

Optimizing PDFs for Search Engines 2008

Contents

Introduction	2
Better rankings & click-throughs	3
Make sure your PDF includes mostly formatable text	3
How to tell if the text is formatable text -	4
Optimize the text in the document	5
Include keyword-rich selling message in the Document Title	6
What happens when you don't create a Document Title	6
How to update the Document Title.....	6
Link to the PDF	9
Increase web site visits from a PDF (Humans & Search Engines)	9
Add contact info	9
Include active hyperlinks to your web site	9
Active Links when a PDF is viewed as HTML.....	12
How to convert URL's into active hyperlinks in a PDF	13
How to turn any text or image into an active link	14
PageRank and PDF's	16

Introduction

Search Engine Optimization (SEO) For PDF's

Acrobat Reader Files (PDF) do get indexed by search engines and they can do well in the search engine results.

PDF = Portable Document Format.

In this document we'll cover -

- **More Click-Throughs & Higher Rankings:** Tips to better rankings & click-throughs from search engine result listings for your PDF's.
- **Increase web site visits from a PDF: Both from Humans & Search Engines**

Get more people to visit your web site or contact you: Including active links within PDF's to increase the number of readers who visit your web site or contact you.

Content Discovery by Search Engines - Active links within a PDF give search engines paths to find content on your web site and may improve search engine indexing & rankings.

NOTE - You are free to republish this PDF or pass it along as long as you retain the PDF in its entirety with all its active hyperlinks.

by:

George Aspland
eVision, LLC
179 East Main Street
Branford, Connecticut USA 06405
203 481.8005 • sales@eVisionSEM.com • www.eVisionSEM.com

Better rankings & click-throughs

Make sure your PDF includes mostly formatable text

Search engines can't read text that's part of an image (You already knew this. It's the same issue with HTML web pages)

Here's an example of PDF file that was indexed in Google. It's a special issue from Scientific American that was put up on a government web site (It has since been moved or deleted).

Whoever put this PDF up scanned the entire issue in from a hardcopy (This is not unusual. We see many PDF's that are scanned from hardcopy).

Here's a screenshot of a partial page from this nine page PDF:



*A PDF indexed in Google.
The entire document was scanned in from a hardcopy*

While this is an extreme example of including all the text in a document within images it will help emphasize the issues of not using formatable text (defined below).

Here are some of the problems when you scan a document from hardcopy:

- The entire PDF is made up of solid, one page images so it's huge, many times larger than it needs to be.
- Because the text is scanned it's very hard to read, almost illegible.
- Most importantly, **search engines can not read the text within this PDF**

Below was a search engine listing for this PDF in Google:

PDF] [pdf file](#)

File Format: PDF/Adobe Acrobat - [View as HTML](#)

Page 1. Page 2. Page 3. Page 4. Page 5. Page 6. Page 7. Page 8. Page 9.

[snap.lbl.gov/news/ScientificAmerica_special_Issue_0904.pdf](#) - [Similar pages](#)

Notice the description Google created for this file, "Page 1. Page 2. Page 3. Page 4. Page 5. Page 6. Page 7. Page 8. Page 9." This happened because there was no text in the PDF that Google could read. Google could not read the text in the PDF so it could not create a more meaningful description.

This listing is not going to help "entice" people to click-through to this document and it will not help this document rank better unless somebody is searching for "Page 1". or "Page 2". etc (I think I'm the only one who does that!)

Compare the above listing to a Google search results listing for a PDF from our site for a search on "improve conversion rates":

[\[PDF\] Improving Conversion Rates with Conversion Point Architecture Design](#)

File Format: PDF/Adobe Acrobat - [View as HTML](#)

"How can **Conversion** Point Architecture. Design. increase my leads. and sales?" Successful websites are all about **improving** your **conversion rate**. ...

[www.evisionsem.com/documents/Conversion_Point_Architecture_Design.pdf](#) - [Similar pages](#)

Google created the listing (duplicated below) from the text contained within the PDF by creating snippets from phrases it found within the document that contains the words and/or the phrases being searched on (just like it does with HTML pages):

"How can **Conversion** Point Architecture. Design. increase my leads. and sales?" Successful websites are all about **improving** your **conversion rate**."

Whenever possible make sure your PDF is created using a method that keeps most of the text as regular formatable text so that the search engines can read and index it.

How to tell if the text is formatable text -

You should be able to tell what text in your pdf is formatable text by selecting some amount of the text (the method to do this will vary depending on the application and version of the application you're using to read the PDF). If you can select sections of text, such as individual words or

phrases, and copy and paste this text into another application such as a text editor, then it is formatable text.

Another way to tell if the text in a PDF is formatable text is to Export it if your application offers this feature. If you can export the text into a format that allows you to use the text in other application and reformat this text, then it's formatable text. For example if you can export the text into an .rtf (Rich Text Format) file then the text is formatable text.

