Boxing clever: how searchers use and adapt to a one-box library search

Dana McKay
Swinburne University of Technology Library
John Street Hawthorn VIC 3122, Australia
dmckay@swin.edu.au

George Buchanan
Centre for HCI Design, City University
Northampton Square, London EC1V 0HB, UK
george.buchanan.1@city.ac.uk

ABSTRACT
One of the major problems users experience searching for information in libraries is the number of places they have to search. It has long been posited that a single search box (like Google) that searched a range of library resources would solve these problems and make users more effective information seekers in libraries. In this paper we use log analysis to compare user search behaviour in a single search box system with that in a traditional library catalogue. We discover that behaviour varies in response to the results produced by the different systems.

Author Keywords
Libraries, information seeking, search, transaction log analysis.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
To have a search that works like Google—offering a single, simple search box that searches a wide range of content and returns neatly relevance-ranked results—has long been a goal for libraries. Evidence shows that users are often confounded by libraries' current fractured and siloed offerings, and the simple alternative that Google presents is both alluring and readily accessible to users of even academic libraries (Buchanan 2005, Griffiths 2005, McKay 2011).

Early attempts to provide users with a single search box focused on federated searching, where queries entered into this single box were performed automatically in the background on a number of search tools. Results were then amalgamated and presented to users. While this approach offered users a single point of entry to a range of information resources, the relevance ranking of results was notoriously poor, and the search process noticeably slow (Hane 2003). Nonetheless, many users preferred the convenience of federated search to searching a range of library resources, each with their own interfaces and peculiarities (Gore 2008).

More recently users' frustrations with traditional library catalogues and journal databases have led to the development of a new generation of library search tools, referred to as discovery platforms or web-scale search. These tools provide many Google-like features, such as spelling correction and relevance ranking, but their largest advantage by far over traditional library systems is the opportunity to search a wide range of library content in one place. Unlike metasearching, searches in web-scale search are conducted over a single, unified index, meaning that the speed and relevance ranking problems of metasearch are greatly ameliorated.

It would seem self-evident that this change would improve library users' experience of search, however results of early studies are far from conclusive, showing that users struggle with the move away from book-focused search (Way 2010) and that the large number of search results (and, concurrently, poor relevance ranking) in web-scale search is confusing to users (Gross et al. 2011). What is missing from the literature thus far, though, is an assessment of how users' search behaviour changes in response to web-scale search, and how the results searchers get change with its introduction.

In this paper we present a transaction log analysis of initial searches from an academic library homepage before and after the implementation of an underlying web-scale search tool. Prior to the change users could access a web-scale search tool with the click of a radio button (see Figure 1), whereas afterwards it was default. This change means that web-scale search was not a completely novel concept to library users, however the change did dramatically alter the number and nature of results users saw in the default search.

Firstly, we will give a literature background to our study, second we will present our methodology. Thirdly we will present our results, followed by a discussion of the implications of these results. Finally we draw conclusions from our research and offer suggestions for future work.

BACKGROUND
The background for this study can be broken down into four broad themes: typical search behaviour; searching library resources; libraries and Google; and web-scale search platforms.

Typical search behaviour
One of the most reliable tools for observing search behaviour in the wild is transaction log analysis. While transaction log analyses cannot tell us anything about users' experiences or emotional responses, they present a clear and accurate picture of behaviour because they allow us to examine how users search when unobserved.

Search log analyses have demonstrated that in both search engines (Bandos et al. 2002; Jansen et al. 2006) and in

OZCHI'13, November 25–29, 2013, Adelaide, SA, Australia.

digital libraries (Jones et al. 1998; Mahoui et al. 2001), users enter short, simple queries, do not change defaults, and do not use advanced search techniques.

These findings also generally hold true in library catalogues (Wallace 1993; Cooper 2001; Lau et al. 2006): users conduct short, simple queries and do not (usually) change search defaults. The same literature demonstrates these strategies to be broadly unsuccessful; why then do users persist in using them? One answer may be that users are notoriously poor at constructing more complex queries (Bandos et al. 2002). An alternative, raised by our earlier work on questions asked in bookshops, is that users do not have a clear grasp of the metadata needed to construct more advanced queries (Buchanan et al. 2011), a finding reinforced by work that demonstrates students and academics alike struggle to create citations (Aronsky et al. 2005; Yu et al. 2006). A third possibility is that users never develop any expertise in catalogue searching because catalogues' limited indexes and the ready availability of the internet mean that users abandon catalogue searching relatively quickly (Fast et al. 2004; McKay 2011). One final possibility is that the extra effort simply isn't worthwhile for users. This hypothesis is supported by our own earlier work on catalogue searching (McKay et al. 2011); the central finding of which is highly atypical: 40% of users did change search defaults. In that interface the default was visible and easy to change. Changing from the default setting produced positive results for users: they were able to (and did) find the items they were looking for. Contrast this with (Lau et al. 2006) where a "desirable" behaviour (longer queries) produces fewer search results. Perhaps, then, the reason users persist in short, simple queries is that although ineffective, short queries represent the best trade-off for users in terms of items discovered relative to effort expended in discovering them.

