This guide offers an overview of advanced online search for investigative journalism. It provides guidance on newsgathering, contributor-finding, advanced search theory and methods and sources for establishing the veracity of digital content.

Over the past ten years an explosion in online activity has opened up the potential for investigation, but has also added to the pervasive sense of information overload that blights modern life. These notes are intended to help investigators get a grip on online research and sources and avoid hours spent sifting through seemingly never ending search results.

by Murray Dick
about the author

Murray Dick lectures in multi-platform journalism at Brunel University, London. Before starting at Brunel, Murray was Information Officer at The Centre for Investigative Journalism, and prior to this he spent five years as a trainer in online research at the BBC. He is a freelance technology journalist, a consultant to the news industry in online search, and a keen(ish) blogger.

Additional information - anonymity, finding web site owners, hidden documents - has been provided by Mike Schrenk: www.schrenk.com

about the cij

The centre for investigative journalism (cij) came into being in 2003 to address a deepening crisis in investigative reporting.

The cij provides high-level training, resources and research to journalists, researchers, non-governmental organisations, academics, graduate students and others interested in public integrity and the defence of the public interest.

The cij is a non-profit organisation and runs international summer schools, training programmes in basic and advanced investigative techniques and organises public meetings – all designed to raise and sustain the standards of investigative reporting. Our handbooks, archive material, web and audio resources have helped bring additional investigative tools to journalists and the community unable to attend cij workshops and training programmes.

The cij offers particular assistance to those working in difficult environments where freedom of the press in under threat and where reporting can be a dangerous occupation.
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>newsgather online</td>
<td>4</td>
</tr>
<tr>
<td>personalised newsgathering and RSS</td>
<td>5-6</td>
</tr>
<tr>
<td>real-time search and social discovery</td>
<td>7</td>
</tr>
<tr>
<td>trend-spotting sources</td>
<td>8-9</td>
</tr>
<tr>
<td>building a master feed with Yahoo Pipes</td>
<td>10</td>
</tr>
<tr>
<td>finding people online: some sources</td>
<td>11-12</td>
</tr>
<tr>
<td>finding people on social networks</td>
<td>13-14</td>
</tr>
<tr>
<td>forums and discussion boards</td>
<td>15</td>
</tr>
<tr>
<td>search engine strategy: basic theory</td>
<td>15</td>
</tr>
<tr>
<td>search engine advanced operators</td>
<td>16-17</td>
</tr>
<tr>
<td>search engine filters (Google)</td>
<td>18</td>
</tr>
<tr>
<td>hidden documents</td>
<td>19</td>
</tr>
<tr>
<td>Google search caveats</td>
<td>20</td>
</tr>
<tr>
<td>verifying online sources and digital documents</td>
<td>20</td>
</tr>
<tr>
<td>how to tell if a website is legitimate</td>
<td>21</td>
</tr>
<tr>
<td>tracking where someone sent you an email from</td>
<td>22</td>
</tr>
<tr>
<td>finding out who created a word document</td>
<td>22</td>
</tr>
<tr>
<td>tracking a website which has gone down or has been deleted</td>
<td>23</td>
</tr>
<tr>
<td>anonymous web surfing</td>
<td>24-25</td>
</tr>
<tr>
<td>tools to help spot photographic hoaxes</td>
<td>26</td>
</tr>
</tbody>
</table>
When it comes to finding news online there are some well-known starting points:

**Google News:**
http://news.google.co.uk

**Yahoo News:**
http://uk.news.yahoo.com

Both Google News and Yahoo News offer a broad range of sources; from regional to international press, trade press and blog output. But the former also includes many direct PR sources, which should be read with a degree of scepticism. Though there is much overlap between these aggregators, in practice each indexes its sources to varying degrees, so it is always a good idea to run news searches across each of them. Both offer advanced search options — allowing you to filter your results by source(s) or by country of origin (which, for example, can help obviate the problem of duplicate place names).

**Newsnow** is a useful alternative to Google and Yahoo News. It is especially useful for browsing news by topic field, something which the other two search-driven aggregators don’t provide with such clarity. However, a number of titles have withdrawn their copy from the Newsnow index, so it is not as robust a source as once it was.

