Search Less Find More: Redefining Search Environments to Accommodate New User Behaviors

Chan Ping Wah
Johnson Davasagayam
Ngian Lek Choh
National Library Board

Abstract

The National Library of Singapore (NLS) has been prototyping new ways of delivering content to its users. This is to enable easier access for users so that time-to-find is more efficient and effective. This article discusses possible models of approaching end-user needs without imposing organised search structures typically designed in library environments. Research points to the fact that serendipitous search executed through popular search engines like Google and Yahoo offer the most effective means of information discovery in an online environment. This paper focuses on how library resources can be deployed to enable easy retrieval so that users can search less and find more.

Keywords: Information discovery, information retrieval, user search behavior, library resources

Introduction

Efficient discovery and retrieval is at the core of most search or browse systems built within the context of the library. Discovery then translates to the retrieval of relevant articles/books commonly referred to as interlibrary lending and document delivery services (ILDS) if they are not digitally accessible or copyright protected.

Current approaches are designed premised on librarian intermediation to enable users locate their finds. These are supported by library designed browse systems that form part of the locator services provided by libraries. The effectiveness of library locator strategies have not been statistically established in Singapore. Users who are unable to locate their find in a library system often turn away without providing their feedback. It is often assumed that the information they are seeking for is not available.

The Internet, on the other hand, has inevitably become the first stop for information seekers as the content base is much larger, unsurpassed by any single library and the prevailing perception of a higher probability of locating their find. The *problematique* for libraries however, is in ensuring acquired collections are first visible and then made accessible via a retrieval process.

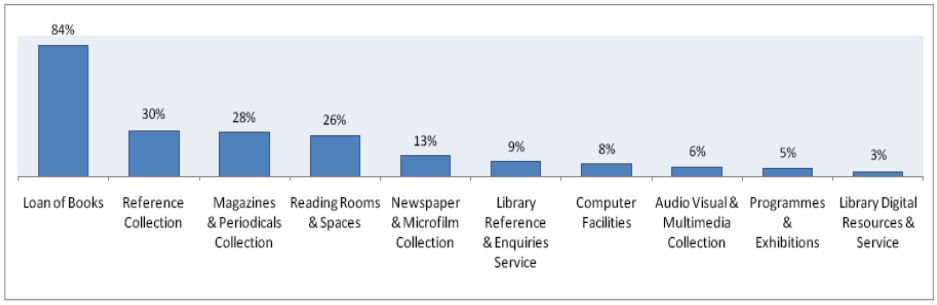
There are numerous studies on user search behaviour suggesting that when end-users look for data relevant to their information needs, they demand search efficiency (Buzikashvili, 2005). Lau and Goh (2006) identified that users are impatient in information seeking behaviour and the lack of perseverance lead them to terminate their searches rather than look beyond the first screen of results (hits).

A similar response was evoked when users have to deal with failed searches. Jones et al. (2000) defined a failed search as one that matches no documents in the collection. A survey on Nanyang Technological University (NTU) OPAC showed that a total of 317.840 queries returned zero hits that accounted for 49.5% of the total number of queries submitted. Users therefore have an equal chance of firing a search that returns no hits (a search failure) or one that returns at least one record (hit) (Lau & Goh, 2006). Existing browse and search systems surface the deep well of collections residing in library stacks somewhat inefficiently and inaccurately. This consequently impacts usage of digital library resources.

In a recent internal study by NLB on the value of libraries (Deloitte & Touche, 2009, p.109, 124), it was discovered that only 3% of the entire population use digital resources and services (see Figure 1) with 71% using less than 5 times a month (see Figure 2).

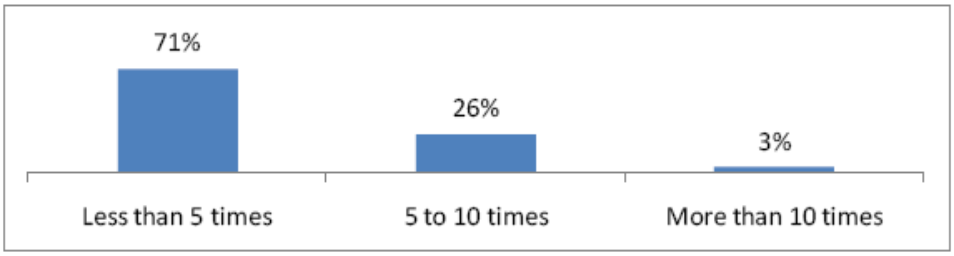
Additionally a significant 56% of the general population (Deloitte & Touche, 2009, p. 93) conveyed that they would not use Digital Library Resources and Services in the future (see Figure 3).

