# Web investigations

Fed up with trawling the web for information that
doesn't want to be found? Robert Blincoe turns his hard-nosed
journalist's skills to web searching and enlists the help of a private
detective to track down those vital details
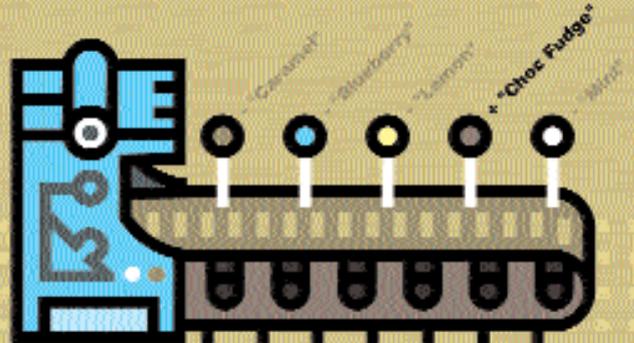
Searching the web for information is something of a broad and nebulous subject, but a few simple guidelines can make it a much less thankless task. If you know how to play about with search engines, look beyond the first item generated by them and question the content of the sites they lead you to – you'll find you're able to reach your objectives much quicker. It's about getting information rather than irrelevant data.

The specific examples in this article look at ways of finding information about people and businesses and making contact with them, but some of the techniques we outline can be used in the pursuit of any kind of knowledge. We asked private detective and expert on data protection Chris Brogan, director of legal affairs with Security International, to find out which freely available websites and

search engines he found useful. His answer: "I think the internet is crap. How do you know any of it is accurate? We use dedicated commercial databases which make the internet look rubbish."

He was being a bit disingenuous. Security International employs someone fulltime to visit and bookmark websites that might be useful to the company. There's a lot of information on the web: a lot of it is useful, a lot of it is rubbish.

Sometimes, to find what you want takes perseverance, imagination and scepticism. You should be mischievous and flexible – you don't have to use a search engine in the way it was intended.

Call in an engineer

The search engine used in our working examples is Google (www.google.co.uk or www.google.com). It's the world's most

popular website and searches through more than 2.1 billion web pages. Most people I know, including professional journalists and researchers, use it as their first port of call. Other options include AltaVista (www.altavista.com), Teoma (www.teoma.com) and AlltheWeb (www.alltheweb.com). The portals, like Yahoo (www.yahoo.com) and MSN (www.msn.com), also include search engines on their sites.

Google delivers unbiased search results from a vast index of websites and does not accept fees for finding and cataloguing sites. But before you go any further you should read *Googleitis* on page 128, to give you an idea of how you can start to rely on it too much.

Our search examples are based around *PC Advisor* and myself, so as not to invade anyone's privacy. The results of web
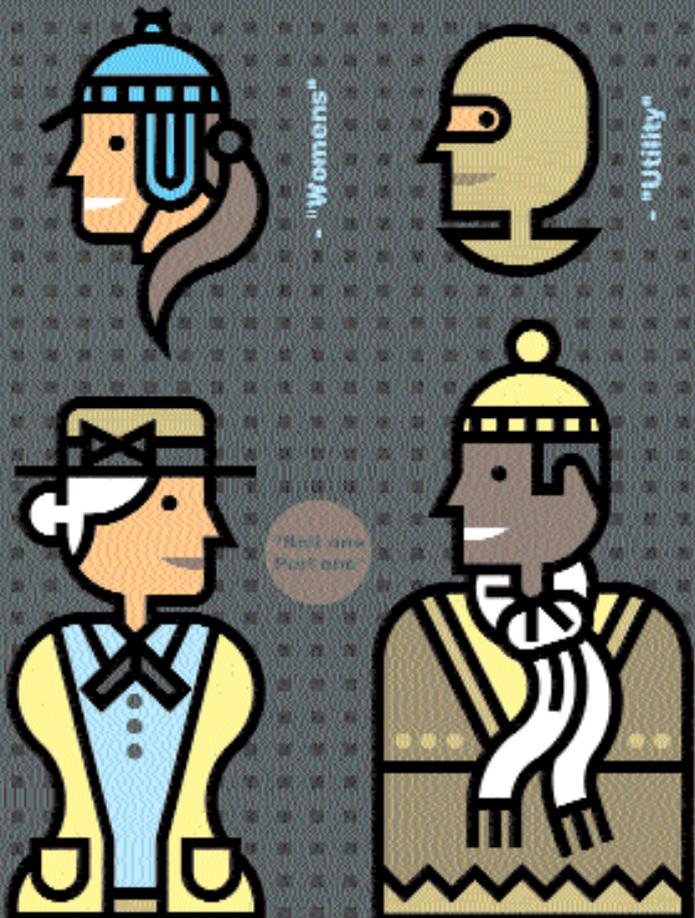
searches can differ daily, as web pages are updated and new ones are added, so if you try and recreate the searches we've carried out you might not come up with exactly the same results.