Another option is **Silobreaker**, a semantic search engine that helps you make sense of the people, places and issues in your newsgathering.

Silobreaker has an innovative tool that allows you to view those entities which comprise the news in terms of their relationship to each other, by citation in the news (via the ‘network’ option, above the search bar). This can be especially useful in the early stages of newsgathering when you need an audit to make sure you are not missing anything out.
While it is often essential to manually gather news around particular topics in an ad hoc way, for long-term investigations and fields of ongoing interest, it is sensible to automate the process. This can be done with Google Alerts (and various other search-results-by-email systems), but to avoid an excessively large inbox, it is better to make use of Really Simple Syndication (RSS) which you can harvest by using a feed reader.

RSS is a distribution technology that allows publishers to instantly alert surfers when they have published new content. News organisations subscribe to press agency news wires which channel in breaking news from around the globe. RSS can be used to replicate this process taking just a little time to establish and maintain the feeds you are interested in.

In order to subscribe to RSS feeds you will need a platform to host them, of which there are a range of options. Most web browsers now offer a means of storing RSS subscriptions; however these will sometimes be stored locally on the computer and so are platform-dependent. This is of no use if you need access to your feeds from other computers - if you hot-desk, or if you also use mobile devices (such as a PDA, a Blackberry or an iPhone).

As an alternative to browser-based subscriptions, you can set up personalised feeds in:

- **iGoogle:** [www.google.com/ig](http://www.google.com/ig)
- **Pageflakes:** [www.pageflakes.com](http://www.pageflakes.com)
- **Netvibes:** [www.netvibes.com/en](http://www.netvibes.com/en)
- **Windows Live:** [http://explore.live.com](http://explore.live.com)
- **Addictomatic:** [http://addictomatic.com](http://addictomatic.com)
- **My Yahoo:** [http://uk.my.yahoo.com](http://uk.my.yahoo.com)

Perhaps the most robust of all the feed readers is Google Reader (free but login required). Once you've logged into Google Reader, there are three areas of the site you should acquaint yourself with:
Google Reader: www.googlereader.com
• the ‘add’ a subscription box on the top-left of the screen
• the reading pane in the middle of the screen
• the ‘subscriptions’ section at the

RSS is near ubiquitous online – here are some of the ways you can make the most of it:

• Initially, try searching within the Google Reader index (in the ‘add a subscription box’) for sources, names or keywords you're interested in. When results come back, click on the ‘subscribe’ button under the search result. Look at the number of subscribers as this can give you a clue as to whether or not a feed is still active.

• Don't be limited by what feeds you can find in this index – it is not exhaustive, and some sources provide subject-based RSS feeds on their own site. Look for the orange (or blue) button or the legend RSS on the page. Click the button and paste the feed into Google Reader's ‘add a subscription’ box.

• Many search engines, including Google News and Yahoo News, offer results as RSS feeds which will keep you up-to-date. Be aware that this process can be less than intuitive in some cases. There is a tutorial on YouTube with instructions for setting-up keyword- or tag-based RSS results: www.youtube.com/rssls

• Lastly, you can even 'scrape' your own updates from pages which don't use RSS by using one of the following tools:

Page-2-RSS: http://page2rss.com
Feed43: http://feed43.com

Google Reader has a clear advantage over some of the other popular feed readers due to the ways in which you can share what you have found. You can share individual items with your Google contacts, you can put some feeds in a folder and share its contents with the world, or you can save all your feeds as an OPML file, and pass it on to others working on the same project (use the ‘import/export’ tab in the ‘manage settings’ section).

When browsing through your feeds in Google Reader, it is often a good idea to opt for the ‘list’ view, as opposed to the ‘expanded’ view (see options on the top-right hand side of the interface) – this will reduce unnecessary and repetitive scrolling.

