Author's version of a chapter which appears in *The connected home: the future of domestic life*, ed. Richard Harper, Springer 2012. The original publication is available at www.springerlink.com.

The web, the home and the search engine

Stephen Robertson

When, in Eric Hobsbawm's account, the 'short 20th century' ended with the fall of the iron curtain in 1991 (Hobsbawm 1994), the revolution in information and communication technologies was just getting into its stride. On the back of such nineteenth century inventions as photography, typewriters, telephones, telegraph, radio, recorded sound, and the punched-card sorting machine, as well as popular publishing, the spread of universal education and services such as public libraries and the postal system, the first half of the twentieth saw a huge expansion of our information horizons. Broadcast radio, film and television became sources of information available to everyone, in addition to newspapers, magazines, and cheaply produced books. For person-to-person communication, the rapidly-expanding telephone system (increasingly based on automatic exchanges) added to the postal service which had reached its apogee around the turn of the century. In business, large scale data processing based on punched cards was making inroads into previously clerical domains. And in the second world war, the demands of the code-cracking community pushed onwards towards the computer.

At mid-century, the digital computer began its vast infiltration of our lives. At first limited to those domains where the punched card already held sway, and to a few arcane scientific endeavours (such as predicting the weather), it gradually found other niches. The symbiosis with typewriting was so complete that the electronic descendant of the old QWERTY keyboard, designed in the late nineteenth century on the basis of severe mechanical constraints, was simply absorbed to become the main method for humans to instruct machines. By 1991 the 'computer on every desk' was beginning to look plausible, and a computer in every home was not far behind; computer-like devices were being hidden away in many other gadgets. But more importantly, it had long been found useful to allow computers to talk to each other directly (rather than through the medium of human beings). The tentacles of the networks were spreading everywhere, and the reification of the global internet as the World Wide Web was just getting off the ground.

The last twenty years have seen the digital world (no longer just computers) expand to take over sound recording and many other sound processing tasks, image recording (photography) and many other image processing tasks, as well as all sorts of textual objects. Not only does your mobile phone contain a computer of sorts, the entire mobile world is inconceivable without computers and digital networks. And although many of us still buy and use paper information resources (books, newspapers, maps, invoices, receipts, scrap paper for notes to ourselves and others), all these are beginning to look increasingly archaic. A year, ten years? Yes, many of them will survive that long. A hundred? It seems unlikely.

Against this background, I would like to explore one particular aspect: the seeking and finding of information, by citizens in their everyday home lives.

Looking for information

Every day, we look for information. The communication-rich world in which we live offers us a thousand ways of receiving information, as well as transmitting it for others to receive. To simplify to a spectrum by choosing one variable that applies to the reception stage, we might receive purely passively, or actively seek out, or function at any point in between. If I read a magazine which deals with my favourite hobby, I am opening my receptors to a variety of information within a closely-defined domain. If I listen to the conversation from the next table at a restaurant, I am not pre-defining the domain in any way (at least in the sense of subject), though I am restricting to a very specific social situation involving specific actors. If I look up someone in my contacts list, I am probably anticipating almost exactly what I will find (a telephone number or an email address – I know exactly what they look like). If I allow myself to read the advertisements opposite me on the Tube, the domain is constrained only by the fact that someone has paid to place a message in that space – I really do not know what to expect, except perhaps in a statistical sense. If I have something in mind that I wish to find, I would not in general look there. On the other hand, immediately beneath the ads is the Tube map, which I may consult for very specific facts. In between the two, I might well scan my hobby magazine with the aim of picking up moderately specific ideas and information.

In this chapter, I will be focussing on the active end of this spectrum. That is, I am not concerned with the activity of looking at the advertisements on the Tube, or at a television screen, simply in the hope of being entertained. But as soon as we begin to direct our attention, to choose what to receive on the basis of some (vague or specific) idea of what we want to find, together with some (vague or specific) notion that what we want probably exists in the outside world, then we are moving into the realm of this chapter. In the world in 2010, among the many ways of seeking information, one form of device has achieved an extraordinary pride of place as a natural starting point: the *web search engine*.