Finally, even assuming a search yields results, users typically do not look beyond the first page of search results (Jansen et al. 2000), so search tools supplying more than one page of results must have excellent relevance ranking. In the absence of good relevance ranking, too many results are as confusing as none at all (Khoo et al. 2012); the number and ranking of results is central to search user experience.

Searching library resources

Within library systems, locating specific resources, for example books or articles, is notoriously difficult (George et al. 2006). Library resources are stored in silos (Gore 2008), and users assume (incorrectly) that any search box on a library homepage searches the resources they want (Swanson 2011). The library homepage is unlikely to help them correct this error, though—even library webpages are prone to usability issues (Swanson 2011).

Library catalogues have been the traditional first touchpoint for library users seeking information, despite the fact that they are notably unusable (Borgman 1996; Lau et al. 2006). Catalogues typically require users to describe the items they seek in very precise ways: knowing specific metadata and using a controlled vocabulary (Borgman 1996). As we saw in the previous subsection, this is too much to ask of most users. Even if library catalogues were the most usable of information systems, though, they would still be unlikely to meet users' information needs: they provide a path predominantly to books (Cockrell et al. 2002), which (especially in an academic library) comprise only a small portion of use (Talja et al. 2003).

Finding articles is no easier for the average library user: many of them do not even know where to begin looking (Cockrell et al. 2002). Even if users do know where to look, all of them except library science students will sacrifice quality for convenience, choosing metasearch over individually searching databases (Griffiths et al. 2005). When users do search for articles, they typically search in the same way they search other information systems: with short, simple queries that return unmanageable numbers of results (Nicholas et al. 2005).

A broader issue with the silos in which library resources are stored, however, is the mismatch with human information seeking behaviour. Identifying resources is a late stage in two major information seeking models (Marchionini et al. 1993; Kuhlthau 1999), and yet siloed resources currently force users to preselect for resource type when searching. Many users will have a clear idea in advance whether they want a book or an article (Talja et al. 2003; McKay 2011), but library resources are even more fractured than this; the library may not have access to books a user wants to read, and articles may be in any one of a plethora of databases.

Traditionally when users failed to find what they were looking for in library resources, they turned to a librarian (Nordlie 1999), In today's online, always-on world, however, users expect to be able to access resources as and when they need them (Fried Foster et al. 2007). In the face of unusable library systems these days, library users will satisfice (Agosto 2002), give up (Nordlie 1999), or turn to Google (Fast et al. 2004; Buchanan et al. 2005).

Academic information seeking and Google

Whether or not librarians or the academe consider it an acceptable strategy, students and academics alike use Google in their academic information seeking (Fast et al. 2004; Buchanan et al. 2005; McKay 2011). Some argue that this is 'lazy information seeking', and that over the long term it will lead to poorer quality academic thinking (Brabazon 2006). Others posit that it is simply the way of things, and that the only way to combat it is to offer better information interfaces in libraries (Bell 2004). A recent study was able to identify the key ways in which Google outperforms traditional library interfaces (Khoo et al. 2012); these include features like smart searching and 'did you mean' typographical error correction. By far the two most important features to users, however, were the single search box that searched 'everything' and the excellent relevance ranking Google offers; without this relevance ranking search tools were perceived to return 'too many' results. Given the desire to search everything, web- scale search seems a self-evident improvement to traditional library offerings. Early reviews in the literature, though, suggest that relevance ranking may be a problem in web-scale search systems.

Web-scale search

Early usability studies of web scale platforms touted them as broadly usable (Gross et al. 2011; Slaven et al. 2011; Foster et al. 2013). On closer examination of the literature, however, these results are not so straightforward. One of the aforementioned studies used only information and computer science students as test subjects; as we noted earlier, this group are unlikely to be representative (Griffiths et al. 2005). Even this group, however, commented negatively on the number of search results returned and their difficulty locating specific resource types, a finding echoed in (Slaven et al. 2011). The 'too many' results problem is supported by studies showing frequent search limiting behaviour (Foster et al. 2013) (Ballard et al. 2011); as seen in the previous subsection this means that the underlying relevance ranking in these systems is not supporting user goals or is mismatched with users' knowledge and skills.