Feedity: http://feedity.com
Feedyes: www.feedyes.com
Just as it is possible to set up RSS feeds for published content, so too can we keep up-to-date with the content people share across social networks. It has taken a long time for **Google** (and the other established search engines) to come to terms with the speed at which news breaks across the social web, but now they are beginning to incorporate results, from Twitter and other real-time sources, into their search results. This can be a great way of finding out news as it breaks. There are several real-time search engines that are helpful:

**Social Mention:** [http://www.socialmention.com/](http://www.socialmention.com/)

**Icerocket:** [www.icerocket.com](http://www.icerocket.com)

**PicFog** (for images shared on Twitter): [http://picfog.com](http://picfog.com)

**Just Spotted:** [www.justspotted.com](http://www.justspotted.com)

**Twitter search (adv):** [http://search.twitter.com/advanced](http://search.twitter.com/advanced)

**Twitterfall:** [http://twitterfall.com](http://twitterfall.com)

All of these 'cloud' services will let you create a personalised RSS space.
You can attain a very different sense of what is news by analysing what readers are most interested in. This represents (at least notionally) the democratising nature of online communications via social web sources. Though soft news topics and consumer issues may dominate amongst what people search most for online, trending keywords and topics can nonetheless help you eke out news which, for one reason or another, may not be carried by traditional news sources. You can keep an eye on what terms are trending on:

**Google Trends:**
www.google.co.uk/trends (UK) (updated daily).

**Google Hot Trends:**
www.google.com/trends/hottrends (USA) (updated hourly).

Some other sources for capturing trends online in social media include:

- **Trendistic:**
  http://trendistic.com (good for looking at trends over past 30 days)

- **What the Trend:**
  http://whatthetrend.com (in-depth trends on Twitter)

- **Trendsmap:**
  http://trendsmap.com (visualises Twitter search terms and trending topics)

- **Twitscoop:**
  www.twitscoop.com (trends on Twitter)

- **Delicious:**
  www.delicious.com/muz038/tools+trends
During the Glasgow North-East by-election in November 2009, rumours circulated that the BNP might sneak into third place, forcing the Tories into fourth, and setting a dangerous precedent during a period when the party seemed to be growing in profile and support. The evening of the election I sat up and tracked several twitter accounts to people present at the count – one of them, a parliamentary assistant to one of the candidates, tweeted that the BNP had come in fourth a full fifteen minutes before either BBC News 24 or Sky got to it.

People are making the news using Twitter, and as long as you have a means of verifying that people are who they claim to be, today we all have access to many news stories.
Once you have a slew of feeds on different topics, it can be difficult keeping on top of them all – and you may want to bring them all together in one stream, whereby you can filter out duplicate (and inappropriate) content, or build a master feed which you could apply for your own uses (such as hosting on your blog or website). This is where:

**Yahoo Pipes:**


an easy-to-use modular utility, comes in very handy (account required) the key elements of a simple Yahoo Pipe are:

- **Fetch feed (in ‘sources’):** paste your RSS feeds in this box – use the + sign to add.
- **Unique (in ‘operators’):** this filters out non-unique content, so if you pull a story from a local newspaper which also crops up in an aggregator feed search, it won't publish all three.
- **Filter (in ‘operators’):** filters out terms associated with places, events and issues you don't want coming back. For example, there is an Uxbridge in Massachusetts often abbreviated to Mass in blogging and news. If you're looking for Uxbridge Middlesex, use this filter to block all references to the term Mass.
- **Sort (in ‘operators’):** this filter allows you to sort your content by publication date – obviously for a newswire we need all content to be distributed by publication date.

Join these together into a logical flow using the drag-and-drop connectors, and connect to the ‘output box’. Once saved, you can lift the RSS for this feed and do what you like with it. Be aware that Yahoo Pipes, as with most freely available online tools, can on occasion be temperamental – but it is nonetheless worth persevering with.
In the UK an essential tool for finding people is 192.com with Tracesmart and Cameo as useful alternatives:

192.com: www.192.com
Tracesmart: www.tracesmart.co.uk
Cameo: http://cameo.bvdep.com

With a combined electoral roll and directory enquiries service, company and director information (from Companies House), as well as house price data (from the Land Registry) 192.com in particular is a robust, bespoke subscription tool for people-finding. Though you need to pay for electoral roll information, searching for phone numbers is free. You can search by street name, cross-reference a partner (or co-habitant’s) name, and search by town or city – all key elements in finding where someone lives.

The availability of historical electoral roll content is particularly helpful. Since 2003 (under the Data Protection Act) people have had the right to opt out of public (or edited) versions of the electoral roll, so past records are often essential in tracing people.