If it's important that your PDF does well in search results you should consider rebuilding the PDF to ensure that it includes mostly formatable text. We've done this for clients who have been given the rights to reprint articles etc but the articles were only available in hardcopy. We scan or export all the images and rebuild the article using formatable text.

Optimize the text in the document

If you are writing or editing the document that will become an online PDF you should optimize it just as you would an HTML web page and include important keyword phrases in the body and headlines etc.

Pay specialize attention to the first Headline as it may be extracted by search engines to be used in the search engine listings (more on this later in the section on Document Titles)

Below, again, is the listing for a PDF from our site for a Google search on "improve conversion rates"

[\[PDF\] Improving Conversion Rates with Conversion Point Architecture Design](#)
File Format: PDF/Adobe Acrobat - [View as HTML](#)
"How can **Conversion** Point Architecture. Design. increase my leads. and sales?" Successful websites are all about **improving** your **conversion rate**. ...
www.evisionsem.com/documents/Conversion_Point_Architecture_Design.pdf - [Similar pages](#)

The occurrences of the phrases "improve/improving conversion rates" as well as sub phrases such as "conversion rates" and the individual words "improving", "conversion" and "rates" throughout this document undoubtedly helped the listing reach a higher position in the search results for this search.

Also notice how Google extracted snippets of text that included the phrase, sub phrases, and words being searched on to build this description for the search listing (just like Google does for HTML pages). Google also highlighted the words being searched on within the description –

"How can **Conversion** Point Architecture. Design. increase my leads. and sales?"
Successful websites are all about **improving** your **conversion rate**.

A description that includes enticing phrases with the keywords being searched on highlighted should encourage more people to click through to the PDF.

Include keyword-rich selling message in the Document Title

Updating the Document Title in a PDF is as important as the HTML Page Title is in an HTML web page. Most search engines will include the PDF's document title as the headline/link for the search listing.

Include a good keyword-rich selling message as the document title just as you do for the HTML Page Title for HTML web pages.

What happens when you don't create a Document Title

If you don't take the time to create a Document Title one of two things will likely happen.

1. Sometimes document titles will be automatically generated by the application used to create the document. The application used to develop the document may extract some text from the document to use as the Document Title.

Don't let this happen. The document title the application generates may not do a good job of enticing people to click through to the PDF from the search result listings. In addition the document title the application generates may not help the PDF rank high for any meaningful search phrases.

2. If you don't take the time to create a Document Title the search engine will likely create link text for the search engine listing by extracting some text from the document. If you've done a good job of optimizing the document and created good keyword rich headlines this may be ok.

Here's an example of a search engine listing for a PDF that had no Document Title

[\[PDF\] VOLUME ONE](#)
File Format: PDF/Adobe Acrobat - [View as HTML](#)
Page 10 of 263 Vol. One - Public Report. MS. GRIMES: Yes. **10** Monegan transcript **page 24**, line 9. 11 Monegan transcript **page 32**, line 23. ...
media.adn.com/smedia/2008/10/10/16/Branchflowerreport.source.prod_affiliate.7.pdf - [Similar pages](#)

Search Listing for PDF that had no Document Title

There was no Document Title so Google extracted text from the document, the first line of text in this case, which was "Volume One".

The link text in this search listing headline, "Volume One", is not going to help entice people to click through to the document or help it rank for any meaningful keyword searches.

How to update the Document Title

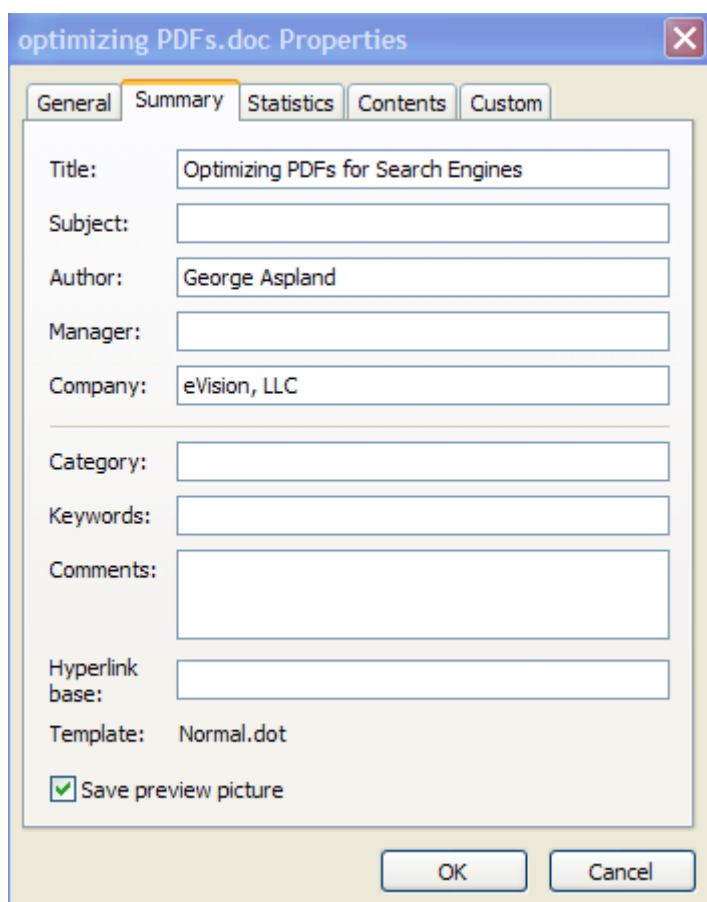
Access to the document title will depend on the tool used to create the PDF. Usually a window will be available from the application's toolbar under the File command.

