

## Search engine optimisation

Version:	1.0
Date created:	13/01/2010
Policy official:	Peter Jordan
Date last updated:	22/01/2010
Date issued:	February 2010
Lead official:	David Pullinger
Guidance number:	TG123

### Purpose

This guidance is an introduction to search engine optimisation aimed at UK public sector website managers, web content managers and editors and metadata managers.

## Table of contents

<b>Background .....</b>	<b>2</b>
Audience .....	2
Introduction .....	2
<b>Structuring your website for findability .....</b>	<b>3</b>
Indexing.....	3
XML Sitemaps.....	3
Robots.txt .....	4
<b>Understanding what users are looking for .....</b>	<b>5</b>
Identifying search terms and keywords .....	5
Baselining search performance .....	7
<b>Improving your website content .....</b>	<b>8</b>
Page titles and meta-descriptions .....	8
URLs .....	9
Headings and navigation .....	9
Content.....	9
Images and other non-text content.....	10
Video & SEO .....	11
Outbound and internal links .....	11
<b>Increasing your search engine ranking .....</b>	<b>12</b>
Inbound links .....	12
SEO and the semantic web .....	13
SEO and personalisation .....	14

## Background

### Audience

This document is primarily aimed at UK public sector website managers, web content managers and editors and metadata managers.

It should also provide informative reading for UK public sector Heads of e-Communications, marketing and communications managers.

### Introduction

Websites tend to be designed on the assumption that users will enter via the homepage and then navigate through the website – but the reality is that most users will come to your website via a Web search engine (e.g. Google, Yahoo!, Bing etc.).

There are a number of website content and design tactics, known as search engine optimisation (SEO), which can help ensure that your website can be found by Web search engines. If your website is well designed, users can go to pages deep within your website to find the content they need and may leave again having never seen or used the homepage.

SEO includes making sure the website is well-structured and accessible, content is unique and relevant, and that there are links to your website from other websites, indicating that your website is a source of trusted and valuable content.

These and other tactics should be considered both in day to day website management (for instance, adjusting content in response to observing common search referrals in your web analytics) and as part of projects (e.g. website development or designing campaigns in other channels, such as print or TV).

There are four main parts to SEO:

- 1) Making sure your website is structured for indexing by search engines
- 2) Understand what your users are looking for
- 3) Using the right words so your content appears in search engine results lists and people recognise it when they see it
- 4) Taking steps to increase your search engine ranking so you place higher in search engine results lists

In addition to following good SEO practice, it is important to make sure that your website design and navigation works regardless of where people arrive in your website.

## Structuring your website for findability

1. Web search engines such as Google, Yahoo! and Bing use automated software called 'crawlers', 'spiders' or 'robots' to trawl the Internet looking for content to index. Typically these crawlers are quite passive; they simply follow hyperlinks. Most of the time this works well, but there are a growing number of cases where crawlers fail to find and index important online content.
2. Good website design and correct use of robots.txt files and XML Sitemaps will ensure that your website can be accessed and crawled by search engines.

## Indexing

3. Web search engines find it difficult to index some types of content, including content that is:
  - inside images (e.g. images of text);
  - inside databases, usually only accessible via a search form;
  - linked to through dynamic code, e.g. Javascript, Ajax, Flash;
  - obscured in some way, eg. through redirection, framesets, iFrames;
  - presented dynamically, especially using multiple, variable parameters;
  - restricted by user authentication, eg. CAPTCHA, log-in or user session;
  - not linked to at all.
4. This un-indexed content is sometimes referred to as the 'invisible', 'deep' or 'hidden' web. Since standard search engines have difficulty finding and indexing it, users will also struggle to access such information.
5. To ensure your website is crawled and indexed by search engines:
  - Provide alternative text for images
  - Avoid having critical content and navigation inside images or Flash movies
  - Ensure that your website can be navigated without Javascript enabled
  - Include an XML Sitemap

## XML Sitemaps

6. XML Sitemaps, now supported by all the main search engines as an industry-wide practice, are a supplementary way of listing the content available on a website in a simple, open and "crawler-friendly" format. XML Sitemaps do not replace existing crawl-based mechanisms, but they enable crawlers to do a better job of indexing a website.
7. Every UK public sector website must have one or more XML Sitemaps (conforming to the Sitemaps Protocol) which cover all the publicly available

content under any relevant domain or sub-domain (possibly including databases). These sitemaps should be exposed through the robots.txt file and submitted to internet search engines and other approved crawlers.