This chapter is about how search engines have infiltrated our lives. It starts with some history. Search engines did not spring fully formed out of nothing with the invention of the Web; as so often with technology, they evolved from pre-existing kinds of systems. But again as so often, the course of evolution took some unexpected twists – unanticipated, that is, by the people who thought of themselves as agents of that evolution.

The Library

Let's start with a thought-experiment. Think of the last time you used Google, or whichever is your favourite search engine, and obtained some information as a result of using it. Now think what would have happened twenty, thirty, fifty years ago. How would you have gone about finding that same information then?

Of course, I don't know what particular information you had in mind, and a variety of answers might be appropriate. However, if you play this game a few times, one source of information that you are

likely to think of is your local library. In Neil Stevenson's novel *Zodiac*, published in 1988, the protagonist ST (a sort of eco-Sam Spade) is a very sophisticated seeker of information. One of his best sources is a librarian called Esmerelda, whose ability to locate relevant things in the library's archive of news material and elsewhere proves crucial to many of ST's cases. ST even thinks in terms of traditional library subject headings when he is commenting on what she does for him. What's surprising about this description is that only a few years later, a similar character would surely have found similar material for a similar purpose via a search engine, using a few well-chosen words.

The library is where this history begins. Libraries have existed for several millennia; they collect information from the world, and organise it in such a way as to make it accessible to people. All sorts of organisational tools and methods have been used, some due to the librarians themselves (catalogues and classification schemes), and some coming from further back in the chain, the publishers and authors (indexes and other forms of internal organisation). A librarian might be expected to direct an information seeker to an appropriate level of tool – for example, a question about an historical event might be best answered by a general reference work or an historical treatise, depending on the level of expertise and sophistication of the seeker.

One form of tool, developed specifically for the sciences, was the abstracts journal, such as Chemical Abstracts. This was a periodical containing abstracts (summaries) of all the scientific articles published in the subject in the previous month, organised under specific headings in a specialised classification scheme, and also indexed in considerable detail. Such a tool demanded some skills, first in its preparation and construction and then in its use in information seeking. Librarians and subject-knowledgeable information specialists developed those skills.

In my youth, library catalogues were usually on cards, and indexes were usually printed in books. But in the second half of the twentieth century, both tools were obvious candidates for the gathering computer revolution. Over several decades, beginning in about 1960, paper-based operations were replaced by computer-based ones. Chemical Abstracts is now a number of different databases hosting a number of different services, all computer-based. The print version ceased production on 1st January 2010.

Online searching

At first, the searching of a database of scientific abstracts was a tortuous process. In the late 1960s, a medical researcher could search Index Medicus (database of medical research papers) by sending off a request to the National Library of Medicine in the US, by post – that is, by what has become known as snail mail. The search would be formulated by an expert in a special query language, coded on punched cards, and run overnight, and the resulting printout returned to the user by post.

Through the 1970s and eighties, it slowly became possible for a user to search such databases online, on a computer terminal of some kind. But formulating queries was still a skill, and searching was often done with the help of a librarian. Around the same time, online library catalogues began to emerge. These catalogues would support the traditional library catalogue function of identifying a particular book held by the library (so-called 'known-item' search), with only very limited support for subject or topical search (the main function of the scientific abstracts systems).

Also at the same time, networks were developing and spreading. So even if the database you wanted to search was on a computer in a library the other side of the world, it might be possible to hook up your terminal to it via the network. By the late eighties, it was possible to search a large number of library catalogues and scientific and business abstracts databases in this fashion. However, it's worth noting a number of limitations, which might not be obvious from the vantage point of 2010.

1. Cost: the abstracts databases (if not the library catalogues) were typically very expensive to search. It was not something you would do on a whim. Long-distance network access, too, was expensive, until the internet became accessible to everyone.
2. Separation: each database had its own interface, maybe a little different or maybe very different from the previous one. Each one had to be searched separately, using knowledge of its idiosyncrasies.
3. Query language: Known-item searches might involve filling in a form (Author, Title, Date etc.). Subject searches would probably involve a complex query language.
4. Output: The result of a subject search would probably be an undifferentiated set of items, of arbitrary size. So if you formulated your query fairly loosely, it would result in a list of thousands of items, with no ranking or suggestion of where to start. If you formulated just slightly too tightly, the result set would be empty.