The name's the same

Imagine you want to know more about the editor of *PC Advisor* and want to contact him. All you know is his name and that he edits a computer magazine called *PC Advisor*. An obvious first step is to type "Andrew Charlesworth" into your search engine. Use quote marks to make sure it brings up the phrase Andrew Charlesworth, rather than lots of independent Andrews and Charlesworths.

On the day of my search the first item brought up by Google.co.uk is 'Agenda – Facing the Legal Challenges of Providing Internet Access', a web page detailing a conference organised by an Andrew Charlesworth. It's an IT-related conference so this could well be *PC Advisor's* Andrew Charlesworth.

Google displays the titles of the web pages it has found in blue with details below in black and the web address printed in green. In this case the address is 'www.ukoln.ac.uk/services/elib/papers/other/copyright/agenda.html'. Stripping the address back to its root (ignoring everything after the first '/') gives the domain name of the site: www.ukoln.ac.uk. From this address, specifically the 'ac.uk' part, we can tell the site is run by an academic organisation based in the UK (see *Evaluating a website*). The 'ac' stands for academic (not Andrew Charlesworth), while 'uk' is the country indicator. An academic domain is likely to contain more reliable information than a personal website.

Evaluating a website

What can the web address tell you?

Before you leave a list of search results, look at the web address and think which are most likely to have good and relevant information.

Is it somebody's personal page? The appearance of a personal name (such as rjones or jones) following a tilde (~), a percent sign, or the words "users" or "members" is a good clue. The address might also include the name of a popular ISP like Freeserve. Personal pages can be good, but they can also contain a right load of old rubbish. Just be aware.

Check the type of domain, the bit just before the first '/' in the address and whether this is relevant to the site's content. For government or military sites in the UK look for .gov.uk or .mil.uk. For academic sites look for ac.uk; non-profit organisations use the suffix .org; businesses tend to use .co.uk, or .com and may start using the new domain .biz.

Is the page from a foreign country and does that matter? Google has a 'Translate this page' option next to the page headings it finds during a search. Country domains include .uk for the UK, .fr for France, .de for Germany and .us for the USA.

When was the site last updated?

The 'last updated' information is usually at the bottom of a web page. Check the date on all the pages you are looking at, not just the site's home page. If the information is old, it might be useless. The date can tell you whether or not the author is still maintaining an interest in the page or if it has been abandoned.

Is the information any good?

Look for links on the site that say About us or Background and see if this information gives you any confidence in the site. Remember, anyone can put anything on the web and it doesn't cost them much money to do it.

Clicking on the page reveals that this Andrew Charlesworth is part of the School of Law, at the University of Hull. It could still be our AC – he might hold down two jobs or this page might not have been created very recently. A quick hunt around the page, right at the bottom, reveals this statement: "Site last revised on: Tuesday, 24-Nov-1998 14:21:20 GMT". It's four years old (see below).

We still don't know if we're looking at the right chap. *PC Advisor's* editor could have left the University of Hull to join the magazine which is based in London.

