

PDF Search Engine Optimisation?

Adobe PDF is a proprietary format developed by well Adobe but is based on Post Script it is a very popular document format. It is also a binary blob, making it a pain for search engines to crawl and rank.

So why am I interested in PDFs?

Well we know like Word documents that all the major engines have a good stab at crawling and ranking them so can we help Google and co along? More importantly could a link in a PDF be followed by the crawler? If its followed will it be weighted?

So this is one of several experiments that I'm running to see how much weight in a PDF and what techniques can be used to optimise PDFs for the big wide web.

PDFs and weighted links

Just like a html document a PDF can have a hyperlink and unlike say Flash or Javascript its not being hidden behind code so potentially it could pass PageRank now that would be interesting wouldn't it?

Let's try a couple of things here like linking to my [top secret page](#) that no one knows about? But will the search engines now crawl it?

We also want to see if it can pass weight to something we already have ranking, so are you interested in [becoming a SEO](#)?

Descriptions and Ranking Factors?

So presuming our links have passed weight how do we improve the rankings of a PDF after all its not like a HTML document with heading tags and all the general SEO goodies, you have headings in your document editor but with PDF this is not a tag rather a font and size so it carries no additional weight.

Meta Information could be one way, back in [2004 Kevin over at Search Engine Guide](#) was looking for various factors that effected PDF SEO he found the title and Author fields in the PDFs meta data was indexed along with the Subject meta data being used for description. So lets try to replicate this feat!

Future of PDFs

With Yahoo and Adobe being new best buddies expect the next few innovations in PDF search to come from Yahoo as opposed to Google in the mean time this could be a great way to grab those long tails. How the big search engines handle formats over then html seems to be changing more and more they are relying on specialist program specific factors which are then combined, and in particular Google seems intent on integrating all these into its main index Google Universal.