Try looking for commands under "File" in the toolbar such as Document Settings, Document Summary, or Document Properties.

You can also update the Document Title if you have a version of Acrobat that allows you to modify the document properties, such as Acrobat Standard or Acrobat Professional (I have not seen a version of Acrobat Reader that allows you to update the document properties).

It's actually best to update the Document Title using a version of Acrobat that allows you to modify the document properties because many applications automatically modify the document title they create. Word 2003, for example, adds "Microsoft Word –" in front of the document title you create.

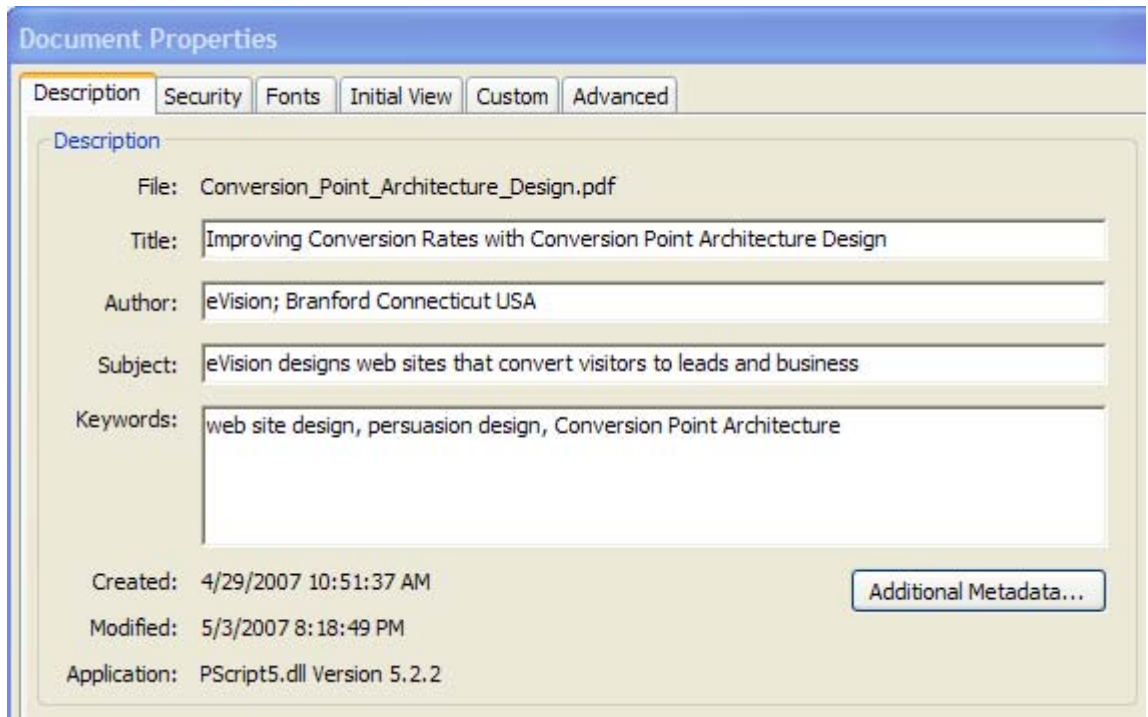
To modify the Document Title in Word 2003 select File / Properties.



To update the document Title in Word 2003: Files / Properties

In Acrobat (the development tool. Not the free reader) you can usually access the document title as follows:

- File > Document Info > General, or
- File > Document Properties > Description tab



To update the document Title in Acrobat Standard 7.0: File / Document Properties

Try to include a good keyword-rich selling message as the document title. Below is the Google search result listing for a PDF from our site for a search on “improve conversion rates” –

[\[PDF\] Improving Conversion Rates with Conversion Point Architecture Design](#)
File Format: PDF/Adobe Acrobat - [View as HTML](#)
“How can Conversion Point Architecture. Design. increase my leads. and sales?” Successful websites are all about **improving your conversion rate.** ...
www.evisionsem.com/documents/Conversion_Point_Architecture_Design.pdf - [Similar pages](#)

The Document Title is the highlighted headline/link in the search result listing:

[**Improving Conversion Rates with Conversion Point Architecture Design**](#)

Just as with HTML pages, because the phrase being searched on, in this case “improve conversion rates” is in the document title this listing likely reached a much higher position in the search results than it would have if the phrase was not in the title.