BT's Residential Directory Enquiries: www.thephonebook.bt.com offers a free alternative, while internationally:

Infobel: www.infobel.com/world/default.aspx

is the best way to go about finding international white and yellow pages. However, the rise of mobile phones (and decline of landlines) will see this site become less useful as time goes on.

Social networks are changing the ways journalists can find contributors. Just as people are asserting their right to privacy by taking flight from official directories,

so too are they increasingly (and voluntarily) publishing details of their personal lives across cyberspace. This engenders a moral dilemma for investigators – but not something which should cause paralysis. All investigations into finding people online should be carried out to the letter (and in the spirit) of NUJ guidelines.

Given none of the major search engines offer people-finding options, a gap in the search market has opened up, leading to the development of a number of little known (and free) alternatives to the big names in search. Three of the more useful ones are:

123people: www.123people.com

In the UK an essential tool for finding people is 192.com with Tracesmart and Cameo as useful alternatives:
Each has their strengths and weaknesses. For example, yoName searches across several social networks (and some other sites). However, it can be less than useful when it comes to Facebook entries. This source applies best match (rather than exact match) search criteria to entries in some social networks (notably MySpace), which can lead to fuzzy results.

123people offers a means of searching a wider spread of the internet than either yoName or Pipl, and aggregates the results accordingly. For example, on the top-left of the results page in 123people, you will find a list of any email addresses found across the web which may (or may not) match your query. This is especially useful given how few email directories are available online, and (as we will see) how unreliable such searches are in conventional search engines.

The second most noteworthy feature in 123people is the Amazon wish lists at the bottom of the results page. This can be a great place to discover which authors, genres, bands, or even computer games, your subject is interested in, and the results often include general details of where people live – perfect for local journalists seeking out an angle.

Web content on 123people is pretty interesting too – it scours the indices of Yahoo, MSN and Google to provide a better spread of the web, and they provide results in other formats, aggregated accordingly (so Word docs, PDFs, and other formats, which often contain useful biographical and contact information, are presented separately).

As with 123people, Pipl lets you specify the (likely) location of your name. Pipl is the only one of the three that allows you to distinguish between first and second names – it’s also the only one that lets you determine them from the outset of your search. But as with all (or certainly some aspects) of all of these tools, it doesn’t run an exact match for your names, just a best match.

Free people finders can lack precision, but they really come into their own when combined with other, more official sources. While consulting for an NGO, we were struggling to find the home address of a popular (and controversial) public figure. Electoral roll results were misleading – the results returned did not match the individual’s social class profile. However, searching for the name in 123people returned an Amazon wish list which indicated that the person in question is known alternatively by his different middle names in different contexts. Armed with this wider knowledge, his identity became clearer with subsequent searches of the electoral roll.
Newspaper groups make great use of Facebook and other social networks when distributing their stories, but their use in newsgathering and contributor finding is equally as significant. Being culturally and socially relative, each network has different groups of users and different purposes:

**Facebook:**
[www.facebook.com](http://www.facebook.com)
for Europe, North Africa, North America and India

**Badoo:**
[http://badoo.com](http://badoo.com)
for Europe but not UK

**Kontakte:**
[http://vkontakte.ru](http://vkontakte.ru)
for Russia

**Nasza-klase:**
[http://nk.pl](http://nk.pl)
for Poland

**Orkut:**
[www.orkut.com](http://www.orkut.com)
for Brazil

**QQ:**
[www.qq.com](http://www.qq.com)
for China

Finding people from these countries in western countries, such as the UK, may best be served by using such sources. Facebook's search criteria are very basic, but as an alternative, there is an application which can be used to refine search considerably:

**Advanced Search 2.2 beta:**
[www.facebook.com/advancedsearch](http://www.facebook.com/advancedsearch)

This advanced search can be helpful where you are looking across a range of interests simultaneously. In order to search Facebook, an account is required. Facebook offers the facility to search by email – a useful feature if you have a person's email address as these are unique to the individual.