On this last point, while systems which ranked the most likely item first had been studied experimentally since the 1960s, the idea of ranking did not begin to penetrate real live systems until the very late eighties.

Words

The world of online searching began a trend which came as a surprise to the trained librarians. Most traditional library approaches to subject or topical searching start with a formalised scheme of classification codes, subject headings, or phrases from a codified indexing language. Each item has to be allocated by a librarian to one or more such formal descriptors. Then at search time, the searcher, or a librarian on the searcher's behalf, chooses which of these descriptors to look under. But as online searching of abstracts databases developed, it was found both feasible and moderately effective to rely on the words of the abstracts. If every word of every title and abstract is indexed, and the system provides a good way to search on combinations of words, the formal scheme and the effort of allocation are no longer strictly necessary. Already in the early 1970s, some such databases began to appear, although there was something of an ideological gulf between the proponents of word-based searching and the more traditional librarians. Word-based searching is commonly, somewhat pejoratively, referred to as the 'bag-of-words' approach.

This notion of searching on words in documents goes along with the idea of ranking the results. Different words differ in importance, both in documents and in queries, and it makes sense for the system to try to assess that importance and therefore the likelihood that a specific document is an appropriate response to a query. If many documents match the query to some degree, the system should guide the user towards the most likely items first.

A research field

As it happens, the notions of searching on words and ranking the results had been the subject of research since the 1960s. Search as a field of scientific research has a long history. The field was named *information storage and retrieval*, subsequently abbreviated to *information retrieval*, by one of the pioneers in the 1950s, and acquired a strong experimental tradition in the sixties. Word search and ranking continue to generate interesting theoretical and experimental work. In fact a major international initiative to advance the state of the art, known as TREC (Voorhees & Harman 2005), began in 1991 with significant US government support, and continues to this day.

A basic assumption of the TREC initiative was and is that many information resources becoming available in the world would not be curated by librarians, nor prepared as coherent databases with built-in provision for searching by publishers, but would rather come in the form of chunks of free-form text. The archetypes for such material are news collections (the texts of news articles derived from newspapers or from newswire services) and legislative material; but in fact some of the interest and funding came from the intelligence-gathering community, which was and is concerned with searching every kind of formal or informal document, including private communications of all sorts. In retrospect, and in particular in the light of the anarchic nature of Web publishing today, this is a very apposite assumption.

In 1991, of course, the World Wide Web was only just beginning. The notion of a Web search engine did not yet exist, though people were beginning to address the question of how to locate sources on the (then) internet. The basic hyperlink notion with which the Web started is itself a powerful device for finding stuff, but works only if you have a good starting point for your search. The scene was set.

Early Web search

Over the first half of the 1990s, as the Web itself began its meteoric expansion, tools to locate resources on the Web also started to appear. Not all were based on word indexing – for example Gopher (essentially a pre-Web technology) worked with filenames only. A little later Yahoo! used a more traditional librarian approach, having editors assign webpages to categories.

The idea of word indexing of the Web required another technology to be developed first: the crawler. This is a program which uses the hyperlink structure of the Web first: from a set of starting pages, each page is downloaded and analysed, embedded hyperlinks are identified, and then followed to obtain new pages.

Given a crawler, the found pages can also be indexed, using the word-indexing methods by now well-established in the information retrieval field. From the same field, search systems can be built to search these indexes. All of these components began to be used together in about 1994, although it took a little longer for web search engines to attempt to index *every* word on *every* webpage they could find. Currently, the big search engines mostly do index every word on a page, but not every page, because they have huge lists of pages that they have not yet visited. These lists are prioritised according to some measure of how useful the page is likely to be, but many low-priority pages never do get visited.

One can argue that the more words you have to describe a page, the better (at least, this can be argued provided that you have a really good mechanism for ranking the better pages more highly). One source of words to describe the page, other than the content of the page itself, is the relevant text from other pages which link to it. Each hyperlink has a piece of text, known as anchor text, which in some way describes the page it points to, known as the landing page. Anchor text is a source of words which in some way describe the landing page – sometimes a sort of summary or heading for it – and therefore may be used to index it. It turns out to be an extremely good source for web search purposes.

