solution

There are two major formats for video feeds, Video Sitemaps and MediaRSS. Google supports both of these formats, but some other search engines such as Bing only support MediaRSS. As it is possible to include all information supported by Google within a MediaRSS feed, we generally recommend implementing MediaRSS to enable you to submit your videos to a wider variety of sites.

* MediaRSS specification: <http://video.search.yahoo.com/mrss>

Include as much video meta data as possible, as the more data provided to search engines, the higher the chance that they will consider your videos to be relevant to user searches. Useful meta data includes the title, description, keywords, category, duration, and a suitable thumbnail picture. Dublin Core metadata can also be attached to MediaRSS feeds, and Bing has provided extensions to the MediaRSS format to provide additional metadata.

The feeds should be dynamically generated and the system should be set up to automatically ‘ping’ search engines each time the contents of the feeds changes.

* The specification for pinging search engines to inform them that an XML Sitemap has been updated is located here: <http://sitemaps.org/protocol.php#submit_ping>

The feeds should also be submitted to a variety of video search engines. We recommend submitting to as many popular video search engines as possible, to ensure that videos can be found by users of these platforms. Once a feed has been submitted to a search engine, no additional work is typically required beyond maintaining the feed. LBi can provide ##clientname## assistance with submitting video feeds.

It is also important to ensure that the feeds validate. This is more important with feeds than is generally the case with websites (although this can be important as well). Feed validation for MediaRSS feeds can be performed on a number of sites, such as <http://feedvalidator.org/> or <http://validator.w3.org/feed/>.

Consider using a feed traffic analysis service such as FeedBurner on MediaRSS feeds. If the feed or feeds use tracking URLs, ensure that these are search-friendly.

The video feed should be referenced with the site’s robots.txt file and submitted to the major search engines.

We also recommend using an HTML markup format such as Facebook Share or RDFa on web pages which embed videos, in addition to an XML feed-based format. Meta data should be consistent across these two formats.

### Alternative solution

Create a Video Sitemap instead of a MediaRSS feed, and submit it just to Google.

Non- hierarchical URL structure

The ##clientname## site is currently not arranged in a ‘hierarchical’ fashion, relying heavily on dynamic URL parameters to create different pages.

This is not search engine friendly for a number of reasons. Ideally, a keyword-driven hierarchy should clearly delineate site sections and sub-sections, as keywords and site hierarchy are used in search engines for ranking and displaying sites in results.

Improvements could also be made in order to make these URLs more useful for human visitors. As URLs are generally read from left to right, a human visitor can ascertain where they are in a keyword-driven hierarchical site from looking at the URL.

Implementing these changes would also help potential visitors to understand a URL appearing in the search results pages, which can in turn increase click through rates.

Here are some examples of the way in which the site is currently structured:

* ##Homepage URL##
* ##Selected poor URL 1##
* ##Selected poor URL 2 (within the URL 1 subdirectory)##

These URLs are:

* The homepage of the site
* ##description of poor URL 1##
* ##description of poor URL 2##

### Solution

Structure the site hierarchically. For example, the above URLs could be structured along the lines of:

* ##Homepage URL##
* ##Optimised version of URL 1##
* ##Optimised version of URL 2##

This should be implemented consistently across the site. Adopting this structure will allow ##clientname## to improve the synergy of the main keyword highlighting areas. Meta data, copy, anchor text, Title and heading tags can be synergised with the current location in the directory. If this is achieved site-wide, the majority of ##clientname## web pages will be sending a strong signal of their topic to search engines.

Additionally, the directory structure of a website typically emphasises the importance of web pages in a tiered manner, e.g. top tier web pages will tend to receive more exposure than the next tier and so on. Adopting the directory structure proposed, ##clientname## will be in a better position to achieve top-level promotion and use the new hierarchy to more effectively pass PageRank value to deeper sections of the website.

It is vital that 301 redirects are implemented on all old URLs to redirect visitors to the new location.

#### Alternative Solution

Redeveloping core areas of the website may not be possible due to restrictions of time or cost. If this is the case, we suggest that you re-write existing URLs. URL re-writing is the process of converting dynamic versions of a URL into a static alternative.

URL Duplicate Content and Brand Threat Issue

Certain URL formats on the ##clientname## website, for example ##type of URL here##, are accessible using any keywords, for example this URL:

##example URL with e.g. a product ID##

Is also accessible under a number of variants, for example:

##example URL with the same ID but different text##

This may cause significant duplicate content issues if erroneous URLs are indexed by search engines and, more seriously, presents the potential for a brand-damaging event.

For example, it is possible for a user to create a ##clientname## URL with a negative brand message, which may then be shared and subsequently indexed by search engines, for example:

##example URL with the same ID and brand-damaging words##

An issue of this nature occurred on the Independent, with 65 fake URLs indexed in Google and one attracting nearly 1500 tweets, attracting significant negative exposure. <http://goo.gl/WCAIg>

### Solution

When a product page is fetched, check that the URL matches the expected name for the page ID in question. If it does not, perform a 301 redirect to the correct version.

This will resolve the issue both for search and for users.

### Alternative Solution

Implement the canonical link element instead. In this instance, as the page loads, check for the correct URL for the page based on the ID and create a canonical link element with the correct product URL.

This will mostly resolve the issue for search – however, please note that the canonical link element is only treated as a “hint” by search engines, and there is no guarantee that it would be followed. In practice, we would expect to see many (but not all) URLs resolve correctly.