Also notice how Google highlights the words and phrases being searched on within the document title (as well as the file description it extracted from the text within the document).

The combination of a well written document title with the words and phrases being searched for highlighted should help entice people to click-through to look at the pdf document.

Link to the PDF

To get the PDF indexed in search engines add one or more links to it from pages on your web site, preferably on pages that are already indexed. That's all it should take to the PDF indexed.

Increase web site visits from a PDF (Humans & Search Engines)

Now you've got your PDF ranking better and people clicking through to it from search engine listings, let's look at how to improve the click-through rates for a PDF (that is getting people to contact you or go to your web site where they may convert into leads or sales).

Add contact info

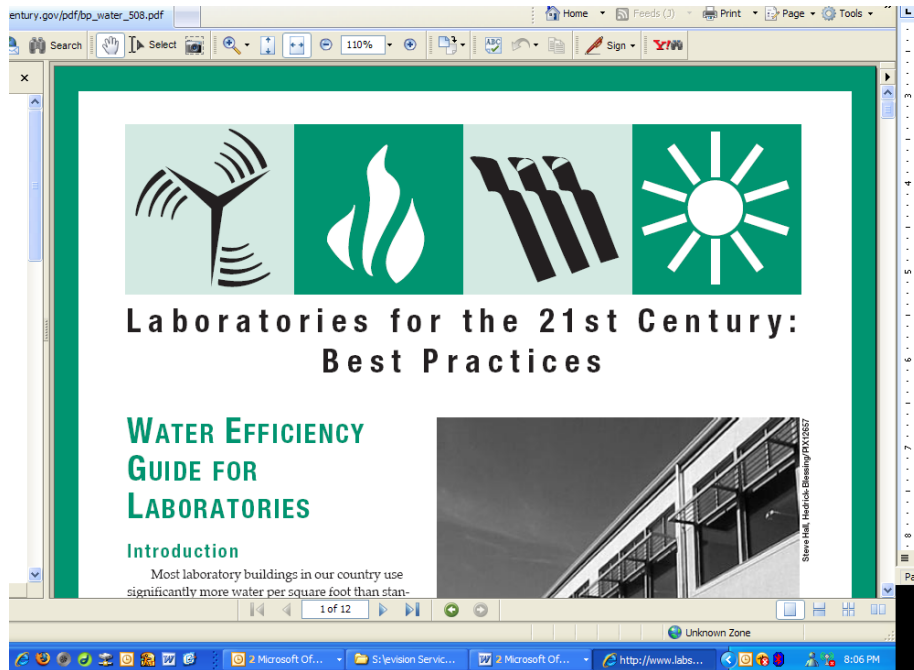
Make sure the document includes all your contact information including your web site address, email address etc, so that people will know how to reach you and your web site whether they read the PDF online or print it out.

Include active hyperlinks to your web site

Place a links(s) to your web site, products & services, and email links within the PDF so that people may contact your or click through to your site while looking at your PDF online.

Content Discovery - These active links within a PDF also give search engines paths to find content on your web site and may improve search rankings of the pages they point to. So include links in PDF's that you host on your web site and on PDF's you distribute to others such as datasheets, whitepapers, manuals and instructions, articles etc.

Here's a PDF being viewed within a browser with no active links to the web site anywhere within the PDF:



