

Web Friendly Text

Why should text be ‘Web friendly’?

Whenever or whatever you write, you’re expecting it to be read by a human. Regardless of language and format (like Braille), ultimately you believe that a human will read it.

When you write for a webpage this isn’t the whole story. You will have two sorts of visitors: human and software. Software frequently visits your site: in fact it’s such an important category of visitor that you may not get any human visitors unless you cater for the software ones.

We’ll look at what a software visitor is in a moment.

Why human readers need special attention

Copywriting is a craft skill under any circumstances. As far as humans are concerned, though, you might think that ‘reading is reading’. Why do you have to take a particular approach because you’ve written a Web page?

Consider: how do you feel about reading from a monitor? If you think about it, many people complain about reading on monitors. We seem to find it more challenging than printed matter.

Frequently we don’t actually *read* a webpage at all: we merely *scan* it. [1] So when we write a webpage, we need to make it easy to scan, that is, make the crucial points stand out.

Visitors may spend only a few seconds on any given page and may not even start to read a long paragraph. [1]

General compositional principles

Keep your sentences short, in everyday language, informative and direct. Use the form of language that’s appropriate to your likely readership: if you’re managing a website promoting activities in Britain, spell check for British English. Either avoid or explain jargon.

Try to reduce every paragraph to one concept.

Overall, reduce your word count as far as possible: you should consider 500 words per page a reasonable limit. If you have a long piece — lets say you’re publishing a transcribed talk or an archive edition of a printed article — either edit it for the medium or break it down into discrete, linked pages. Avoid one long, scrolling page.

Given that people seem less willing to read a web page, *summarize your most*

important points in the opening paragraphs, since they are the most likely ones to receive a thorough reading. Elaborate on your key ideas in the rest of the page, breaking your topic down into sub-sections with *clear, descriptive headings*. Remember: your reader is probably just scanning the page for words that leap out, so headings are a crucial way to emphasize primary ideas.

Examine a wikipedia page [1]. Notice that each entry cross-links to other pages heavily. Also notice that the words that make up a hyperlink stand out. Consider your topic in terms of a group of interlinked pages that can be read in any order (either in serial or at random). Each page should focus on a specific point within a topic: when you've covered your point sufficiently, make it easy for people to find the next topic to move on to.

Don't assume that visitors will enter your site at any particular page (even the homepage), or that they will explore it in a linear order. A reader might have been using Google to search for a term that appears in a page in the middle of a section of your site. Once they've finished the page they arrived on, give them a reason to explore further: you can even slightly 'tease' the reader by hinting at more interest elsewhere.

Above all write every page to be an 'authority' on it's topic. That doesn't mean it has to be long and exhaustive — accurately summarizing the bullet points will do. Your in-page links can add to the sense of authority if they always pass the reader on to a page that expands on a point. For example if you write a page on the history of domestic animals, every time you mention a new animal—say, horses—link to a page specifically about it. That page could link on to further pages about specific breeds of horse.

Software visitors: Robots and Spiders

Software visitors are programs written purely to examine your webpage. They are usually referred to as either 'Robots' or 'Spiders' (because they follow the links on the Web). Robots are the front-line tool in the process of creating the index that makes up a Search Engine like Google.

Believe it or not an awful lot of text on the Web is genuinely aimed at non-humans: spam is written to grab the attention of a Robot, with the aim of rigging a favourable page ranking. Spamming is an incredibly tempting approach to Search Engine Marketing, but it's deceitful and should be scrupulously avoided, if no other reason than that it's unethical. Very few people are capable of effectively tricking a Search Engine algorithm. Programmers at companies like Google play cat-and-mouse with spammers on a daily basis. When you get caught out your site will be penalized.

Grabbing the attention of a Robot is actually a good thing, but we can do this very well by playing fair.

Writing for a new website

The content of a website is what visitors arrive for, not the layout. Layout is an important support to the content, but gets a disproportionate amount of the development attention. *An ugly website hasn't necessarily failed*, but a website where the visitor can't find key information has.

You should really have your text content resolved before your layout designer starts work. If you haven't, then it's likely that there's been insufficient preparation.

Start by creating a list that covers every topic and subtopic. This is a dummy website www.animals.net:

```
-----  
First level pages      | Second Level pages  | Third Level pages  
=====
```

Domestic pets	Dogs	Dachshund
		Lurcher
	Cats	Persian
		Devon rex
Farm animals	Cows	Freisan
		Angus
	Pigs	Large white
		Large Black

Keep this list to only three levels if possible. This list should become your site's navigation. Make sure every page in your site appears in this it.

Go through each of the pages in this list and compose the following:

- A document title
- HTML page title
- Meta description and meta keywords
- A summary opening paragraph

Google Search Engine Returns: a brief interlude

This is a sample Google return. Notice it has three basic sections:

[Search engine optimization - Wikipedia, the free encyclopedia](#)

Search engine optimization (SEO) is the process of improving the volume and quality of traffic to a web site from search engines via "natural" ("organic" or ...

en.wikipedia.org/wiki/Search_engine_optimization - 94k -

[Cached](#) - [Similar pages](#) -

The blue coloured, underlined phrase is the HTML page title. The black coloured line is *often* either the first 20-30 words of text in the page, but it *can* be the content of your meta description tag. The last, green, line is the URL (Web address) of the page. Note that, on a live search engine return page (SERP) the words you searched for will be rendered in bold, indicating that that's partly why Google chose to return that page as a result.