Figure 1: Proportion of Users who used Library Services



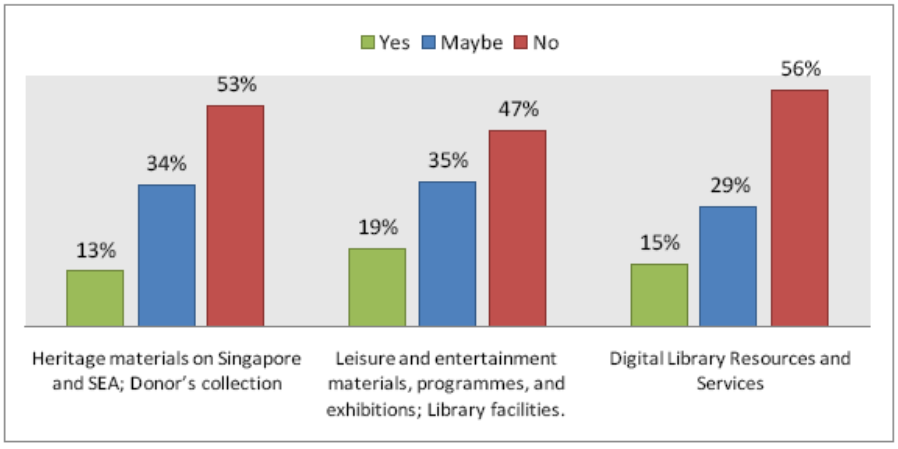
N=1,202 (General population; Onsite library users)

Figure 2: Number of Times a User is Connected to Digital Library and Services in a Month



N=38 (General population users; Onsite Library users – Digital Resource and Services)

Figure 3: Potential Future Use of Library Services



N=309 (General population non-users)

The above corroborates OCLC's (Online Computer Library Centre) 2005 search habits survey which highlighted that 84% of information seekers go to Internet search engines first, and only 1% use library online catalogues. This has grave implications for libraries and librarians. Hence a re-examination of our digital services is imperative to attract the 'M" generation of information seekers.

While our existing users are likely to continue enjoying our digital suite of services, non-customers are unlikely to be bought over. British Library's 2007 survey of researchers concluded that younger researchers demand instant information gratification and they use Internet search engines for broad scans before searching in subscribed databases: this however may not accurately reflect user behavior in academic environments.

The Internet also offers additional benefits to information seekers in providing social networking and collaboration tools with seamless access to content and notifications over mobile devices. These new media interactive platforms and devices offer added dimensions of access and delivery previously unimagined. It therefore became imperative in our redesign exercise that end-to-end search and retrieval experience were necessary. We also concluded that the access-delivery dichotomy has to be deconstructed to be able to reach out to a wider audience. User search patterns and behavior, both online and onsite, offered significant insights about new strategies for effective content search and discovery.

Revisiting User Search Behavior

User reliance on library catalogues as a guide to effective search is no longer the dominant logic in information seeking environments. Anecdotal evidence strongly suggests that trusted and organised archives like libraries are not necessarily the first stop for information seekers, much less their last stop.

The author (Ngian) had communicated to her 18-year-old daughter about the National Library's massive digital resource collection and digitised newspapers with finding aids like pathfinders. This was with the intent to excite the latter and her friends to make use of the extensive digital collections that was publicly available. Even marketing brochures were given for distribution in the school of the 18 year old. Months later, the author had sought feedback on the use of digital library resources online only for her daughter to remark "We can get all our information from the Internet and we know how to authenticate the content by comparing the information with more

than one source"! Moreover she felt that it was not convenient to log-into NLB's digital library's space while having to perform "Google" searches and to communicate online with her school colleagues on their project.

Even in the university library in Berlin, students have been reported to use Internet to search for information, even though the library has acquired a good collection of digital resources. In a survey done by the university, undergraduate and post-graduate students indicated that Internet search is the most convenient and commonly used. (Unpublished trip report, Berlin, 2008). Defining and understanding information search environments are important for libraries to ensure effective use and retrieval of digital resources. Three predefined environments shed light on how search environments can be best structured for public users in the National Library of Singapore.