Looking further down the original Google search list (you can navigate backwards and forwards during your search by using the Back and Forward buttons on your web browser) the second

item – the web page for a series of conferences called *The Freedom of Information Act 2000 and its implications* – reveals Andrew Charlesworth is a "Senior Research Fellow, Centre for IT and the Law, University of Bristol". This information is written in the black text below the web page title. Following the link and scrolling down to the bottom of the page, shows this was last updated on 20 September 2002.

It would appear that the Hull University information is up to date. But still no mention of *PC Advisor*. It looks less likely that this is our AC; any conference involving a magazine editor would have the magazine's logo slapped all over it. This AC is the equivalent of the Danish ska band (see *Googleitis* on page 128).

Page version: 1;
Web page maintained by Philip Hunter and hosted by UKOLN - feedback to elib@ukoln.ac.uk .
Queries about the programme itself should be directed to the eLib Programme Office at elib@jisc.
Site last revised on: Tuesday, 24-Nov-1998 14:21:20 GMT
DC Metadata

Logic box

Phrase searches

Search for complete phrases, particularly useful for quotations and proper names, by enclosing them in quotation marks. The words "cheese and ham", for example, will appear together in all results exactly as you have entered them.

OR searches

To find pages that include either word A or word B, use an uppercase 'OR' between terms. For example, type "sandwich cheese OR ham".

+ searches

Search engines ignore common words like 'and' because they slow down the search. If the word is essential to the results, use a '+' sign in front of it and include a space before it. However, any words in a phrase search will be included in the search.

- searches

Sometimes a search term has more than one meaning: 'bass' can refer to fishing or music. You can exclude a word from your search by putting a minus sign immediately in front of the term you want to avoid. Include a space before the minus sign, but none between the sign and the word you want to exclude. So to find web pages about bass that do not contain the word 'music', type: "bass -music".

Googleitis

A web page further down the list, boldly entitled *Andrew Charlesworth* is another page with an academic domain (http://trapdoor.glos.ac.uk/el/gemru/staff/ac.htm), but visiting the page reveals this AC is a reader in cultural geography at the Cheltenham and Gloucester College of Higher Education. Neither our IT academic nor our *PC Advisor* man. It's time for a more intelligent search.

Boolean search

Always start by typing as much information into a search engine as you have. Entering "Andrew Charlesworth" + "PC Advisor" into Google immediately brings up *PC Advisor*, along with its address and telephone number. Clicking on the link brings you to the *PC Advisor* contact page and a link to Andrew Charlesworth's email. Now you can write and tell him what you think of his efforts.

The '+' sign means search the web for all occurrences of "Andrew Charlesworth" *and* "PC Advisor". Visiting Google's Advanced Search and Search Tips sections (links are on the Google home page) will fill you in on many clever searching options, but all you really need most of the time are specific search terms, in quotes, linked by a '+'. If something

you're not interested in comes up a lot, add the '-' sign to say and *not* that.

Google also allows you to search the entire web or just UK pages. Obviously *PC Advisor* is a UK-based magazine, but doing the all-web search brings up more results. This is because the magazine is part of a large US publishing company and Andrew Charlesworth's articles have been syndicated to other magazines. This illustrates that you should think about how you limit your search. On one hand you can miss out on information, on the other you can get too much.

We know where you live

Addresses and postcodes are just strings of numbers and letters. A search engine will interpret them as well as any other search term. Entering "85 Tottenham Court Road" – the address of *PC Advisor* – into Google, making sure you use the quote marks, will bring up several links relevant to the address.

Top of this list is 'EPR Architects' and a click on its link (http://www.epr.co.uk/architect_offices_tottcrtrd.html) will reveal that it redeveloped the office block at 79-85 Tottenham Court Road. It offers you several pictures, inside and out, of the beautiful building occupied by the *PC Advisor* team.

Next item down on the search list is a link to property company LMS's list of properties (http://www.lms-plc.com/html/property/properties.cfm?property_item=139&action=all). Clicking on this link will reveal a list of addresses, with 80-85 Tottenham Court Road highlighted.