By the time Google came to dominance at about the turn of the millennium, these principles were well established (Croft Metzler & Strohman 2010). Currently the big search engines all do something similar, but with many tweaks and bells and whistles. At some level, the model used by all is the old bag-of-words model, although it has now reached a level of sophistication unguessed at by the originators of word indexing. In part, this sophistication comes from a process of adaptation, to which I return below.

The uses of Web search

The historical origins of the web search engines, sketched above, give little indication of the explosion of uses to which they have been put, and in particular, the uses to which home users – citizens in their everyday lives – put them. Already in the late 1990s, researchers began studying search engine use, and reporting results which sometimes startled them and their colleagues. I well remember being startled myself to learn how many searches were being made for celebrities such as Britney Spears, or had an explicit or implicit sexual connotation. Although these studies start from the actual queries – *what* words people search on – they can also be quite revealing about some level of intent – the *why* of search.

The idea of search which has dominated information retrieval research for most of its life, and in particular the model which the early web search engines borrowed from the abstracts databases, is what might be described as *subject* or *topical* search. That is, there is an assumption that the user is looking for documents *about* something, or documents which contain or convey certain information, or documents which answer specific questions. This may be contrasted with the *known item* model on which much library catalogue search is based, where the user is trying to establish whether or not the library contains a specific book that she knows to exist.

What emerges very quickly from an examination of search engine use is that these two represent only a part of the wide range of uses to which search engines are commonly put. Both these purposes are well represented among search engine users; but in addition, there are both fuzzy areas between these two poles, and also other rather different kinds of purposes. There is a three-way distinction between search types which goes some way towards a more comprehensive view (Broder 2002):

1. Informational: roughly the subject or topical query described above;
2. Navigational: trying to locate a page that you know or think or expect to exist (for example, the home page of a person or company; or a tax-return form);

3. Transactional: trying to *do* something, such as order a service or product from a supplier, or download a program or a piece of music or a movie.

Within these classes, and in the spaces between them, many variations are possible.

This variety itself came as something of a surprise to many information retrieval researchers. To take the known-item extreme: if I have visited a page before, possibly many times, I *might* remember its URL, and type it directly into the address bar. Or I *might* have saved a bookmark for it, so that I can click on my Favourites list to get back to it. But URLs are hard to remember, and bookmarking requires a conscious act at a time when I am thinking about something else. It is very likely much easier to do a search on a search engine. It may indeed be much easier to remember the one- or two-word search that will get me there than to remember the URL or even to find it in my Favourites list.

So this is exactly what many people do. For many real users, the URL is gobbledegook, and Favourites are a hassle, and why bother with either when the search-engine route is so much easier. As a result, search engines see a huge number of navigational queries, often repeated many times by the same or different users.

Indeed, for many users, there is little or no distinction between the browser and the search engine. If when I open the browser it goes straight to my home page, which is one or other form of a search engine front page, then I have no need to make a distinction. For many non-technical users, such a distinction would actually get in the way of understanding. The browser+searchengine is simply the single device that allows them (with luck at least) to get where they want to go.

Web search engines in context

So far, I have talked about search engines as a one-way process: they were invented, developed, produced, and offered to the users of the Web. But the interactions between the various parties involved in the Web in one way or another have been complex, and the developing notion of a Web search engine has been very much influenced by, as well as influencing, the rest of the Web world, including its users. These interactions deserve much deeper analysis.

In the rest of this chapter, I will focus on some of the interactions. These come into three categories:

1. The business model of the Web search engines: advertising.
2. Efforts on the part of website owners to get exposure: search engine optimisation.
3. Direct responses of the search engine owners to users and usage.

Advertising

Web search engines are a very profitable business, and they make their profits from advertising. The model on which most engines work is as follows: given a query from a user, serve up some relevant advertisements as well as the regular results of the search. (These regular results, coming from the search engine's own crawl of the web and the resulting index, are commonly referred to as the 'organic' search results.) (Wikipedia, *Search Advertising*).