The Three Approaches

The Restaurant Model

The first model is likened to that of the restaurant that offers a menu driven approach. The user selects and the order is served, very much like an order made at a restaurant.

In this approach, libraries acquire and organise content; build delivery platforms and subsequently market the service. This traditional content delivery approach is very often confined within the library space. The model assumes that the content that the user needs is available or can be acquired by the library. The user has options and is fairly clear about the information that is sought.

It is a "patrons demand, libraries deliver" model. Librarian intermediation is required in organising content for patrons to make informed choices, retrieve them if they are not readily accessible and manage rights associated with the distribution and re-use of the content. For such a service, the libraries will usually build a platform to respond to enquiries and deliver on-demand. The process involves requestor's authentication and entitlement, delivery channel and payment mechanism. This is often the most expensive part of the made-in library content delivery approach.

To design a seamless service delivery, libraries invest in elaborate technology driven platforms. The cost of building such an infrastructure can be quite

prohibitive. The model is almost always successful if there is a ready and voluminous demand for the service, or if the value of the content delivered is significantly high. The British Library's Document Delivery service (Inside Web) is a case in point. Inside Web contains over 20,000 leading research journals plus access to over 400,000 conference proceedings and other periodicals. The system also delivers British Library's legal deposit collections. The breadth of content available for search is necessary for the approach to work successfully. Libraries with much smaller archival content base will have limitations in executing a full-fledged search and deliver service based on this model.

The Fishing Model

The second model christened the fishing model, postulates the open seas, with the fisherman (information seeker) knowing the kinds of fish in the sea and dropping his bait. He then filters the find and discards what is not useful.

In this approach, libraries will have to adopt the "long tail" philosophy: ensuring that all content is made available; making sure that they can be retrieved by users ("Find Me") and access is affordable (Anderson, 2004). The user operates from a platform with tools and utilities navigating the sea of information. With search (Google, Yahoo and MSN) and social network spaces (You-Tube, Flickr, Wikipedia, My-Space, Face-Book, etc.) the user extracts information from libraries that have made their catalogue and content visible on Internet platforms.

While the library space is still a popular place for authenticated information, it is clear that the popularity of the Internet user spaces is rising rapidly. Libraries cannot underestimate the reach of these spaces to Generation "M" information seekers. Surveys demonstrate that libraries have to make deliberate plans to leverage on them or face the risk of losing their users (Rainee, 2009). Users are moving towards peer-to-peer sharing of information as it is the case online in "scribd.com" and "slideshare.net".

Information seekers are part of social networks where "good enough" content is sufficient. These are open spaces, less regulated and peer authenticated but it offers timely and "useful-to-me" content value as against "trusted" and dated information. Libraries therefore need to "sprinkle" packaged content into these user spaces with the help of new technologies. Netizens who inhabit these spaces will serendipitously stumble on these packages.

They do not need to migrate from their existing spaces to another platform to search for information. In other words, libraries and librarians can become part of social spaces online. Social netizens conversely share and exchange content with libraries, enriching the value of repositories that are exposed in social spaces. Libraries then repackage user generated content and replenish the sea of information with new content. Libraries and librarians can manage the information ecology in a more constructive, engaged and holistic way.

There are added advantages of leveraging on the e-spaces provided by other service providers. This strategy avoids costly marketing campaigns as the "Google", "Yahoo" and "MSN" (gym-able) engines provide the visibility and more importantly, global access to local content.

In embedding content in social networking sites, owners of social spaces have vested interests in keeping their spaces vibrant and attractive. Membership and traffic have a viral effect on users. Community owners build the appropriate capacity to handle traffic and user management, and their interactions. Penetration into new communities takes nothing more than membership into these sites by embedding content.

Secondly, social networks are naturally self-seeking and will target its own audiences hence generating intensity of use. The cost of generating intensity through promotion is significantly reduced. Hence libraries can cast their net as widely as possible to acquire varied customer segments, without exponential increases in operating cost.
The scalability of customer reach can be realized in this model of service delivery. "We build you come" is then substituted with "Netizens lead, Library Follow" model, where the user drives the information environment. Libraries invest their content in the most frequented e-spaces to generate highest access. The approach offers and efficient eco-system of information delivery to users where they are, incurring minimal overheads. Libraries can then focus on creating useful information packages targeting at different user groups.

