

Get the most from Google

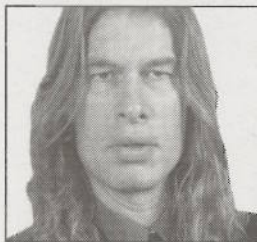
Search for digital objects, not just webpages, to save time and receive richer business results

A popular urban legend contends that peoples of the North have hundreds of words for the term "snow."

While there might not be hundreds, linguists have determined that Arctic cultures have certainly developed a more extensive and richer palette of words that describe different types of snow.

Just as snow surrounds and embeds Eskimo culture and its language, another thing is equally ubiquitous in our own North American business culture and vernacular. This thing has taken on so many variations in our society that we've developed dozens, perhaps hundreds, of terms to describe this thing as well. This thing is information.

Searching online is a case in point. When Google and other search engines index the Internet, they don't search for "information" per se. That's too generic a description and a technically inaccurate term. More precisely, they search for (and through) specific varieties of information. They scan digital objects – html files, ASCII text, xml files, rich text format documents, PowerPoint files, Word documents, Excel spreadsheets, pdf files, Shockwave Flash files, along with numerous image, video, audio files and other digital artifacts.



GARRETT WASNY Internet Intelligence

Herein lies a valuable search secret. Start looking at the web the way Google sees the cyber world – not as a morphing glob of nebulous information as we humans often (and wrongly) perceive it, but as a collection of distinct digital objects, each with its own unique information "snowflake" pattern and signature. Excel spreadsheets, for example, are just that: spreadsheets. If you're looking for numbers, doesn't it make sense to look specifically for Excel spreadsheet files? Need an overview of a subject in bullet-point format? Wouldn't you know it, but that's what most PowerPoint presentations are, so why not zero in on PowerPoint files in that search? Scanning for in-depth discussion papers? What a coincidence. Many pdf files are just that, so why not hunt specifically for pdf files when seeking detailed reports on a subject.

How is this done? Easy. Whenever you do a Google search, scan not only for key-

words, but for different digital objects or file types as well.

Say you wanted to hunt for information on **CHC Helicopter Corp.**, a Richmond-based company that's the world's largest provider of helicopter services to the global offshore oil and gas industry. Before we start, let's clarify by what we mean by "information." In this case, we'll define information as five types of digital objects: html files, pdf files, PowerPoint presentations, Excel spreadsheets and Word documents.

To start, we fire up Google and type "CHC Helicopter" with the quote marks (which tells Google to look for that specific phrase) in the search box.

For that term, Google returns "about 87,700" results. All the top 10 results are html pages and all highly relevant.

These include links to the CHC home page and press release pages, the CHC entry in Wikipedia, the online encyclopedia, along with profiles of

the company in *Globe Investor*, *Hoover's* and *Yahoo Finance*. This is an excellent start and many users would end their Google searching right here. They'd shift their focus to these first five links and explore them in more detail.

A more discerning web searcher would certainly flag these initial links and circle back later to explore them further, but they wouldn't be finished with Google just yet. They'd dig a little deeper using the Google file type command which allows a user to search for a specific file type.

To search only for pdf files, you'd type the following in the Google search box: "CHC Helicopter" filetype:pdf. This tells Google to scan for pdf files – and

Start looking at the web the way Google sees the cyber world – as a collection of distinct objects each with its own unique "snowflake" pattern and signature

only pdf files – that contain the term "CHC Helicopter." This pdf-only search yield 921 results, the top 10 of which are also highly relevant but completely different than the leading results for the term "CHC Helicopter" without the filetype command.

That's good. You're working Google from a different angle and uncovering new links and leads, but don't stop there. To search just for PowerPoint presentations only, you'd type "CHC Helicopter" filetype:ppt (ppt is the file extension for PowerPoint files). For Excel spreadsheets, type "CHC Helicopter" filetype:xls (xls is the file extension for Excel spreadsheets). For Word documents only: "CHC Helicopter" filetype:doc (doc is the file extension for Word spreadsheets).

You'll discover in each of these searches, the leading results are all highly relevant, but totally different from each other, even though you're using the same keywords ("CHC Helicopter").

Reason: when you scan by file type, you rearrange the results in a fresh way and bring forward digital objects to the top that are normally buried deep down in the stack.

Here's where it gets really interesting: substitute your own keywords in this search string and start incorporating the file type command in all your searches. Guaranteed, when you cycle through different file types – html, pdf, PowerPoint, Excel, and Word – you'll uncover new and highly relevant results that were always there, but were previously buried too deep to access.

What was once difficult to find is now easy as pie. Eskimo pie. ♦

Garrett Wasny MA, CMC, CITP is an Internet search consultant at www.garrettwasny.com.