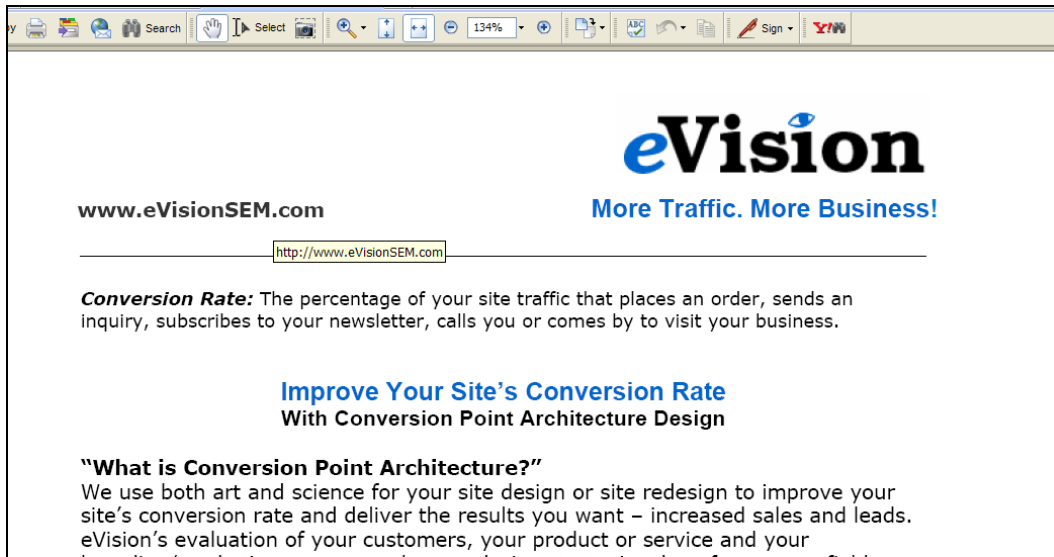
A pdf with no active links

There's no way for any one to click through to the web site if they view this PDF online. In fact the address to the web site isn't in the PDF at all (this is not unusual).

Here's a search engine listing for a PDF on our web site.

[\[PDF\] Improving Conversion Rates with Conversion Point Architecture Design](#)
File Format: PDF/Adobe Acrobat - [View as HTML](#)
"How can Conversion Point Architecture. Design. increase my leads. and sales?" Successful websites are all about **improving your conversion rate.** ...
www.evisionsem.com/documents/Conversion_Point_Architecture_Design.pdf - [Similar pages](#)

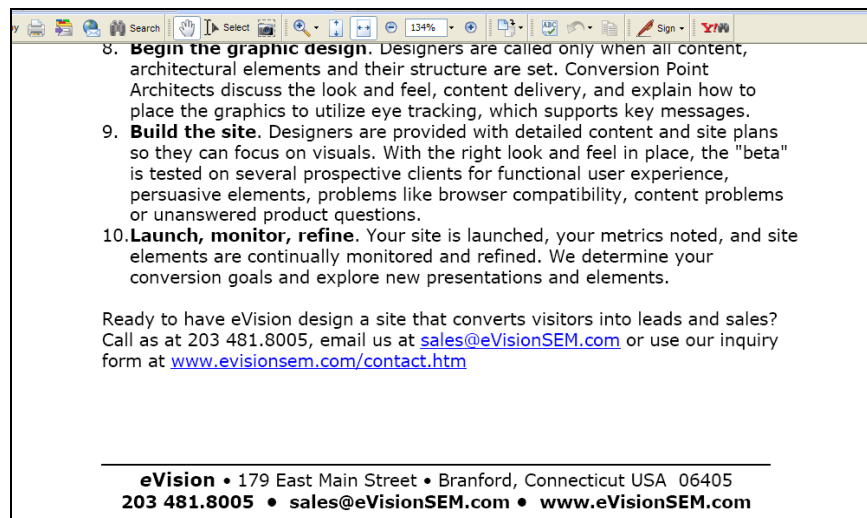
When searchers click through to the pdf from this listing most will see the following pdf in their browser:



A pdf with active links

Notice the hyperlink at the top of the page, www.eVisionSEM.com. It's an active link to our web site (The logo is also an active link to the web site. More on this later).

Below is a screen shot of the last page of this 2-page PDF:

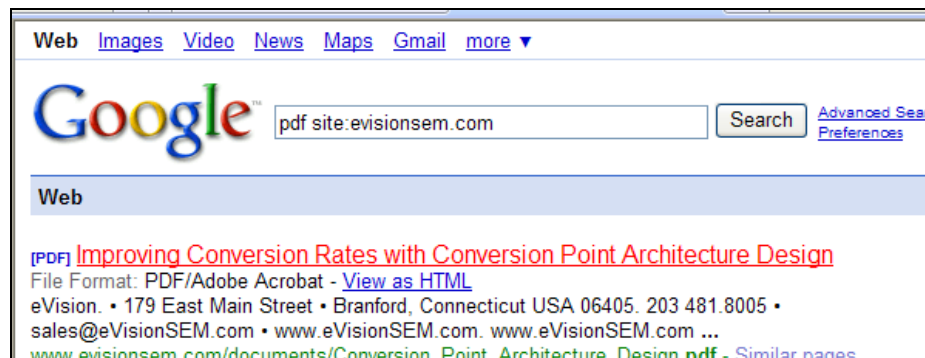


Active links throughout the PDF

The links to the web site and the email address at the bottom of the PDF and within the body of the pdf are all active links. This PDF includes plenty of opportunities for people to click through to the web site or send us an email.