NLB had adopted this approach quite extensively in the past two years, primarily in the dissemination of e-content. Singapore content is embedded in the Google space, in order to reach more users.

For many years, NLB has been compiling encyclopedic, byte-sized information about Singapore called Singapore Infopedia. Each entry in Infopedia comprises a factual brief with an appended list of related reference resources. The intent is to encourage users to read more about the topic, after

browsing the facts. By November 2006, NLB had compiled 1,000 records that were meta-tagged and mounted on its corporate website. (www.nlb.gov.sg) It generated 200-400 accesses a month initially.

In November 2006, a microsite was created to expose the content for Google to crawl and index them. From November 2006 to March 2008, NLB found that the usage increased from 400 a month to 30,000 to 150,000 in March 2008. The encyclopedic records have increased to 2,000 with an article viewership of nearly 1.6 million per year. Customers can comment on the articles and resident hyperlinks lead the user to related articles or content within NLB's depository/site. Similarly NLB is present in authenticated librarian group-sites that are gym-able. For example, topical information search guides are uploaded in "libguides" (http://www.springshare.com/libguides/index.html) for users to locate NL content without having to visit the NLB site. Users are then led to NLB's website if they want more content.

A similar approach has been undertaken to expose NLB's database of images. Apart from making the images available to Google, they are also exposed to Flickr to engage our users. Currently newspapers are being digitised. The intent is to make them available to Internet users with a built-in charging mechanism. More of the NLS' local content are packaged in micro-sites and exposed to Google and other Internet sites for users to find and locate library content. When users locate the library's content in the open seas, they are gently led back to the NLB digital library with the hope of exploring and discovering other library e-resources. NLB works towards finding more channels to reach its users, and exploit search engines and Internet spaces to allow users to find its content.

The limitation however is with copyrighted, subscribed content and the permissions required to make content visible and searchable on the open Internet platform. The recourse is to deposit summary content or at its worst, just the metadata, abstract or first 50 words of the articles for access on the Internet. Such information will allow netizens to review and request for document delivery.

In this approach, the delivery process must be kept as simple and convenient as possible. Where local content needs maximum visibility, this approach will sufficiently reach its audience with minimal investment. There are however hidden costs associated with this approach. Back-end support to customise and package content for access through public search engines incurs manpower cost and technical expertise. Sufficient understanding of search engine retrieval methodologies and experiments with Google Apps would be required

to enable this to happen cost effectively. Commercial enterprises are known to invest in huge sums of advertising dollar to become visible on online search platforms or to be embedded in social networking sites.

The Franchise Model

One of the key customer segments the National Library serves is the policy and enterprise research community. The content that researchers need is usually specialised and domain specific. Researchers usually confine themselves to their own circle of peers and research interests and utilise institutional rather than public resources for research. Dedicated platforms are not as useful as they would expect to find what they want from the Internet or from their own institution.

The franchise model proposes the use of API services (e.g. Web Services) to embed content in localized and institutionalised spaces. With this, researchers at their respective institutions do not need to navigate across different portals to request or access content. Instead, within their own portal, they will be able to search and access content from the National Library or from other external organizations. Access and user privacy are protected within their relatively autonomous portals or digital spaces, yet work together to meet the information needs of researchers.

One of the challenges for the National Library is to tier content access and delivery services for both the general public and research community in a centralized and convenient platform. However information seeking behavior and subject matter depth are fundamentally different for the two distinct communities requiring different search engines to cater for general users and researchers.

The compromise is to enable research communities to deploy their customised search engines on content from the National Library or other public domains. The National Library needs only to develop APIs that can be reused by as many research institutes and communities. To the user, the combined content delivery service is presented as one integrated service, rather than from two different service providers. This is currently available to subscribed users of the business community (www.ebis.sg) of the Singapore Business Federation where content is embedded on their website.

Since collaboration happens “behind the scenes”, the marketing of the service and content is fronted by subscribed communities to their customers. These

communities have vested interest to make the portal attractive to their own customers. The originating library only needs to market their delivery service (via API) to these research communities. While this is fairly new in application, another example of this can be found in a collaborative book club portal called "Bookjetty" where the NLB online catalogue is embedded on their site. On the average, there are about 1,000 accesses to the online catalogue per week. By offering the online catalogue to "Bookjetty" users via an API service, the cost of the new acquisition was negligible while enabling NLB to extend its reach to "Bookjetty"'s clientele with relative ease.