Fourth on the Google search list is IDG.com, with *PC Advisor* mentioned in the second line of information. A click on this

When you read the words Napoleon Solo what comes into your mind? Have you no idea what it might mean, or does it conjure up the suave image of Robert Vaughan's character from the TV spy series *The Man From Uncle?*

If you typed the words into the Google search engine (www.google.co.uk or www.google.com), the two top results reveal that Napoleon Solo is the 'No 1 ska-band in Denmark'. You have to get down to the third listed result to get even a sniff of *The Man From Uncle.*

Unless you're a bacon-loving rude-boy, the Danish ska band is probably not what you're searching for. The fact that the group topped the list on the world's most popular web search site shows just how easily you can be led astray.

Napoleon Solo is no random example. The *Wall Street Journal*, quoting the *Sunday Times*' description of Labour MP George Galloway, printed this: "He may think like Che Guevara but he dresses like the [Danish band] Napoleon Solo." The introduction of the reference in brackets, [Danish band], was all the *WSJ's* work. A *WSJ* sub-editor who'd never heard of the character had typed the name into a search engine and accepted the first answer it came up with.

It's an example of being over-reliant on Google and not using your brain. The *WSJ* suffered an attack of Googleitis.
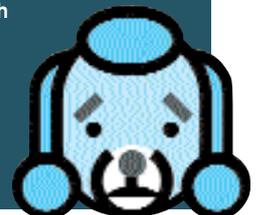
link reveals that IDG (International Data Group) is the parent company of *PC Advisor*. You can click through to the magazine's website or make a note of its phone number. The fifth item on the search list is a direct link to *'PC Advisor: Expert Advice in Plain English'*. Both the items reveal *PC Advisor* is based on the fifth floor of 85 Tottenham Court Road, London, W1T 4TQ. You've found quite a lot of information about *PC Advisor* and where it's based from a simple search.

If you'd entered just the postcode "W1T 4TQ" into the search engine, Google would have led you straight to *PC Advisor*.

Try doing a search on your own street name and postal area, for example, "Acacia Avenue" + "W10". You might be surprised by the results. I found a council report describing a mugging of an 86-year-old woman which happened just outside my house. The incident happened two months before I moved in. If I'd checked before I might have thought twice about buying my home. Checking out postcodes with Upmystreet (www.upmystreet.com) can also be illuminating.

Calling your name

It might surprise you that you can type a phone number into Google or any other search engine and get a result. It doesn't work with every number, which is a good thing for privacy reasons, but it does turn up info on commercial operations.

Try typing in "020 7291 5920", *PC Advisor's* editorial contact number, into Google. It brings up a link to *PC Advisor*, together with its address and other contact phone numbers. Use the quote marks or you'll bring up individual occurrences of the three groups of numbers '020' '7291' and '5920' and the results are less than fascinating even though, in this case, *PC Advisor* stays at the top. Be accurate with the way you arrange the numbers. Typing "0207 291 5920" (with the seven misplaced in the area code) will not bring up anything.

The commercial UK directory enquiry service

192.com (www.192.com) has suspended its own reverse phone number search, concerned that the service infringed data protection laws.

You might wonder why you'd want to do this kind of search if you're not a private detective. (TV detective Jim Rockford once spent an entire episode trying to steal a phone company book which allowed reverse searching, so he obviously thought it useful.) But getting an address from a number can be handy in business, particularly as some websites don't give full contact details.

Another route for tracking people and businesses is a directory search. Both BT (www.bt.com/directory-enquiries/dq_home.jsp) and 192.com run simple services. BT allows 10 free searches a day, and an additional 200 a month if you register. It's £14.95 a month for 100 searches on 192.com, but you can get some free if you sign up to direct marketing campaigns.

BT turned up six results for A Charlesworths in the London area, which is not too many to call and check, but he might not live in London or he might be ex-directory. The electoral roll, published

---

Going back to your roots

A classic reason for searching the web is to trace your family history. A first port of call, even for professional genealogy companies, is the electoral roll, so a visit to 192.com (www.192.com) is a good place to start.

Next would be accessing the 1901 Census. The UK's public record office has put this back online, having taken it off in January 02 when it was swamped and collapsed under the weight of web traffic. There are charges for anything beyond the most basic search, which is a minimum of £5 if you're using a debit or credit card.