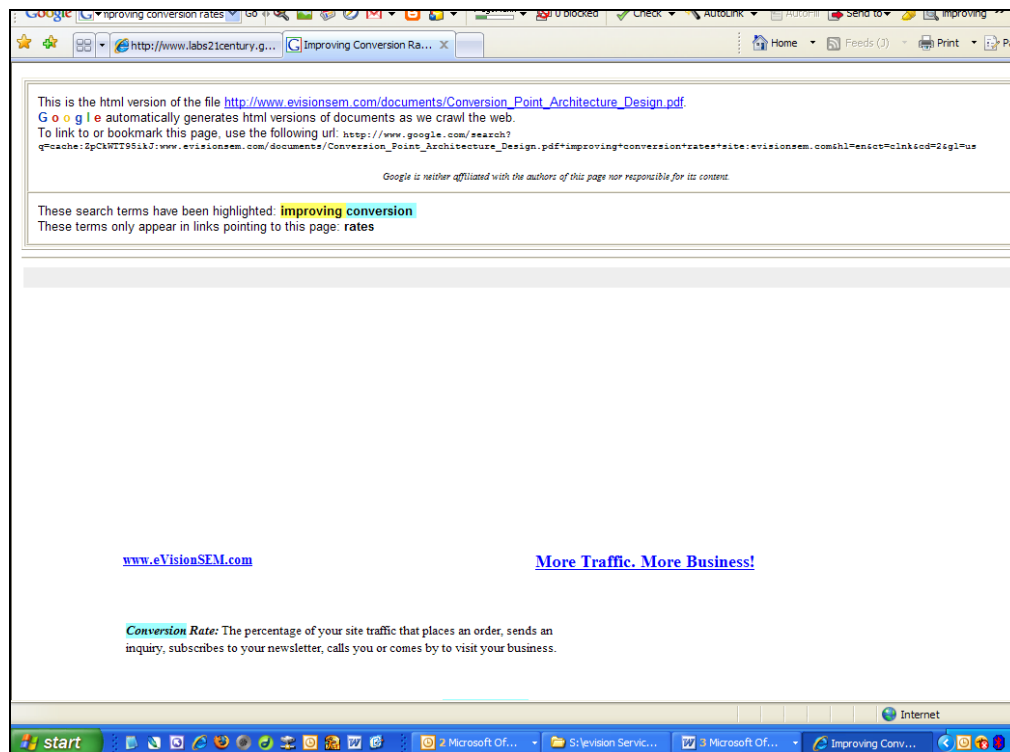
Active Links when a PDF is viewed as HTML

Most search listings for PDF's include an option to view the PDF as a web page. Here's a search engine listing for a PDF on our web site again. Notice the [View as HTML](#) link:



Option to View as HTML in Google SERP

I don't know how many people actually view pdf's this way. This feature does a terrible job, but here's the above PDF viewed as HTML:



A PDF Viewed as HTML in Google

Again when viewed as and HTML it's pretty horrible, but at least the embedded hyperlinks work and people will likely click through to the web site.

How to convert URL's into active hyperlinks in a PDF

The way to imbed active links within a PDF will vary depending on the tool used to create or modify the pdf. Some tools may allow you to specify that you want web addresses to be active in the pdf that is created. If the tool you're using does this, great.

However we'll make the assumption that you'll layout your document using a tool, such as Word, and save it as a pdf without the web addresses automatically being converted to active links. We'll show you'll how to turn the web addresses into active links using the Acrobat development tool (Acrobat Standard 7.0 in this case). Note - I've seen Word to Acrobat tools turn URL's into active links automatically at times, but often it does not or it doesn't do all the URL's.

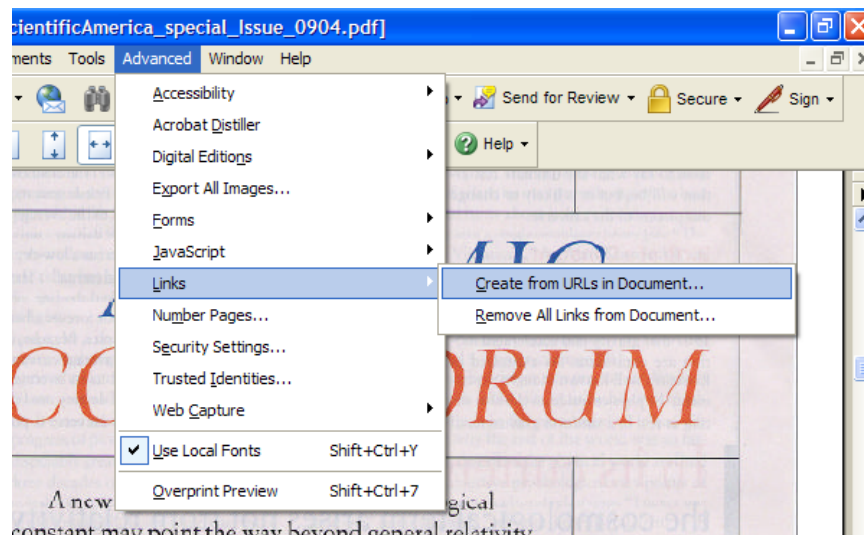
Tip: Highlight your Links

Acrobat 7.0 Standard will not change the text color to blue or add an underline when it converts the text URLs to active links.