**Twitter Advanced:**
[http://search.twitter.com/advanced](http://search.twitter.com/advanced)

that allows you to search for people by location, in addition to other Boolean and advanced operators and filters. Use postcodes rather than place-names to avoid duplicate place-names. However, this approach is fuzzy at best – many people do not specify where they tweet from or do not use mobile devices to access Twitter. If you are interested in filtering tweets so that you see only those linked with pictures, videos and other media (which can be very useful in terms of following breaking news), you can try the following search:

`yfrog OR twitpic OR tweetphoto OR snaptweet OR twiddeo OR twitvid near:edinburgh within:10mi`
Twitter allows this type of proximity search by means of co-ordinate geolocation. You can find the longitude and latitude of a place using:

**Google Maps:**
http://maps.google.co.uk

- enter the address details in the search box
- zoom in until you locate the place you are looking for
- right click the location on and select “center map here”
- click “link to this page” and copy and paste the URL—look for the value of ll parameter in the URL these are the latitude and longitude co-ordinates.
- copy and paste that value in the Google Maps’ search box to confirm the location

Once you have co-ordinates, you can search for tweets to within fractions of a mile by manually editing your search terms in the search box (so within: 0.1mi, rather than within: 1mi). It’s not clear just how accurate this is (given the nature of GPS), and results in built-up areas can be markedly different.

Perhaps the most intuitive place to search for Twitter users by location and/or profession (whether experts or other newsworthy people) is via a bespoke directory like:

**Tweepz:**
http://tweepz.com

In terms of other social networks,LinkedIn is probably the best resource. LinkedIn is a social networking site aimed at the professionals and for those seeking contacts and jobs in their industries. Although there’s a heavy US presence here, LinkedIn contains an impressively broad range of UK-based professionals.

Registration is not necessary to search, albeit it will let you make contact, and seek out introductions to people. If you want to send mail, request more than five introductions at a time or see expanded results, you will need to buy a pro account (there are different levels of access).

LinkedIn can be especially useful for seeking out people by employer — present or past. This can be a good means of approaching a sensitive issue, or an issue where employees (or potential whistleblowers) are being discouraged from talking to the press.
forums and discussion boards

Some useful sources for finding people where conversations are taking place online (whether experts, hobbyists, members of the public, or whistleblowers) include:

**Google Groups:**
http://groups.google.com

**Yahoo Groups:**
http://uk.groups.yahoo.com

**Omgili:**
http://omgili.com
(often anonymous discussion boards)

**Boardtracker:**
www.boardtracker.com
(often anonymous discussion boards)

**Google Blog search:**
http://blogsearch.google.com

**Blog pulse:**
www.blogpulse.com

**Twingly:**
www.twingly.com

For help in finding bloggers by field or interest, use:

**Omgili**
http://omgili.com
(often anonymous discussion boards)

search engine strategy: basic theory

Although proprietary, newspaper archives such as *Nexis* and *Factiva* are the most useful and reliable sources for finding news and backgrounders, there are some advanced functions you can use to speed-up (and tighten-up) your search.

The following are two key issues to bear in mind when undertaking search:

- Think of what you want from the results, rather than the question and how these results might be presented, where on the page or in the document would they be found? Remember – search is culturally (and linguistically) relative.

- Once you’ve done this, choose your terms – use a thesaurus, and think around your subject – try to avoid words with multiple meanings. If that’s not possible, try incorporating those terms as part of a phrase.
Some useful sources for finding people where conversations are taking place online (whether experts, hobbyists, members of the public, or whistleblowers) include:

**AND:** the AND operator in Google is implicit when you put spaces between terms.

**OR:** use this operator when either one or another word is desired – blair wmd OR weapons will return results where either wmd or weapons is found in conjunction with blair. Alternatively, use the pipe symbol – blair wmd | weapons.

**NOT:** use this operator to remove certain words from your results, but be wary not to use it too much, or you may miss vital results – rangers - qpr will remove references to Queens Park Rangers.

**Phrase search:** use to restrict a search to a particular phrase. This is especially useful when searching for names or a turn of phrase – “richest bank in the UK” (try with and without quotes)

**Wildcard:** Google doesn’t support the wildcard in the way it is conventionally used in other search engines – because it uses automatic stemming to find alternate endings and spellings for your terms. But you can use it in:

- **In Google you can use a ** when phrase-searching in order to serve as placeholder for a particular word (which could be one of many things). For example, the query Norway oil ** said** will return results where any industry commentator, politician, or other vested interest has contributed their opinion on the oil industry in Norway, within news results. This can be a useful way of finding experts on a given topic.