8. More information about sitemaps can be found in: [TG122 Exposing your website to search engines](#).

## **Robots.txt**

9. A 'robots.txt' file gives instructions to search engine crawlers about what, if any, part of a website can be accessed, crawled and indexed. The robots.txt file should be placed in the root directory and compliant crawlers will take notice of its instructions.
10. Check that the file is not preventing compliant crawlers from accessing important parts of your site or conversely exposing parts of your site that are not for public.
11. Remember that some crawlers may ignore the robots.txt file - especially malware robots that scan the web for security vulnerabilities etc. The file is also easily readable, so anyone can examine it. So use more secure methods to protect sensitive information.
12. Remember to point to your site's XML Sitemap from the "robots.txt" file. The following sample robots.txt file excludes several directories and points to the sitemap.

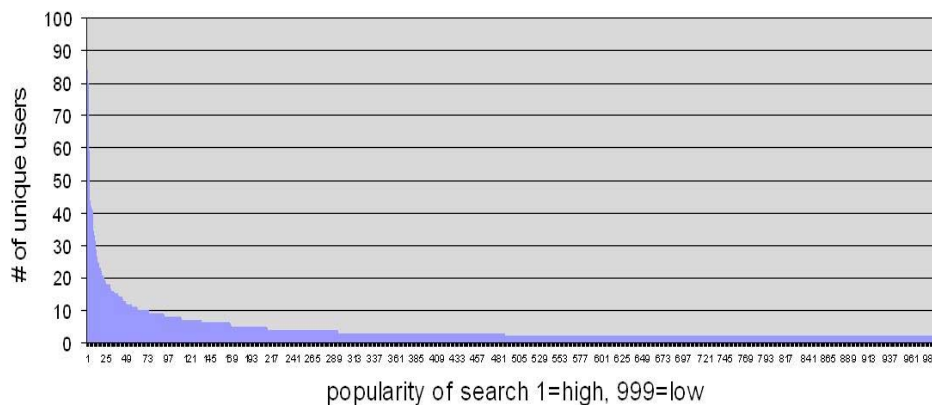
```
User-agent: *
Disallow: /cgi-bin/
Disallow: /latest-updates/
Disallow: /a-z/
Sitemap: http://www.jusice.gov.uk/sitemap.xml
```

## Understanding what users are looking for

13. The first part of implementing an SEO strategy to improve the findability of your website content should be to undertake research to understand:
- What people are looking for?
  - How findable is your content currently?

## Identifying search terms and keywords

14. Analysis of data about search provides extremely valuable insight into your site users and what they are looking for. It is a direct interaction with customers and helps you understand what visitors want from and expect to find on your site.
15. More detailed analysis of customer journeys will help you to relate search terms to specific pieces of content. This provides further insight into how language is used by people to describe their particular informational needs.
16. Search queries normally show a Zipf curve distribution, known as the 'long tail'. Only a small number of queries are searched for a lot, with a large proportion of queries searched for only a few times making up the bulk of search queries



17. Data on search terms comes from two sources: referrals from external search engines (keywords) and internal website search (search terms). Both can be obtained from web analytics packages. You should also monitor which search engines are referring traffic to your website and how much of your traffic comes from particular search referrals.

**Keywords Searched**

OGE Reports: available from 2/22/2004 to 10/23/2004

User Type: All By: Weekly Week: 10/17/2004 - 10/23/2004 Top: 100

Show only keywords with at least 1 unique users.

Search for keywords containing the following text:  Refresh

Rank	Keyword	Unique Users	Visits	Page Views	Visits Per User	Page Views Per Visit
1	Council Tax	302	308	420	1.0	1.4
2	jobs	192	198	290	1.0	1.5
3	car tax	150	150	193	1.0	1.3
4	education	143	144	179	1.0	1.2
5	passport	128	128	172	1.0	1.3
6	MP EXPENSES	110	111	175	1.0	1.6
7	expenses	106	106	164	1.0	1.5
8	health and safety	106	106	166	1.0	1.6
9	maternity leave	105	108	154	1.0	1.4
10	child benefit	102	105	151	1.0	1.4
11	paternity leave	100	101	147	1.0	1.5
12	DVLA	93	93	140	1.0	1.5
13	bullying	92	92	175	1.0	1.9
14	housing	91	92	138	1.0	1.5
15	housing benefit	91	91	136	1.0	1.5

