

IET WHITE PAPER

Cookies, fake news and single search boxes: the role of A&I services in a changing research landscape

November 2017



INFORMATION DISCOVERY IN THE RESEARCH WORKFLOW

In preparation for this paper the IET interviewed engineering professors, researchers, and librarians, asking them to share their thoughts on the way different online tools are used in the research workflow. There are many different aspects to research workflow including the desire for trusted information, fast access to information, the need to assess and validate sources, the serendipity of discovery, and the fact that researchers are often searching for highly technical and specific information. There are also a multiple of tools available to researchers to support their search.

Take the PhD literature search for example. When interviewed, professors and librarians agreed that more complex search of high quality resources would strengthen the work of their students, but felt that this sophisticated attitude to searching for information was one that had to be taught and learned. Some expressed frustration with the dominance of one search engine, Google, which has become the defacto search tool for both undergraduate and post-graduate students. But this frustration is not limited to the academic market: we learned how one particular engineering company had attempted to solve the reliance on general search engines by building its own bespoke database of high quality, trusted sources, complete with a single search box entry point. Their users *still* used the open web.

The ubiquity of general search engines was widely acknowledged. There was a common perception that students and early career researchers are too reliant on search tools designed for general and consumer use, and are not proficient at using the more complex, high quality scientific search tools and databases that are available to them. For example one participant commented that, when asked to create a search for their chosen topic, their PhD students would usually rely on the open web alone, until such time that they were given advice and guidance about other tools they could use to search.

Another noted that the method of searching for scientific sources was typically down to the experience of the person undertaking the search. While his students did know how to use different tools, they still started out with a generic search engine but went on to use the citations found to identify further sources in the databases provided by the library. Others simply typed a full article or publication title into the standard search box, presuming that results from Google Scholar would be returned, when in fact, this wasn't always the case.

As well as generic search engines, researchers are increasingly turning to scholarly research networks such as ResearchGate and preprint servers to locate relevant information. For one professor, this practice was born from the slow process when ordering articles through official routes at his institution. This led to the inclusion of preprint papers on personal publications lists and emailing authors directly to ask for copies.

Overall, the conversations with professors seemed to indicate a gap in the knowledge of early career researchers. While they can see that reliance on general search engines leads to information overload and low quality, irrelevant search results and they want to narrow down results and filter information, it isn't resonating with them that their library has tools that provide precisely this service. The idea that library resources are harder to use than a general search engine is an entrenched position, and a challenge for libraries to overcome.

THE VIEW FROM THE LIBRARY

From the librarian perspective, the experience is that researchers and students seem unwilling to take several simple steps to get information they need; the general search engine has created an expectation that a single search box and one-click to results will deliver all the necessary literature. This means that when researchers see a specialist abstracting and indexing service, they presume they are adding unnecessary steps into their workflow rather than crafting specific searches that will lead them to the most authoritative information.

Librarians naturally see it as their role to promote the best resources to their patrons and provide the necessary guidance and support. But students and researchers seem to have allowed a single search box interface to become their comfort zone and can find adjusting their behaviour very difficult. One respondent noted that while search engines do make searching for information much easier than it used to be, these tools also present researchers with a new hurdle to overcome, namely filtering the sheer amount of information they are presented with:

‘Early career researchers need to manage the redundancy of the material, and be sure that

it’s of sufficient quality. Take early stage PhDs. They start with a single search engine and drop the results down in front of you thinking that what they have found is all of a high enough quality. This is where peer refereed and specialist resources can help them learn the skills of research. Further down the line they’ll gain an appreciation of the quality of publication which is fundamental to a good PhD and gaining a deeper understanding of the subject.’

This new challenge of information filtering is an essential one, with even the most niche topics returning tens of thousands of results on the open web. Abstracting and indexing services add value by removing this step from the researcher. Instead, subject specialists have reviewed the material that is entered onto the database, to ensure that it is high quality, and then use their experience to apply relevant terms from a thesaurus or controlled vocabulary that help researchers find specialist information. Researchers who expect a single search to return the information they need may be making work for themselves by not taking advantage of the filtering expertise that A&I databases can provide. Meanwhile, relying on the open web is becoming more and more complex.



LIBRARY RESOURCES AND THE OPEN WEB

In recent years the principles that drove the development on the World Wide Web – access to information across international boundaries – have been increasingly eroded by commercial and political concerns. Students have become increasingly dependent on global general search engines as gateways to information, but at the same time these gateways are becoming less and less helpful.

The challenges for professional researchers using the open web include the way cookies are used and the increased level of personalisation in search results, both of which can make search less replicable; the skill of filtering information; and identifying the provenance of material they find.

The theory is that a search engine treats all information fairly, and returns the most relevant results, but this is not always something that happens in practice. The use of cookies on a machine to make a judgement about the relevancy of a topic means that there is no level playing field when researchers begin their research. Their internet browser will already have decided what topic interests them based on the search history of the machine.