An intelligent document title

We don't always have complete control over how our documents are named. Some models of website software (e.g. Content management Systems) force a label for the page that is ostensibly gibberish. This is a pity, but not a disaster. Search engines can still evaluate these pages—we've just lost an opportunity to feed them some clues.

What we'd like is for each page to have a keyword friendly name: e.g. domestic-pets.html *not* dompet.html. Google recognizes the hyphen as a word separator, but can't make any sense of compound contractions.

This logic applies to directories (folders) too. So:

www.animals.net/domestic-pets/

and

www.animals.net/farm-animals/

not

www.animals.net/dompets/

or

www.animals.net/frmanmls/

Thus, a good URL would be

www.animals.net/domestic-pets/dachshund-breeds.html

and a bad one

www.animals.net/dompets/daxies.html

HTML page title

Great care should be taken over this, because it *is* used to evaluate the content of a page. Keywords and phrases are a must, but full sentences are unnecessary. You can separate 'clauses' with either a vertical bar | a double colon :: or a simple hyphen with a result something like:

Animals | Domestic Pets | Grooming Dogs

or

Animals :: Farm Animals :: Pigs' Diet

Meta tags

Meta information is ‘information about information’. Meta tags appear in the <head> section of a webpage and are usually invisible to page visitors. There are several types, but only two are pertinent here. They are used by some search engines, but not all. Google’s Webmaster Tools certainly reports on the existence and quality of both meta description and meta keyword tags, so it’s advisable to treat each with due care.

Meta description

Each page should have its own short description of no more than thirty words. This is sometime used as a summary of the page in search engine returns

Meta keywords

Meta keywords are a collection of the words that best summarise the content of the page.

Traditionally these have been seen as the crucial way to inform a Search Engine. They were abused heavily in the early days of search engine optimization, so they are of dubious value in modern websites. However their status may change and Google’s Webmaster Tools does remark on their existence. So it’s *probably* worth putting effort into composing them. In a sense this is work you have already done: you should already have listed the words and phrases that most summarize your site and page; you may as well collate them in the document itself.

A keyword list should be composed like this:

“dogs,cats,friesan cows,farm animals”

notice that phrases are separated by a comma with no space after it.

Loads of links

Yes, getting links is a good thing. But like everything else I’ve said above, it’s quality not quantity.

If I bother to link from a page on my site to a page on yours, the logic goes that I think your page is important for some reason. Thus a link to your webpage implies a ‘vote’ for it. Obviously the more people that link to your page the greater the implication that it’s significant in some way.

This doesn't mean that your site will immediately come out on top in a search. The Search Engine still has to work out what you're significant *for*. Buddhism? Footwear? Sheep?

Consider an example. Lets say we all think that Wildmind is a great place to learn meditation. We create a page of links with a list of sites we'd like to recommend:

www.wildmind.org

So Wildmind gets a vote. But a vote for what? How is Google to work out *why* we think it's important? The domain name is poetic, so doesn't help all that much. We could give Google a hand by wrapping the link around a phrase someone might actually search for:

[Buddhist meditation](#)

Or even better, weave the recommendation into some instructive text:

Learn [Buddhist meditation](#) online with Wildmind.

By doing this we can legitimately influence search returns. The important point in all of this, though, is that context is critical. One should be as specific as possible: ideally a page listing beginners' meditation classes links directly to pages with dates and times of classes, not just to a homepage.

The page a link is placed in needs to say something coherent about *why* the link is there. That's why categorization is also important, it emphasizes context.

Since acquiring links is important, give other website owners a reason to link to you: as I said above, make sure each page is authoritative on it's subject. This benefits you, the other author and the visitor.

Notes

[1]

Nielsen, Jakob. "On the average Web page, users have time to read at most 28% of the words during an average visit; 20% is more likely." <http://www.useit.com/alertbox/percent-text-read.html>

[2]

<http://en.wikipedia.org>

Further reading

Kissane, Erin. "Writing Content that Works for a Living", A List Apart Issue 271, November 04, 2008 <http://www.alistapart.com/articles/writingcontentthatworksforaliving>

Nielsen, Jakob; “How Little Do Users Read?”, *Alertbox*,
<http://www.useit.com/alertbox/percent-text-read.html>, (May 6, 2008)

Nielsen, Jakob; “Long vs. Short Articles as Content Strategy”, *Alertbox*,
<http://www.useit.com/alertbox/content-strategy.html>, (November 12, 2007)

Nielsen, Jakob; *Designing Web Usability*, ch. 3 “Content design”, (New Riders, 2000)

Google Webmaster Guidelines:

<http://www.google.com/support/webmasters/bin/answer.py?answer=35769>