18. Once you have compiled a list of search terms and keywords, analyse the top 50 – 200 top search terms by grouping them together into categories, looking for misspellings and alternate terms, to build up a true picture of relative popularity.
19. Over time, this should become the list of words and phrases that are important to your users, which you can incorporate into your content management design.
20. Some useful questions to ask:
  - What are the most frequent search queries (both referrals from external search engines and internal site search)?
  - Which popular searches bring back zero results?
  - Which search queries bring back irrelevant or misleading results?
  - Which search queries bring back popular content?
21. In addition to gathering data about what keywords are currently being used, there are tools designed to help you analyse other related terms being used on external search engines, that are not currently being directed to your website. Google Adwords and Search-based Keyword Tool are just some of the tools that can help you identify other keywords you could be using to attract target users onto your website.
22. Other sources of user research (surveys, focus groups, etc) may also provide insight into other words being used by your audience.
23. Good practice:
  - Prioritise the high volume, most important terms for SEO

- Look for patterns in the 'shoulder' of lower volume terms, to identify any other high value content for SEO
- Look outside your website, to identify opportunities to provide new content or to improve SEO for your existing content
- Feed insights about users choice of words and phrases back to content providers and communicators in your organisation

## **Baselining search performance**

24. By comparing your list of popular search terms with analysis of how effective they currently are, you can create a list of content areas to prioritise for SEO. Once you have a list of keywords, you can start to benchmark visibility by looking at how those terms are currently performing.
25. Search for each of the keywords and phrases on your list, primarily in the external search engines and also in your internal website search (look at your web logs to identify the main referring search engines for your website)
26. Check the rankings for each keyword or phrase in search results
  - In web search, are your pages at the top/near the top of results?
  - Is your content overlapping with other public sector websites, creating confusion for potential users?
  - In site search, are the 'right' pages at the top of results?
27. Check whether the queries bring back good quality results – clear titles and descriptions that give a good 'scent of information' for the user.
28. Check click-through rates for your content pages - how much traffic pages are receiving from internal website search and external search engines?
29. Do some search keywords and phrases have null visibility?
  - This might indicate you have no content that covers these terms
  - Or that your content is poorly optimised for terms – i.e. your content does not use the titles and words that your users do
  - Or that appropriate content is located on a different public sector site
30. Check the bounce rate of your search results page to identify where users are so discouraged by the results that they leave your website immediately.
31. Check how much of your site is indexed.
  - Are web search engines indexing all the content you would like?
  - Does your site search index all the content it should?

## Improving your website content

32. Once you have an understanding of the keywords and phrases your users are employing and how well your site is indexed by site search and web search engines, the next step is to develop and implement an SEO strategy that builds on what you have learned and focuses on your business priorities.
33. Starting with the most popular but least performing content areas, work on improving metadata, headings, content and URLs and monitor before and after performance, to show improvement and identify what changes have the most effect (so you can apply those lessons to the next most popular set of content).
34. Content improvements have two functions:
  - 1) to increase the likelihood of your content appearing for particular search terms; and
  - 2) to ensure that your users recognise the right content when they see it.

## Page titles and meta-descriptions

35. Page titles and meta-descriptions are used by search engines in results displays.
36. An informative page title, that incorporates user focused words and phrases, is more likely to be clicked on by a user. Titles should be unique for each page and should provide a stand-alone summary of the content of that page.
37. Remember, only the first eight or so words will appear in the results so make the first words count. Put keywords as close to the start of the title as possible.
38. Where possible, avoid the use of Stop Words in page titles. Stop Words are frequently used words commonly ignored by Web search engines including prepositions and conjunctions (e.g. of, in, at, for, and etc.)
39. A page title should be concise, informative, keyword rich and give people a good idea of what to expect from the page – it may be the only thing they read – and hence help them to decide if it is appropriate.
40. Meta-descriptions are also important. Although web search engines usually generate a “snippet” which includes the keywords searched for, if the description is relevant it is often displayed. A description is *not* an introduction – try to summarise the content of the page and value to the user of clicking through to it. Include important keywords.

41. To fine tune your metadata, look at what is currently displayed in the search results page for the search engine you are targeting and use the 'view source' function in your browser to see whether the search engine is picking up your metadata or some other type of content.
42. Meta keywords can be added for your pages; they may help fine-tune your site search results but web search engines generally ignore them due to abuse by search engine spammers.

## **URLs**

43. When your website is listed in the results returned by a search engine, it is important that people recognise your site and that it would be helpful for them to click on it.
44. The URL used should indicate the purpose or owner of the website.
45. Where possible, use human readable URLs containing real words. URLs that are obscure or dynamically generated may appear complex to users and difficult for search engines to index.