At the Census site (www.census.pro.gov.uk) it helps to know a little about your family history. Name, age, place of birth and where they were living in 1901 produces better results. You could receive hundreds of hits if you only offer a surname.

Fortunately, I have an uncommon surname and know my family was settled in Rochdale by 1901. The site turned up some relatives, but others proved harder to find.

If you're not sure on a spelling of a name or feel it might have changed over the years, you can replace a letter with an underscore, or use an * where there could be several endings to a surname.

Know thyself

Extending your search, you could try Ancestry.co.uk (www.ancestry.co.uk), owned by the US firm MyFamily.com, which has been granted licence to the UK 1891 Census. Naturally, you must pay.

A word of caution on all this comes from Janet Carter, genealogical research manager of Ancestors of Dover, a professional genealogy company. She says she uses these sites "as a finding aid". "It saves a huge amount of time searching through parish registers, but you can never be sure what you have is accurate. You can draw up a very nice family tree, but can't be sure whether it's the right one."

The main flaw in online searches is that there are few resources available compared to libraries and church records. People also make mistakes when keying in information. Don't forget, much of the genealogical information you come across may have been compiled by keen amateurs such as yourself, rather than professional archivists. The view of Chris Brogan from Security International is, "All information is data, but not all data is information. And all information is not intelligence."

As well as the sites listed above, Carter recommends: www.genuki.org.uk; www.cyndislist.com; www.ancestry.com; and www.familysearch.com.

---
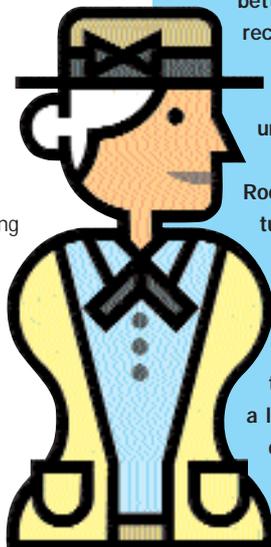
on 192.com, carries over 44 million UK adults' names and addresses. It even has a relationship search that allows people to look for every John and Mary Smith, for instance, that live in the same household.

Who are you?

If you wanted to contact me and did a search for my details, "Robert Blincoe" + "PC Advisor", you'd be out of luck. The search brings up two articles I've written, mentioning *PC Advisor*, but none I've written for *PC Advisor* (those are contained within *PC Advisor's* site and you must register to get them). Besides, I am a freelance writer so the site has no contact details for me.

Searching on just "Robert Blincoe" brings up a lot of results, but the first page contains my forefather, born in 1792 and placed in the workhouse, along with articles I've written and information about another Robert Blincoe who has written about missionaries.

Another way to search might be to consider whether or not I own a web domain, such as robertblincoe.co.uk or blincoe.com. Set your browser to the WhoIs site (www.whois.co.uk) and enter robertblincoe into the search box. It shows no domains registered for robertblincoe.co.uk and robertblincoe.com (no cybersquatting now), but it's a different story for blincoe.co.uk and blincoe.com. Both these are registered to a Robert Blincoe. Since October 02, Nominet (www.nominet.org.uk/ref/whois3.html), the UK internet names

organisation, has restricted the amount of information it gives out about personal domain holders.

If you visit www.blincoe.com you'll see the page is under construction. Never mind. Let's investigate blincoe.com by visiting SamSpade (www.samspade.org). This site offers a lot of complex search tools but we'll stick to the very basic. Scroll down to the first available box in which you can enter text and enter blincoe.com, then press 'do stuff'. Lo and behold it produces addresses and phone numbers for the organisation and admin behind blincoe.com.

If you'd entered blincoe.co.uk SamSpade would only have returned the same info as WhoIs. But this search can be very useful when you need to get in touch with one of those websites which won't print a phone number.

Blincoe.com is, in fact, owned by me. So is the information SamSpade turned up accurate? Is it heck. This is the internet. I typed in a made-up address. Do you think I'm crazy? ∎