- **+:** the plus sign allows you to stop Google from stemming your words – if you are interested in a word in a particular case. It can also be applied to stop Google finding reference to certain words which link to (rather than feature in) the pages you are searching from, when viewing cached content (in your search results).
**Synonyms:** the tilde operator can be used to take advantage of Google's internal thesaurus. "~marriage" will return related concepts.

It's worth bearing in mind that other engines offer an even broader range of search operators. **Exalead**, for example, permits **atleast** and proximity searching. Their **atleast** function allows the searcher to find pages that feature a term prominently, which can be useful when you are searching for backgrounders on people or issues.

The proximity search function allows the searcher to find terms which occur close to each other, this can be useful when trying to unearth connections between people and events in the news.
search engine filters (Google)

**the domain filter - site:**
This allows you to filter results by domain or domain type. It can be especially useful when you are looking for information from a particular type of source. All UK universities share the .ac.uk domain – so seeking out academic expertise will always benefit from this type of filter. Likewise, many UK charities, NGOs, action groups and other non-profits use the .org.uk domain. For a comprehensive list of national and top-level domains, check

**NORID domains:**
www.norid.no.

To find academic experts, it helps to incorporate three factors into your search:

- your subject term/s
- a term connecting the subject to his/her profession (ie expert, department, professor etc.), and
- the domain function: site:.ac.uk (for UK universities)

So, to find an expert in solvent abuse (UK), try comparing the following examples:

- **compare:** expert solvent abuse with: “solvent abuse”
- **professor site:ac.uk**

To find a (US) expert on presidential spouses:

- **compare:** expert political spouses with: “political spouses”
- **professor site:.edu**

Don’t forget to check Google’s cached option in your search results – it will show you the page as was when it was first indexed, so you won’t miss out on your terms if the page has changed in the meantime. In addition, you can use **<Control + F> or <Apple + F>** in your browser to locate terms in long documents.

You can even use the domain function to find discussions in Facebook try searching:

- **lehman brothers site:facebook.com/topic**

You can’t do this within Facebook.

**the occurrences filter - intitle:**
This filter can be useful for finding backgrounders as it is a particularly restrictive field and exploits headline-writing standards across the media (but it can be hit-and-miss). If you want to find background (analysis, not news, professional rather than amateur) on the Somali war in 2007, try comparing the following searches:

- **compare:** somalia war with: somalia war intitle:Q&A

Alternatively you could try:

- **depth/comment/analysis/brief/background in the intitle: field.**

**the file-type filter - filetype:**
The file-type filter can be useful for locating information according to common usage in certain types of file. For example, if you were searching for an expert in greenhouse gas emissions, it is likely the experts in this field will have published PowerPoint slides to present at conferences.
People often leave sensitive and confidential information online. Sometimes this information is carelessly deposited in forums and social networking websites. At other times, information is made available because of system errors or carelessly configured servers. In any case, search engines find it – and so can you, if you know where to look.

Using this search you can look within a specific document for information, for example, you are more likely to find financial information by looking through excel files. You can use any file extension to narrow your search – jpg, wav, ppt, avi etc.

Combine search terms to narrow results further: filetype:xls “house prices” and specify the location with + London.

Or exclude terms, for example filetype:doc “security plan” - guidelines will exclude guidelines from the results adding site:gov will restrict the search to government sites.

However, be very careful what you download as the documents could have viruses – to be safe, open any documents in:

Microsoft Writer:
http://windowslivewriter.spaces.live.com
or
Open Office:
www.openoffice.org

The searches listed below are of the most interest to journalists.

The Google guide of advanced search techniques:
www.googleguide.com/print/adv_op_ref.pdf

Provides detailed ways of searching for information and you can find more about search techniques on the web, including YouTube, by searching for “Google hacking” or “Google hacks”.