## **Headings and navigation**

46. Search engines use complex algorithms for relevancy that will decide whether your website appears high in the results list or not. They take information inherent in website structure and content, to interpret which words are 'important' relative to other words on the page.
47. Correct use and mark up of headings (<H1>, <H2> etc.) and emphasis (<em>, <strong> etc.) assist search engines to understand the structure of your content and identify the most relevant parts.
48. Headings should be informative and include, where possible, the words and phrases you know your users search for, as headings are likely to be weighted slightly higher, in terms of relevance, when your website is indexed.
49. Consider also how well your navigation structure works for users who arrive deep into your website via a search engine. Will they be able to find their way around your website from there?

## **Content**

50. Use the language of your audience to ensure that content will be returned in search results for those keywords (identified in the analysis described earlier).

51. Content should be unique to your website and thoughtfully written. Write for humans – do not add too many or poorly matched keywords to your page in the hope of attracting search engines. “Keyword stuffing” – will reduce the effectiveness of your keyword strategy overall.
52. If you know that users are familiar with your brand or campaign messages, make sure they are incorporated. If you think users are familiar with outdated brands or terminology (e.g. ‘dole’), try to incorporate them into content to ensure the correct content is returned in search terms (e.g. ‘This benefit was formerly known as \_\_\_\_\_’).
53. Long paragraphs of text and overuse of official government words and descriptions may make it difficult for users to decide whether the information is relevant to them and may result in a ‘bounce’ or one-page visit.
54. Information from search term and keyword analysis will also help you to structure the content on the page, so that popular content is placed in the first place people look on the screen. It is not enough to simply get users to your site. If they cannot recognise the information they need on the page, they will look elsewhere for something that meets their expectations. Well written content is succinct, easy to read and broken into small, digestible chunks of information.

### **Images and other non-text content**

55. The primary purpose of the ‘alt’ attribute for images and other non-text content is to provide information for automated screen readers so that the user understands what is being shown on the screen. It is also a very useful way of ensuring that your non-text content can be found by search engines.
56. A search engine will build up a picture of content by looking at the alt text and page elements around the image or (other non-text content), such as content and headings. Alt text, headings, captions and other descriptive elements should be meaningful and use the words your users expect.
57. Most search engines offer an image search facility and increasingly display images in the main search results.
58. Alternative text should be clear, easy to read and relevant.
59. Alternative text is not required if normal text already conveys the information provided by the image or for any images that are purely decorative - for example the rounded corners of boxes.
60. Be succinct and focused and don’t use too many keywords.

## Video & SEO

61. Video is becoming an increasingly important content format. Users may expect to find it and it can also benefit your site's visibility as search engines start to include video in the main 'blended' search results.
62. If video content is a traffic driver for your site, consider hosting the video on your site. If you wish the video to be viewed as widely as possible, consider also hosting it on public video sites. Web search engines are generally unable to index the content of the video, so focus on the title, description and consider hosting a transcript which can be indexed.
63. One major search engine also supports Video Sitemaps – an extension to the Sitemap Protocol.

## Outbound and internal links

64. Search engines pay a lot of attention to links. Indeed, Google's unique insight was to pay attention to the volume, relevancy and quality of links to a page in determining the relevancy and authority of the content.
65. You can control the text used in internal and outbound links to other sites. The text tells users and search engines very useful information about the linked to URL and also provides information about the scope and nature of your website through building a picture of related information and, in the case of internal links, the natural hubs within your website.
66. Use link text that is:
  - clear
  - easy to read
  - relevant
  - concise
67. Provide a good 'scent of information' – what is the destination content about? Avoid using link text like 'more' or 'click here' or using too many keywords.

## Increasing your search engine ranking

68. Web search engines use complex, secret algorithms to determine the formulas for ranking and relevancy that decide whether your website appears high in the results list or not.
69. Search engines have an ongoing battle with website owners who try to exploit how the algorithms work, in order to artificially improve their website ranking. Good SEO is not about trickery and short term wins – such approaches are the resort of the traffic generating websites which have minimal value to users.
70. For government websites, which have a legitimate purpose and audience, most SEO activity will focus on good website design and content management, as described above. Increasing the quality of the website and developing unique and trusted content is the best approach for increasing website ranking in search engines.