As you lay out your document, highlight your web and email addresses so they will stand out as active links. Ideally use the expected blue color text and consider underling the text too, e.g. www.domain.com

Create links from URLs using Acrobat 7.0 Standard

1. Export your document as a PDF
2. Open the document with Acrobat
3. Choose Advanced > Links > Create from URLs in Document (see image below)
4. In the Create Links Dialog window either select All or enter a range of pages and click OK.



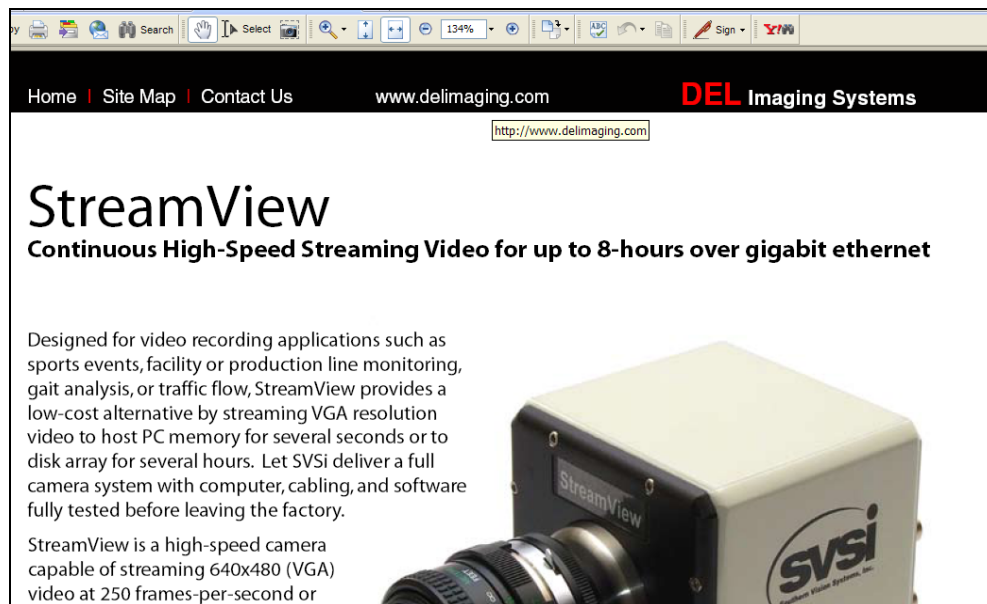
Choose Advanced > Links > Create from URLs in Document

That's it. Acrobat will convert all the web and email addresses it detects within the text of your PDF into Active hyperlinks.

How to turn any text or image into an active link

In addition to converting URLs to links you can turn just about any text or image in a pdf into active hyperlinks. This gives you the flexibility to turn any text within the body of your document into an active link. You can also incorporate a navigation system similar to that found on a web site allowing you to provide links to pages in a site without having to include long URL's within your PDF.

Below is a screenshot of the top portion of a PDF we created for a client:



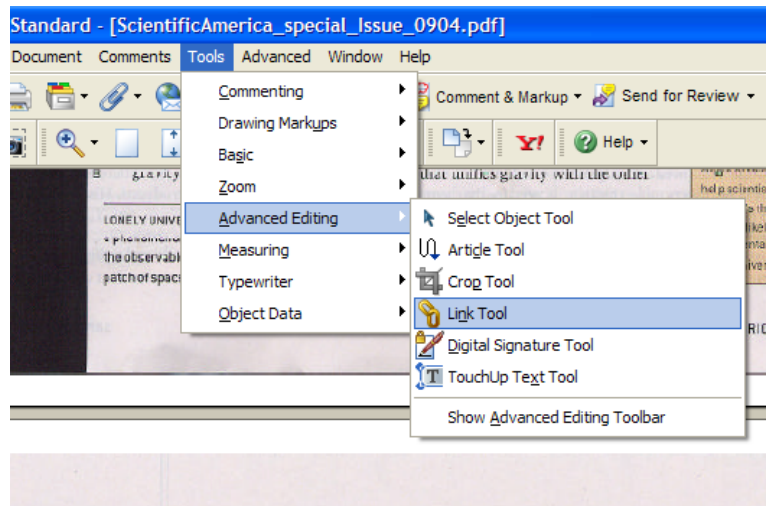
A PDF with text and image links

In addition to the active web address in the banner, the phrases “Home”, “Site Map”, and “Contact” are links to pages on the web site as is the logo in the banner. This banner looks very similar to the banner on the client’s web site.

Using Acrobat (Acrobat Standard 7.0 in this case) here’s one method to turn just about any text or image into an active hyperlink.