The strength of this approach is in its simplicity in embedding content via APIs and widgets. Many commercial sites already deploy this technology as part of a larger content syndication strategy. The franchise model requires a different modality as content syndication require approval from content providers first. Portal owners too may impose restrictions on content and accessibility.

Conclusion

Using the flexibility of modern information technology (IT) systems allows for all three approaches to be executed from a centralized or federated digital platform. While NLB is prototyping the integration of these approaches on a single digital platform, the approaches in themselves are evolving. Systems in themselves are feasible, however it remains unclear whether they can also be regarded as economically and legally feasible options.

The unknown issues revolve around the size of demand for these approaches, and the agreements that libraries and publishers would enter into to regulate content distribution, accessibility and re-use. The reinvention of the information cycle would need libraries to revisit copyright provisions not only in their respective legislations but also of those in destination countries.

Last but not least, the notion of relevance is at the centre of information retrieval. Approaches towards making content visible and accessible in user environments should not compromise document-query relevance as the lack of it may pose a distraction to users who require efficient retrieval. Positioning content in significant user spaces and public search engines needs to also ensure quality of retrieval by end-users in these spaces.

References

Anderson, C. (2004, October). The long tail. *Wired*. Retrieved from http://www.wired.com/wired/archive/12.10/tail.html?pg=5&topic=tail&topic_set.

Brazier, C. (2007, August 19). *Resource discovery at the British Library: New strategic directions*. Paper presented at the 73rd World Library and Information Congress, Durban, South Africa.

Buzikashvili, N. (2005). Information searching behavior: Between two principles. In F. Crestani & I. Ruthven. (Eds), *Lecture Notes in Computer Science: Vol. 3507*. Information Context: Nature, Impact, and Role (pp. 79-95). Berlin: Springer. doi: 10.1007/11495222_8

Chan, P. W. & Ngian, L. C. (2008, May 8). *Back to the future: Library competencies for Library 2.0*. Paper presented at the Library Association of Singapore Conference 2008: Innovate to Serve, Singapore.

Chan, P. W. & Ngian, L. C. (2007, October 31). *When is a library not a library? When is ILDS not ILDS?* Paper presented at the IFLA Interlending and Delivery Services Conference, Singapore.

Deloitte & Touche. (2009). *Provision of consultancy to conduct the return of investment study for the National Library Board*. Unpublished report.

Jones, S., Cunningham, S. McNab R. & Boddie S. (2000). A transaction log analysis of a digital library. *International Journal Digital Libraries, 3* (2), 152-169.

Lau, E. P. & Goh, D. H. L. (2006). In search of query patterns: A case study of a university OPAC. *Information Processing and Management, 42*, 1316-1329.

Ngian, L. C. (2008). *Learning from 2nd IFLA Presidential Meeting and library visits*, Berlin, February 2008. Unpublished paper, National Library Board, Singapore.

OCLC. (2005). *Perceptions of libraries and information resource*. Retrieved from http://www.oclc.org/reports/ 2005perceptions.htm

OCLC. (2007). *Sharing, privacy and trust in our networked world*. Retrieved from http://www.oclc.org/reports/sharing/default.htm

Rainie, L. (2009, March 30). *Friending libraries: Why libraries can become nodes in people's social networks.* Paper presented at Computers in Libraries conference, Arlington, VA. Retrieved from

http://www.pewinternet.org/Presentations/2009/8-Friending-libraries.aspx

Research Information Network. (2006, November). *Researchers and discovery services: Behaviours, perceptions and needs*. Retrieved from http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/researchers-and-discovery-services-behaviour-perc.

About the Authors

Chan Ping Wah, Senior Director, Research and Service Innovation, National Library Board, Singapore
Email: pingwah@nlb.gov.sg

Johnson Davasagayam, Deputy Director, Publishing and Research, National Library Board, Singapore
Email: johnson_paul@nlb.gov.sg

Ngian Leh Choh, Director of the National Library and Deputy Chief Executive, National Library Board, Singapore
Email: lekchoh@nlb.gov.sg