## Inbound links

71. One of the best indications a search engine has that a website's content is trusted is that other trusted websites link to it. Building effective links to your site is a key way to boost your site's visibility, particularly when links come from other 'authoritative' sites (websites which are respected and offer high value e.g. other government sites, education sites and news sites).
72. It is good practice to check which sites are currently linking to yours, for example by typing, "link:www.sitename.gov.uk" into web search engines.
73. Many links to public sector information will occur naturally as other websites, blogs and social media link to your content. Also, content on the gov.uk domain has a high degree of trust associated with it.
74. Once you have built up a system of links into your website, avoid putting this at risk through unnecessary website structure or domain changes. These can damage ranking in the short term, while the website reputation is rebuilt.
75. If you are making large scale changes to the structure of your website (or are converging content into another website) it is advisable to alert key stakeholders ahead of time and ask them to link directly to the new location, in addition to setting up temporary and/or permanent redirects.
76. Ensure you use the correct type of redirect to tell search engines whether your content has moved or is no longer available:
  - 301: content has moved
  - 410: content no longer available

77. Key pages should have persistent URLs. A persistent URL will point to key content even if the content has moved. For example, [www.coi.gov.uk/webguidelines](http://www.coi.gov.uk/webguidelines) will always point to the web guidelines home page, even if the underlying architecture of the site is changed.
78. Use web analytics tools to identify where users are following wrong or out-of-date links from other websites. If you detect that another website is sending users to the wrong place on your website, send the correct URL to web master of that site and ask them to replace the faulty link with the right one.
79. Although there are systems for alerting search engines to changes of domain, a change of domain is damaging in the short term and should never be undertaken unless absolutely necessary. An existing domain should be used for new functionality or content wherever possible, to minimise the time taken to build up the 'reputation' and ranking of the website.
80. Identify key URLs for documents and other content that others may wish to link to. Each important document or service should have its own unique URL.
81. Identify key stakeholders and communities of interest that will find your site relevant to their needs. Suggest they link not just to the home page, but to relevant deeper content
82. If you reorganise your site, advise key stakeholders about new URLs and ask them to link directly to the new location, rather than just relying on redirects.
83. Engage with blogs/social media and RSS. Remember that people can pick up information from your site via blogs, micro-blogging services and RSS feeds. Links back to your site improve SEO.
84. Ensure relevant URLs are included in press releases and other communications.
85. Consider providing a 'link to this page' function to enable users to easily bookmark your website using social bookmarking tools such as Digg or Delicious.

## **SEO and the semantic web**

86. Whilst using extremely complex algorithms, search engines have not had any true intelligence until recently. They succeeded by clever processing of text matches, rather than understanding concepts. However, increasing numbers of websites are using semantic web technologies such as RDFa and microformats. These are open standards that add machine-readable information to existing HTML markup, so that browsers and search engines get useful hints about the structure and meaning of the content.

87. Search engines are starting to use this information to add value to users' searches by adding additional information about for example a restaurant or book – number of reviews, user rating, author etc.
88. Government Departments are starting to use RDFa to mark up public sector consultations and Directgov is able to use this data to show the publisher and start and end dates of consultations.

**Legal aid: funding reforms**

A consultation seeking views on proposals to reform the fees paid to litigators and advocates in publicly-funded criminal cases and measures to reduce spending on experts' fees in all legal aid cases.

Ministry of Justice

Starting Date: 20/08/2009 Closing Date: 12/11/2009

[www.justice.gov.uk/consultations/legal-aid-funding-reforms.htm](http://www.justice.gov.uk/consultations/legal-aid-funding-reforms.htm)

**Freedom of Information Act 2000: designation of additional public authorities**

The aim of the consultation document is to seek views as to whether the Government should look to use the powers under Section 5 of the Freedom of Information Act and extend the coverage of the Act and if so, which organisations it should consider.

Ministry of Justice

Starting Date: 25/10/2007 Closing Date: 01/02/2008

[www.justice.gov.uk/consultations/cp2707.htm](http://www.justice.gov.uk/consultations/cp2707.htm)

More information [on structuring information on the Web for re-usability](#) is available online.

89. Whilst it is early days, the semantic web will increasingly allow search engines to bring together more extensive and richer information from a variety of sources and users will increasingly expect this and expect to get data within their search results rather than just a link to a page.

## SEO and personalisation

90. Search engines are increasingly offering the opportunity to personalise results, by for example promoting results you particularly like. As more users choose to personalise their results, it will be harder to know what a page's 'average' ranking is for all users. Whilst it may be high for the generic results, it may be lower for some users who have personalised results.
91. The good practice described in this guide still holds good, but you may need to undertake deeper analytics – how much traffic are you getting to a target page for each key word or phrase and looking at competitors' content: How does it rank and why? Is there a specific context or meaning to a phrase in a public sector context that you can use to advantage?

END