However, be very careful what you download as the documents could have viruses – to be safe, open any documents in:

Microsoft Writer:
http://windowslivewriter.spaces.live.com
or
Open Office:
www.openoffice.org

The searches listed below are of the most interest to journalists.
Google search caveats

While the above and other filters, operators and functions (such as the number range filter) can be useful, Google doesn't deal very well with punctuation. Often you will find phrase searches include results where the name you have searched appears with a comma separating the first and second names – clearly not the person you were looking for.

verifying online sources and digital documents

tracking who owns a website:
Through the DNS (domain name system) you can find information such as who a web url is registered to and when it was registered. You can do this using the IP (internet protocol) address of the site you want information about.

What is my IP: www.whatismyip.com will tell you your IP address.

Network Tools: http://network-tools.com by typing the url of a website under look-up you can find the IP address and some basic information such as where the site is registered. You can then use one of the regional internet registries (right) to find out more information.

Domain Tools: www.domaintools.com is another site for finding out about a site based on the url. By using the ‘who is’ function, you can find out who owns the site, the contact details and when the ownership expires etc.

Look up Server: www.lookupserver.com is another option to find information on a DNS as is All Who Is: http://allwhois.com and Reverse DNS Lookup: http://remote.12dt.com
how to tell if a website is legitimate

First consider the purpose of the site you are looking at. Is the site trying to sell you something? Are they trying to entertain, or to advocate a particular viewpoint, or a combination of these? You might want to ask who is funding the site, and why? Fundamentally you need to ask: why does this site exist?

Consider also the author of the site (assuming there is one). Are they an expert in the field they claim? Do they have 'previous'? What else can you find out about him/her/them which might bring into question their reliability? It is very easy to phrase-search an author's name in any search engine to find out what else they might have done, who else has taken note of their work or activities, and what the general consensus is about them.

You can use

**Yahoo Site Explorer:**
http://siteexplorer.search.yahoo.com

to establish which other sites link to the site you are concerned with – which might suggest an affiliation.

When searching for the official websites of corporations and other large organisations, it is always worth looking for them within Open Directories:

**DMOZ Open Directory:**
www.dmoz.org

**Yahoo Directory:**
http://dir.yahoo.com

The information about websites in these directories is managed by real human beings, who are 'experts' in the topics they manage. Search engines cannot offer such protection, so there is at least a modicum or reliability about the results you find in here.

Alternatively, seek out those online sources that deal in online hoaxes for guidance, including:

**Snopes:**
www.snopes.com

**Urban Legends:**
http://urbanlegends.about.com

**Museum of Hoaxes:**
www.museumofhoaxes.com

In summary, there are many ways to check the veracity of an internet source, but there is no substitute for the traditional, and entirely healthy scepticism most journalists will take to finding any source they might consider using.
tracking where someone sent you an email from

Whichever email system you use, you can access the email's header which will more often than not divulge the IP address of the sender. For more information about tracking emails see:

Email Tracker Tutorial: www.visualware.com/resources/tutorials/email.html

finding out who created a word document

While some basic metadata is often available via the (right-click) properties menu option on your desktop, there are tools that can help locate further information; in some cases the username and file-path used in the creation of a document, which can in turn provide valuable clues as to the provenance of a document:

Clean content: www.oracle.com/technetwork/middleware/content-management/overview/index.html (program)
Docscrubber: www.javacoolsoftware.com (program)
tracking a website which has gone down or has been deleted

The **Way Back Machine**: [www.archive.org](http://www.archive.org)

is an excellent way to pry into previous versions of web pages. This can be useful when pages disappear, move to new locations, or the content changes. Sadly it is possible to design websites in such a way that their pages do not get into the servers of the **Way Back Machine**, and site owners can approach the webmaster and request that content be removed. Nonetheless, this source can be a very useful tool for uncovering lost or unavailable material.

A similar (albeit far more restricted) means of accessing past pages of websites is in the cached option you often see associated with Google results. Click on these links to see what the website looked like when it was last indexed, this goes some way to explaining why from time to time you click on a page from your search results which don't actually feature the terms you've searched for – that page is no more. This is incredibly useful for keeping track of political bodies, meeting-minutes, terms and conditions and corporate social responsibility charters.

While consulting at BBC News 24, the duty editor was trying to find suitable contributors for a story about firemen being attacked by members of the public while tending to their work. He recalled someone with experience in the field, who used to work for a charity set up by firefighters in Islington to smooth relations with the local community. But he couldn’t remember the person's name, and had no contact details. The charity had long since disbanded, and the hosting for the website lapsed.