A much exercised debate in the search engine community is the relationship between organic results and ads. Generally, the big search engines try to maintain a strong distinction, to make it clear which are the paid-for ads – though depending on the search, these may be given more or less prominence. Thus if the user's intention seems to be some form of shopping, to find a supplier for some product, then ads for this product might reasonably be regarded as good candidates for the user.

In an extreme, a search engine might make no distinction, simply promoting in the ranking paid-for entries. However, this is commonly regarded as underhand, and search engines seem to lose credibility by doing this – and the business model clearly requires substantial numbers of users who trust the search engine enough to make use of it. Hence the usual distinctions (position on the page, colour, style, and/or direct label such as 'sponsored site'.

The detail of the business model, and the mechanisms provided by the search engines for advertisers to place ads, are increasingly complex, and not directly relevant to the present discussion. However, some aspects are worth extracting and generalising (with a complete disregard for detail!).

First, the usual trigger for payment is clickthrough. That is, the advertiser does not pay on submitting his ad; he pays when a user clicks on it. (How much he pays depends on an auction process, whose details are not needed here.) This provides a strong incentive for the search engines to serve up relevant advertisements, as well as relevant organic results. In fact, advertising through a search engine is one of the most focussed forms of advertising available to many organisations today. An ad is shown only to those people who issue particular queries.

Second, many users of the Web and of search engines are in fact undertaking tasks to which ads may be relevant. A great deal of shopping, and/or preparatory work for shopping, is done via the web.

Third, despite both the above, only a small proportion of users do actually click on ads. As in so many other domains of advertising, the business model relies on having a large number of users, only a very few of whom will respond directly to any particular ad.

Fourth, in contrast again, there is evidence that the credibility of the search engine is affected by the relevance of the ads. Many users do look at the ads, and are less likely to trust the rest of the results if the ads are not relevant (Buscher Dumais & Cutrell 2010).

Web search engines ride a somewhat tricky line. On the one hand they are completely dependent for their existence on advertisers and advertising. On the other, they have to impress users with their impartiality – they are also completely dependent on maintaining a vast number of satisfied users.

Search engine optimisation

As the author/owner of a website, you probably would like many people to visit it. Whether it represents your CV, a hobby, a business, a charity; whether it is purely informational or can be used to order goods or services; whether it is very specialist or of interest to a wide range of people – in all these cases, it was almost certainly created with a view to people seeing it. And it is well known

that the single most important route for users to reach any of the billions of pages on the web is the search engine.

It therefore follows that it matters very much to the website owner how his site is indexed by the search engines, and for what queries it is likely to get both retrieved and ranked near the top. (It is also known that most users, most of the time, look no further than the top few ranked items when they search – it matters very little what is at rank eleven, and not a jot what is at rank 101.) So savvy website owners often try quite hard to ensure that their sites are well represented by the search engines (Wikipedia, *Search engine optimization*).

This process requires some skills, akin to but a little different from those of the librarians or information specialists discussed earlier. This requirement has spawned an entire industry, the SEOs or search engine optimisers. Quite unlike the search engine industry itself (megalithic, with a small number of large players) SEOs form a cottage industry comprised of a large number of individuals or small organisations. The most obvious client group for this industry are businesses whose web presence may be crucial to their survival, but at least some of the SEO ideas are pervasive and even influence home users, in ways which I will explore a little further below.

The motives for web presence in general, and therefore for search engine optimisation in particular, might range from the most lofty (I am trying to provide authoritative information on this medical condition) to the most base (I am trying to sell you pictures of nude underage girls). The tactics used have a similar range, from making sure the website describes itself well, containing text which is appropriate in quantity and quality, to what is commonly described as spam. In general the search engine providers encourage some level of optimisation activity, because it helps them present good results, but fight against spam, which has the opposite effect.

Most people are familiar with spam email. There are all sorts of forms of spam website. Because hyperlinks are so important to search engines, one common form of spam is known as link spam. If you can arrange for many other pages on the web to link to your own page, the chances of your page being returned for a search are greatly increased. This can of course be legitimate linking from related sites, but there exist many sites whose sole purpose is to provide many links to spam pages. Another form, term spam, is to fill your page with the kinds of words or phrases people commonly search on, perhaps with no regard at all for their relevance in a content sense.