How to create text and image links using Acrobat 7.0 Standard

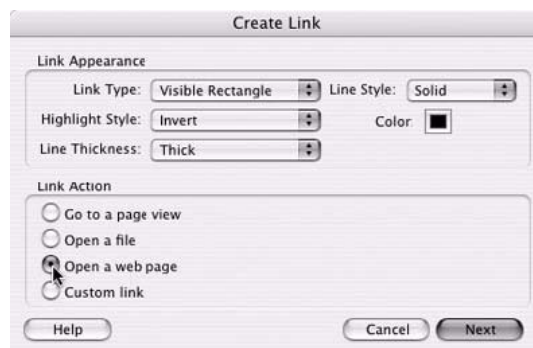
- 1 Go to the page in the PDF where you want to place the link.
- 2 Choose Tools > Advanced Editing > Advanced Editing Tools / Link Tool.



Choose Tools > Advanced Editing > Advanced Editing Tools / Link Tool

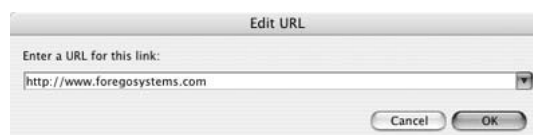
3 Drag a rectangle around the area you want to define as a link. When you release the mouse button, the Create Link dialog box appears.

4 Click the Open a Web Page radio button in the Create Link dialog box.



Link to the Web by choosing Open a Web Page button in the Create Link dialog box.

5 Click the Next button and the Edit URL dialog box appears. In the Address field, type the URL of the website to which you want to link and click OK.



Enter a URL for the link.

Tip: You can add parameters to your links

Consider adding unique URL's or adding tracking parameters to the URL's so you have more tracking options with your analytics.

PageRank and PDF's

The same link popularity analysis (Link Pop, PageRank, whatever you want to call it) appears to apply to PDF's as it applies to HTML pages.

For competitive keyword phrases you'll want to apply the same techniques as you would for a HTML web page such as including links to the PDF from other pages both on and off your web site .

Here's an example of a PDF that ranks high against lost of competition –

In a Google Search for “w4” or “w-4”, the listing below was the first listing in the search results from among over 30 million results. This is the IRS's PDF version of the W-4 form.

[PDF]

2008 Form W-4

File Format: PDF/Adobe Acrobat - [View as HTML](#)

Form W-4 (2008). Purpose. Complete Form W-4 so that your. employer can withhold the correct federal income. tax from your pay. Consider completing a new ...

www.irs.gov/pub/irs-pdf/fw4.pdf - [Similar pages](#) - [Note this](#)

Google shows about 500 pages from other web sites that link to this page. This page is probably reaching position 1 in this search, from among millions of competing web pages and files, because of the number of times “W-4” is mentioned in the pdf, the number and quality of the links pointing to it from other web pages, and due to the link popularity (trust, etc.) of the irs.gov sub web and the pages that link to this PDF from within the irs.gov sub web.

In our experience we don't see lots of PDF's reaching top positions for meaningful competitive searches. This likely has to do with the link popularity of most PDF's, they typically are not linked to as much as HTML pages. This is especially true of internal site linking. If you put some effort into “promoting” you PDF's, both from within your web site and from other web sites, you can get PDF's to reach high positions in the search results.

In summary, PDF's do well enough in search results especially for less competitive searches. And again, the links within the PDF's do get crawled which can get new content indexed and may help improve the rankings of the pages pointed to.

About George Aspland

George Aspland is founder and President of eVision a firm providing online marketing services to organizations around the world. eVision increases targeted traffic to its clients web sites while improving the ability of the web site to convert those visitors into inquiries, leads, and new business using its proprietary Conversion Point Architecture process.

Mr. Aspland speaks regularly on the subject of Search Engine Optimization (improving results from search engines) and has written a book, The Guide to Search Engine Marketing, a step-by-step guide to search engine marketing used by organizations who wish to optimize their own web sites.

Aspland also writes an online marketing blog/newsletter, which is sent to webmasters and marketing executives. Visit the blog to sign up for the newsletter, www.evisionsem.com/blog.

About eVision

Since 1998, eVision has been delivering dramatic, measurable results for companies and organizations, from small to large, around the world.

eVision increases targeted traffic to a web and dramatically improves the site's ability to convert those visitors into leads and sales **More Traffic. More Conversions. More Business!**

We drive targeted traffic to a web site through legitimate search engine optimization (SEO), pay-per-click (PPC) search engine marketing tactics and online promotion while working to ensure the web site turns those visitors into inquiries, qualified leads, and new business.

eVision uses its proprietary Conversion Point Architecture process to design user focused, search-engine friendly web sites that help convert visitors into customers. Conversion Point Architecture is a proven combination of art and science drawing from many fields including psychology, the usability field, web-buying behavior research and web technology and typically includes usability/conversion rate evaluations, target audience analysis, defining site objectives, user testing, advanced site analytics, complete content reviews, site architecture design, and site development.

eVision, LLC
179 East Main Street
Branford, Connecticut USA 06405
203 481.8005 • sales@eVisionSEM.com • www.eVisionSEM.com