By finding a dead link to the charity website on another web page, we resurrected the site in Internet Archive, and found a name and phone number.
Why anonymity is useful to journalists:
Anonymous browsing techniques may protect your identity and thereby provide greater access while conducting research. Anonymity may protect you if performing automated or repetitive research tools. With certain techniques, you can conduct research while appearing to be doing so from another country.

Your sources may use anonymity techniques to either protect their identity or to make their story possible. Anonymity techniques can defeat national firewalls and get information out to the rest of the world.

anonymous email
Anonymous email is possible via a product called NYMS: www.anonymizer.com/nyms/, which allows the creation of disposable email addresses via the Nyms network.

proxies
Proxies act as intermediaries and protect your identity. There are different types of proxies from different sources. Open proxies are servers that either intentionally, or because of misconfiguration, allow people to connect through their network, and assume one of their network IP addresses. An example of website that lists open proxies is Xroxy:
www.xroxy.com/proxylist.htm
However, open proxies are best avoided. There are also commercial proxies that do a better job. For example:

**Anonymizer:**
www.anonymizer.com

Anonymity becomes more important as regimes place added restrictions on journalists’ ability to speak freely. Regardless of the measures governments take, however, journalists are still able to publish stories through the use of “proxies”.

**Tor**

An alternative proxy is the **Tor Project:** www.torproject.org.

Tor is the proxy network that facilitates journalism from some of the most hostile environments in the world. It is free software and an open network that helps protect against a form of network surveillance that threatens personal freedom and privacy, confidential business activities and relationships and state security known as traffic analysis.

Tor was originally developed for the US Navy for the primary purpose of protecting government communications. Today, it is used every day for a wide variety of purposes by the military, journalists, law enforcement officers, activists and many others.

It protects you by bouncing your communications around a distributed network of relays run by volunteers all around the world: it prevents somebody watching your internet connection from learning what sites you visit, and it prevents the sites you visit from learning your physical location. Tor works with many of your existing applications, including web browsers, instant messaging clients, remote login, and other applications based on the TCP protocol.
There are a few things you should consider when happening across digital photographs which you might want to use in your investigations. For example, you can check the stills against news wires (stills) in Yahoo or Google image verticals.

Moreover, watch out for particular ratios – Yahoo news stills are often 380 wide x 345 high. You can check the metadata attached to photos in

**Exif Reader:**
[www.takenet.or.jp/~ryuuji/minisoft/exifread/english/](http://www.takenet.or.jp/~ryuuji/minisoft/exifread/english/)

which gives details such as the camera used, the time it was taken, the scale and the geo-location. If the image has been doctored, there will be no exif data available – again this does not necessarily mean it is a hoax but should alert you to exercise caution.

Lastly, use stills editing software to zoom in on the pictures in question – look for the areas where tones meet – this should give an idea of whether manipulation, or layering has taken place. One final caveat on this last option – it can sometimes be difficult to spot such manipulation in jpegs, because they are compressed, and pixelate significantly when viewed with zoom.

If the pictures have been sent via Powerpoint, beware. Check the image size by using photo editing software (there are free online alternatives to Photoshop such as **Picnik:** [www.picnik.com](http://www.picnik.com)) to establish the photo's original size. Most cameras shoot stills at a ratio of 2000 x 1200 pixels or higher. Any smaller, and the still has been resized – which doesn't mean it's a fake, but which should none-the-less raise concerns.
This handbook was made possible by a grant from the Open Society Institute [www.soros.org](http://www.soros.org).

We would really value your thoughts and feedback on this handbook. Please write to us at [info@tcij.org](mailto:info@tcij.org).

If you would like to suggest a new topic for a handbook, or know journalists/authors who could help write one, drop us a line at the above address.

For more in the handbook series, please visit our website [www.tcij.org](http://www.tcij.org).

City University London, Journalism & Publishing, Northampton Square, London EC1V 0HB

[info@tcij.org](mailto:info@tcij.org) | +44 (0) 207 040 8220 | [www.tcij.org](http://www.tcij.org)