The war between the spammers and the search engines is a continuing one, a sort of guns-and-armour-plating contest. The search engines try very hard to detect and avoid spam, and the spammers discover new techniques to circumvent the defences of the search engines. Nor is the problem likely to disappear – the spammers have a lot at stake, as well as the search engines.

Usage and response

The search engine industry inhabits a marketplace, and must of necessity pay close attention to the needs and requirements of its users. Over the period of the industry's existence, the engines have been adapted in many ways to this user context. An important feature of this adaptation relates to the mix of kinds of user, as well as of kinds of material and information available on the web. In particular, as the Web itself has expanded from serving mainly academic purposes, into a general-purpose source of all varieties of informational resource for all varieties of citizen, and as the search

engines have consolidated their role as the main portal to this general-purpose source, the needs, requirements and searching habits of home users have played a huge role in this adaptation.

Search engines (that is, the organisations that run search engines) adapt to their users in all sorts of ways. One is as follows: every so often, the queries received by the engine over a period will be sampled, and the results of each search carefully examined. Individual pages will be assessed as to their relevance to the query (or to what might have been the intent of the user in issuing this query). This data will be used to evaluate possible changes to the system; a modification which generally improves the result ranking by pushing up the good stuff is likely to be accepted (modifications are being tried out all the time, possibly affecting any component of the system at any stage). So good modifications are retained, and over a period the system evolves to be more effective. The same data may be used in a more direct way, to train some part of the system in a machine learning fashion. The component which ranks the results may be trained in this way, to promote good results nearer the top or the ranking.

This method of adaptation depends to some extent on the assessors being able to guess the possible intent of a user in issuing a query – which may indeed be quite obvious, but might be more subtle. But there is another piece of evidence which helps to get around the problem of interpretation: clickthrough (that is, what the user clicks on, on each search results page) is recorded. For some users, search engines even have access to extended information, such as sequences of queries issued by the same user, or the time spent on a page (did the user return immediately to the search page?), or where she went next. Such data can be used in various ways by the search engine.

At the simplest level, if the same query is seen many times from different users, and almost all users make just one clickthrough, all on the same link, then it is very clear that this link ought to be ranked first for this query, if that is not already the case. Thus if it is apparent from the clickthrough evidence that 90% of the users who type the query 'amazon' are interested in the bookseller (as opposed, say, to the river, or the female warriors of classical mythology), then probably the a link to the bookseller should come first. This may be learnt by the system in a number of ways, including purely automatic ones – which may not require any human being to notice the evidence or take action on it. In other words, some forms of adaptation are programmed into the search engines.

On another level, the same evidence also supports the observation that many queries are navigational or transactional rather than informational – these same users do not want to read *about* Amazon, but to go to the Amazon site, probably for shopping purposes. Such evidence informs the judgements of the assessors mentioned above, both in relation to the specific queries to which the evidence relates and in their interpretations of other queries for which there may be less evidence. It also informs the design of search engines in other ways: for example, knowing that many queries occur in a shopping-related context affects the ways in which advertisements may be presented, as discussed above.

In between, much may be learnt much about the kinds of things people are interested in searching on and the usage of language in searching. Many queries are or contain the names of people – often celebrities or people involved in entertainment. Many other queries are about entertainment or leisure activities of all kinds. Discovering a restaurant (or finding the phone number of a restaurant you know about), finding out what's on in any location in any one of a huge number of categories, booking tickets for the same, finding how to get somewhere, a recipe, the weather forecast, the

state of the traffic, opening hours, finding and booking holidays, and so on and so forth – all these are tasks for which we expect help from a search engine. If this list is compared with the kinds of search task that were in the minds of the original search engine designers, and arose from the usage of library-type searching methods, it is very clear how far the world has moved.

How we present ourselves

As I have indicated, the SEO industry serves (on the whole) business users of the web. But at some level the notion of presenting ourselves as individuals has been influenced by the same ideas.