Additionally, this also does not resolve the brand issues described, as users will still be able to create URLs with negative brand messages and share them with others (e.g. via social media).

Broken Links Issue

A number of pages on the site contain ‘broken’ links which lead to invalid pages.

Here are a few examples:

## EXAMPLES – A SCREENSHOT FROM XENU MIGHT LOOK NICE ##

Broken links are not just detrimental to search performance: users, as well as search engine spiders, will encounter the ‘page not found’ error, which is bad for the brand experience.

### Solution

Find all broken links and either correct or, where appropriate, remove them. ##A list of errors has been supplied along with this document.##

One way in which to find broken links is to examine server log files with an automated process to identify URLs listed as 404 errors which have an ‘HTTP Referer’ located on the same domain.

Another technique for finding broken links is to use a site spidering tool, such as the free Xenu’s Link Sleuth: <http://home.snafu.de/tilman/xenulink.html>

Google Webmaster Tools can also be used to locate broken links. <https://www.google.com/webmasters/tools/>

LBi can provide you with an up-to-date list of broken links at any time.

Pages Missing Analytics Tags

The ##clientname## website uses an analytics service, ##Google Analytics##, to track visitors as they enter, explore and leave the website. This tracking is performed via the inclusion of a small piece of tracking code inserted within each web page on the site.

Some pages, however, are missing this analytics tracking code. Users who visit these pages will not be identified and tracked by the analytics service – this will result in gaps in the visitor tracking, and may mean that ##clientname## is potentially underestimating the site’s current performance.

Examples of pages lacking tracking tags:

* ## example ##
* ## example ##

## IF THEIR 404 PAGE ALSO LACKS TRACKING ##

The site’s 404 Not Found error page also lacks this tracking code. This means that the number of visits to pages which do not exist are also not being tracked correctly.

Example 404 page without tracking:

* ## example ##

### Solution

Ensure that the analytics tracking code is correctly implemented on all pages on the website.

#### Cross Digital Impacts

If analytics tracking is incomplete this may impact a number of other areas. The tracking of paid search performance may be affected and affiliate referrals may not be fully measured.

RSS Feeds Issue

## THIS SECTION APPLIES TO NEWS, BLOGS, PRESS RELEASES, AND SIMILAR REGULARLY UPDATED CONTENT ##

## IF THEY HAVE A FEED BUT NO <LINK> ELEMENTS MENTION THAT INSTEAD ##

Using web feeds is a very effective method of enabling frequently updated information to be discovered and indexed by search engines rapidly. The use of web feeds also enables this information to be found in a variety of blog and feed search engines, such as Google Blog Search or Technorati.

Although the site features a news/press releases section, it does not use web feeds (such as RSS or Atom).

URL:

## LIST EXAMPLE SECTIONS HERE – IF THERE ARE SEVERAL LIST AT LEAST A FEW DIFFERENT ONES ##

### Solution

Provide RSS and/or Atom feeds (using recent versions of these formats) for the most recent ##press releases, news, articles and other regularly updated areas of the site##.

It is important to ensure that the feeds validate. This is more important with feeds than is generally the case with websites. Feed validation can be performed at this site: http://feedvalidator.org/

Wherever feed data is embedded within a page, it is also usually suitable to link to the relevant RSS feed. Include <link> elements with appropriate titles within all pages that link to the feed so that, should a visitor be using a more modern web browser that can identify the presence of feeds, they can easily determine which feed they are subscribing to. Some search engines such as Google also use these <link> elements to discover feeds.

#### Example

<link rel="alternate" type="application/rss+xml" title="##SUITABLE TITLE##" href="http://##URL OF RSS FEED##" />

Also, provide a visible feed button which links to the feed, as some users will look for these when attempting to subscribe to feeds. We recommend using a recognisable feed icon such as the one on the left.

The feeds should be dynamically generated and the system should be set up to automatically ‘ping’ one or more different ‘pinging’ services which are appropriate for the content as each new article is created.

We also recommend submitting these feeds within search engine webmaster consoles.

## Content “Below The Fold” Issue

##clientname## displays a large portion of the main page content below the fold of the page. The majority of the initial page real estate is taken up by ##DESCRIBE e.g. an excessive amount of advertisements and/or an excessive amount of additional links which are not part of the main navigation and/or another situation##.

#### Example

##SCREENSHOT OF CLIENT’S SITE##

##You could also use <http://browsersize.googlelabs.com/>##

When a user navigates to a web page with a lot of visible content obscured below the fold of the page it can result in a poor experience, as the user is not immediately presented with what they were initially searching for. As well as being a poor user experience, this can negatively affect the site’s ranking in search engines.

SpeechMarkOpen.png

*If you click on a website and the part of the website you see first either doesn’t have a lot of visible content above-the-fold or dedicates a large fraction of the site’s initial screen real estate to ads, that’s not a very good user experience. Such sites may not rank as highly going forward.*

Source: Google

SpeechMarkClose.png

#### Solution

Ensure the primary real estate on each web page (areas “above the fold” of the page) include content relevant to the user and the main purpose of that page. For example, if the web page is an informational page about “##client specific example##” then the user should be presented with information relating to that subject, without having to navigate or scroll to another area of the page.

There should also only be a limited amount of ad copy, if any, placed above the fold of the page.

##any additional client specific suggestions##

#### Cross Digital Impacts

Usability plays an important role in how users interact with a site. Improving the overall layout of a web page, and ensuring relevant content is easy for users to find, will improve their experience of the site.

LBi can provide additional support with usability issues where required.