Librarians expressed some concern around a perceived lack of awareness among students and researchers of how far search results are personalised. This group is also felt to be much less aware of the difference using scholarly databases, including A&I services, can make to their research and that searches carried out on generic search engines are much harder to replicate. Scholarly databases make search much more visible.

Of course, general search engines devote enormous resources to developing their algorithms so that they return good results, but broad definitions that are useful to marketing at

consumers don't necessarily work for a highly specific, academic, search environment.

Once they have search results, students and researchers must make a judgement about whether or not they trust what they are reading. It is here that the skill of filtering information is crucial. This is particularly true for post graduate research students, for whom literature search is an essential aspect of their projects.

Recent changes to the way online content is presented indicate that it may become increasingly difficult for readers to assess the quality of the material they are accessing. Services such as Accelerated Mobile Pages from Google and Facebook Instant Articles are designed to increase the speed of delivery, but they do this by stripping articles of any article branding. This means that readers have no instant, visual way of establishing the provenance of what they are reading. Metadata, including author, publisher and journal information is important to help researchers assess the relevance and quality of content.

This is especially a problem because the search engines and social platforms that dominate the web are for-profit enterprises that accept advertising, and may well return a result because the article and end-user are a good fit for their advertising criteria. These non-neutral barriers are becoming an increasing problem¹. An article that becomes popular through advertising can then retain a high ranking in organic results. This undermines the natural flow of academic research, an environment where one discovery or technique can lead to another, where theories are published and then overturned. Researchers value the serendipity of discovery in their

The theory is that a search engine treats all information fairly, and returns the most relevant results, but this is not always something that happens in practice.

1. Detours and Diversions — Do Open Access Publishers Face New Barriers? By Kent Anderson

https://scholarlykitchen.sspnet.org/2017/05/31/detours-and-diversions-do-open-access-publishers-face-new-barriers/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed:+ScholarlyKitchen+%28The+Scholarly+Kitchen%29

literature search and the articles they didn't know they were looking for are sometimes the ones that lead to new and fruitful avenues to explore.

All this means that the techniques that search engines and social platforms have developed to manage the sheer volume of digital information are unfortunately creating barriers to the provision of honest, relevant search results.

When students and researchers use an A&I database curated by experts in their field, they know they are accessing the latest, high quality

information. In recent years universities have invested considerable time and effort in ensuring their students have information literacy skills to cope with the open web, but the web is evolving even faster than these skills allow for. Some universities now explicitly teach the differences between general search engines and scholarly search tools, including A&I databases, to equip their students with the knowledge they need to make informed choices about the sources they are citing.

ABSTRACTING AND INDEXING; HOW PUBLISHERS ENSURE HIGH QUALITY INFORMATION

Libraries make significant investment in specialist resources, and so do publishers in the curation and provision of these resources.

For example, for over 50 years, content in Inspec has been selected on its scientific merit and perceived importance as determined by scientific experts. These experts then manually index the literature, using their own expertise of the subject to determine which terms and classification codes from Inspec's 20,000-term thesaurus and chemical, numerical and astronomical indexes

should be applied, even if the article text itself does not explicitly mention a particular term or subject.

As a result, Inspec helps researchers find content that other databases and search engines may not retrieve, simply because of the way the content has been indexed. This makes a search of Inspec more direct and powerful, filtering the literature on behalf of the researcher and delivering access to the most authoritative and relevant information available.

CONCLUSION

Digital publishing has always evolved quickly, but recent changes, including the consolidation of search, the increasing importance of advertising, and the sheer number of irrelevant search results that any search phrase will return, have created barriers for researchers looking for information.

In this changing landscape, filtering information for quality, relevance, and accuracy is becoming an increasingly important skill for students and researchers. But they do not have to manage the sea of information alone. Libraries invest in high quality resources for students and researchers to use, and A&I databases can play a vital role, allowing users to rely on the expertise of experienced indexers who have already assessed the information. As one librarian commented,

'Major A & I services provide trusted sources of information from quality document types -conference papers, journal articles, ebooks, book chapters, technical reports, and patents. The ability to search and refine by controlled vocabulary, author affiliations, source title, and International Patent Classification codes are some of the search facets that engineering researchers find extremely useful.'



Librarians want students and researchers at their institutions to be confident and skilled at dealing with content from different sources. Professors recognise that young PhD students may rely too much on a simplistic attitude to search, and want to see their students' research skills to develop through the first stages of their career.

Despite this appreciation that specialist search tools are beneficial, it seems that researchers are finding it difficult to change the behaviour they have learned when searching the internet as consumers. Incentivising them to use the resources that libraries provide may be about highlighting the indexing expertise they are accessing when they use these services. A more complex search may feel like an extra step in a

researcher's workflow, but using a high quality search tool actually removes the increasingly difficult task of filtering content for basic accuracy and relevance. Material accessed through A&I services can be trusted, is high quality, and holds subject specific information. While students and researchers may not see it at first, with only the most relevant sources to read, these services help make their research much faster.