If I have a Facebook entry, I compose a profile of myself for it. It my be more or less detailed, more or less descriptive, more or less accurate; but what it consists of is words (and maybe a photograph). In the back of my mind, I am more-or-less conscious of the fact that other people may want to find me, and that in order to do that they will have to use either the links from other people's pages, or the words. (Even if they know me, and know that I have a beard, there is as yet no known way of searching for photographs of people with beards.) So I know that the words matter for this purpose.

The words do not necessarily have to be descriptive in the usual sense. The name of the school I went to might be supplemented by the name of some secret society, or some code or catchphrase, that only my immediate friends at that school would recognise. Word-based search engines are at some level purely superficial – they care not a jot for meaning or sense. And we, the citizens, have not only got used to them, we have learnt to think like them.

Googling

Many commentators have noted how search engines have infiltrated our lives, in so many ways – the coining of the verb 'to google' is just one of many examples. The concurrent adaptation of search engines to the population of users of the web, and of these users to search engines as the single most important starting point for general information seeking, over a decade and a half, has formed a positive feedback loop with extraordinarily far-reaching effects. In the process, the library and other such resources seem to have been left far behind.

To over-generalise grotesquely, we (citizens all) have come to believe not only that everything (the entire gamut of information resources in the world) is available on the web, but that we can all find it, any particular thing we need, via a two- or three-word query typed into a box. To many people born after about 1993 (sometimes known as the Google generation, CIBER 2008), this view of the information world is the only one they have ever known, and they will have been introduced to it at home before encountering it in any formal educational context. Even when, as some of them filter into higher education and academic research, they find it necessary to use more formal, librarian-curated databases of research papers and other resources, they will and do carry with them assumptions about search derived from the use of search engines in the home and elsewhere.

In James Blish's novel *A Life for the Stars*, published in 1962 but set in the far future, the protagonist Chris has occasion to interrogate the Librarian of the city of New York about some matters of history and mythology. The Librarian is one of the City Fathers, who/which are machines. Chris asks a question, the sort of question one might ask a human being, rather than choosing a subject heading in a traditional library catalogue or index, or issuing a two-word query to a web search engine. The

Librarian nevertheless chooses a traditional-looking subject heading and constructs/reads an answer to the question, from its Wikipedia-like store of the sum of human knowledge. Then: '...the Librarian, which spent its entire mechanical life substituting free association for thinking, had a related subject it would talk about if he liked…' – and this turns out to be exactly the answer to the question he hadn't explicitly asked. Actually, that is a very good description of what web search engines do: give them some words and they will freely associate. Commercial pressures and adaptation have brought the art of free association with bags of words to a high point of utility, and to a central role in our lives as citizens, that are nothing short of astonishing.

References

Blish, James: *A Life for the Stars* (part of the *Cities in Flight* series). 1962.

Broder, Andrei: *A Taxonomy of Web Search*. SIGIR Forum, Fall 2002, Volume 36 Number 2 (http://www.sigir.org/forum/F2002/broder.pdf).

Buscher, Georg; Dumais, Susan; Cutrell, Edward: *The good, the bad, and the random: an eye-tracking study of ad quality in web search*, SIGIR 2010 (http://portal.acm.org/citation.cfm?doid=1835449.1835459).

CIBER: *Information behaviour of the researcher of the future*. CIBER briefing paper, University College London 2008 (http://www.ucl.ac.uk/infostudies/research/ciber/GG2.pdf).

Croft, W Bruce; Metzler, Donald; Strohman, Trevor: *Search Engines – Information Retrieval in Practice.* Addison Wesley 2010.

Hobsbawm, Eric: *The Age of Extremes: The Short Twentieth Century, 1914-1991*. Michael Joseph 1994.

Stevenson, Neil: *Zodiac*. Atlantic Monthly Press 1988.

Voorhees, Ellen; Harman, Donna: TREC: Experiment and Evaluation in Information Retrieval. MIT Press 2005.

Wikipedia: *Search Advertising*. (http://en.wikipedia.org/wiki/Search_advertising).

Wikipedia: *Search engine optimization*. (http://en.wikipedia.org/wiki/Search_engine_optimisation).