

Search 2016: Human not Artificial Intelligence Needed

Karen Blakeman, RBA Information Services

karen.blakeman@rba.co.uk

One of the advantages of running search workshops is that it forces me to keep up to date with changes and new developments. For many people, weird search results are something that they have to cope with but do not necessarily have time to investigate. They are providing information and support to other people, and once one project is over another takes its place. There is very little time to look into why the search engine - often Google - is behaving bizarrely. For me, keeping abreast of what the search engine companies are doing is a large chunk of my job and what I discover is sometimes disconcerting and worrying.

Single, small changes in algorithms build up over time to effect bigger changes in the way a search is analysed, processed and presented. For example, Google recently stopped showing advertisements to the right of the results on desktop search. This was not altogether unexpected since Google and its competitors have been steadily moving towards a single, simplified interface that works on all types of devices. Don't think though, that there will be less advertising. Google is already pushing extra advertisements to the top of the search results, which means scrolling down further to get to the more reliable results. And it is all too tempting when using a mobile device to click on the first vaguely relevant link.

It is not just the advertising that one has to be wary of. A major trend with all of the major search tools is to offer "facts" and quick answers, extracted from one or more websites, both at the top and to the right of results. No need to click through to a document to find the answer to your query because it has already been found for you. The problem with these "facts" is that the source is not always given and the overall quality appears to be going downhill rapidly. Run a Google search for court fees for the UK small claims procedure and you'll probably see a four-row table that starts with a fee of £205 for claims up to £5000. A note at the bottom of the table tells you that there are five more rows, the implication being that there are higher fees for higher levels of claims. When you click through to the web page there are in fact more rows at the top of the table showing lower rates for claims below £5000. This is not an issue for those who take the time to click through to the website to see the full table, but those who do not could be deterred by Google's answer from pursuing their claim.

Where the search engines can seriously start to go wrong is when they combine information from multiple sources. Examples pop up with alarming regularity. Type “*frugivore*”, for example, into Google and it tells you that cats, lions and killer whales are examples of fruit eating animals! Some errors, such as this, are obvious howlers. It is the almost-but-not-quite-right answers that are potentially dangerous. Clicking through to the source, if it is provided, and double-checking the information with another site is time consuming but vital if one is to be sure of the accuracy of the data.

The most significant, recent development in search is the use of artificial intelligence (AI) as part of the mix. Google, in particular, has invested heavily in AI and in October 2015 confirmed that [RankBrain](#) is now an integral part of its web search. As if to further emphasise its commitment to AI John Giannandrea, who had been leading the company's research into artificial intelligence, took over as Senior VP of search in February when Amit Singhal retired. There is much discussion as to how the AI component actually functions and some have attributed the increased variability in the quality of results to its activities (Google: [RankBrain Doesn't Use New Signals But May Adjust Weights Of Existing Ranking Signals](#).) What is certain is that all bets are off when it comes to predicting how our searches are likely to turn out. Knowledge of advanced search commands will help to a certain degree but critical appraisal of what pops up on the screen is now more important than ever.

The impact of new technologies on research is a big enough headache in itself, but we now also have to consider recent developments in the so-called “right to be forgotten” legislation. To summarise what has happened to date: an individual in the EU/EEA has the right to request a search engine to remove links to information about them from search results generated by that search engine. Under the EU legislation, this affects any search engine that is based in the EU/EEA. It is up to the search engine to decide whether or not to comply with the request taking into account public interest as well the concerns of the individual. If agreed, the information remains on the original website but it is not visible to those viewing search engine results on a European version of the search engine. Until now there has been an easy way to circumvent the restriction, which was to use a non-European version of the search tool, for example Google.com. This is no longer possible for those of us identified as being located within Europe.

To comply fully with legislation, Google has announced that it will now “use geolocation signals (like IP addresses) to restrict access to the delisted URL on all Google Search domains, including google.com, when accessed from the country of the person requesting the removal.” See Google Europe Blog: [Adapting our approach to the European right to be forgotten](#) and Search Engine Land: [Google Agrees To Complicated Worldwide “Right To Be Forgotten” Censorship Plan](#) has summarised it thus:

“Assume that someone in Germany files a Right To Be Forgotten request to have some listing removed for their name. If granted, the censorship will work like this for searches on that person’s name:

- Listing censored for those in Germany, using ANY version of Google.
- Listing censored for those in the EU, using a European version of Google.

- Listing NOT censored for those outside Germany but within the EU, using non-European versions of Google.
- Listing NOT censored for those outside the EU, using ANY version of Google.”

The obvious way around this is to use a VPN or proxy server that gives you an IP address outside of Europe. For many people this will probably not be an option. Alternative search engines such as StartPage.com and DuckDuckGo may be another solution, but is there a problem anyway? Does it really matter if links to some stories about an individual disappear? Yes, it might. Some of my research work involves due diligence on companies and individuals, and on two occasions I have discovered information that had been excluded from European searches as a consequence of the right to be forgotten. In one case, the extra information was deemed non-essential in the context of the enquiry but in the other it was critical. For both, I found the information by searching directly the databases and sources that held the original data. The data is there but not accessible via a general web search tool, again highlighting the danger of over-reliance on Google et al.

It is no longer enough to know how to use advanced search commands. We also have to understand how the results are generated and manipulated, and the restrictions that may be imposed on the output. Knowledge of alternative tools and the relevant, primary sources is vital. It is not artificial but human intelligence that is needed in 2016 and beyond to find information and appraise it so that it is fit for purpose.