Log analyses of web-scale search are fairly limited; nevertheless they demonstrate similar patterns of searching behaviour to log analyses of other systems: short, simple queries (Meadow et al. 2012). The evidence as to whether having everything in one place is beneficial to users is ambiguous: an early analysis shows users going out of their way to use an old catalogue (Way 2010); but in a more recent study users accessed a variety of resources via the web-scale search (Lown et al. 2012).

None of the aforementioned studies, though, compare user effectiveness in a traditional catalogue and against a web-scale search, nor do they leverage search results to attempt to understand user experience. What remains for this paper, then, is to examine the specific impact that the introduction of web-scale search had on search behaviour, and to understand that impact by examining search results. This is an approach that is thus far unique in the literature.

METHODOLOGY

As seen in the previous section there is a long history of using search log analysis to understand user behaviour in library catalogues, digital libraries and the broader internet. As such it is a logical and well-tested approach to examining user behaviour in web-scale search systems, especially in comparison with other search tools.

We begin this section with a description of the search tools from which the data was collected, outline our data collection and sampling, and finally report our analysis.

Search interfaces

The library where this work took place is situated in a small but research-active university in Australia. The library is well used, both online and in person.

The search interface from which data was collected changed slightly over the course of the study. In 2012 it offered the library catalogue as the default search, which primarily targeted books. Alternative search tools—web-scale search, EBSCOhost (a large journal article database) or Google Scholar—could be selected using radio buttons (see Figure 1). In 2013 the standard library catalogue is no longer available and the search box defaults to a web-scale search (see Figure 2). Users have the option to search all metadata (this is the default setting) or they can elect to search titles or authors only. Google Scholar and EBSCOhost are available via links.

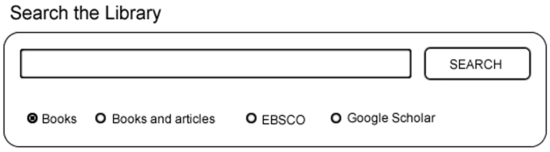

Figure 1: homepage search interface 2012: catalogue search default

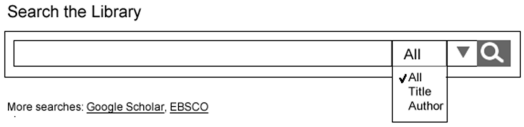

Figure 2: homepage search interface 2013: web scale search default

We will not identify the underlying search technology in this paper, as our findings likely apply to any web-scale discovery platform. While small details may differ by specific platform, the underlying differences between web scale and traditional catalogue searching will not.

Data collection

The search box on the library home page logs a number of data points for each search: the query terms, the date, the time, which search target was selected (in 2012, i.e. the catalogue, discovery platform—also referred to as web-scale search—EBSCOhost, or Google Scholar), search options (in 2013 only, i.e. all metadata, title search or author search) and if the user was on or off campus.

Data for this study was collected on four days, two each in the first semesters of 2012 and 2013. The first date for each year was in the second week of term, the second date ten weeks later near the end of term. All data was collected on a Tuesday, which is one of the busiest days of the week for the library for both in-person and online visits. The first collection in 2013 happened three weeks after the default search was changed to web-scale search and the interface changed to that shown in Figure 2. Our two-per-term collection was to determine whether there was any change in behaviour as users (presumably) became accustomed to web-scale search. Finally we examined library statistics to see how many users followed links to search Google Scholar and EBSCOhost.

Data analysis

We analysed the data in two ways: calculations over the entire population, and close scrutiny of six sample sets.

For the entire population of data, we were able to automatically calculate how many people selected each search target or option, and how many words they typed in. We then compared the differences within and between years for the types of resource targeted in searches. We also compared the number of terms in each search, within and between years, for different target resource types, and for different search tools.

We randomly selected six sample sets of 100 queries each to examine more closely. Four sets were comprised of

searches that used the default search setting on each of the four days in our data set. The two other sets drew searches from the combined search (which was not the default, and had to be deliberately selected by users) on each sample date in 2012.

Each of these queries was coded according to whether it was for a known-item search or a topic search. If the search was known-item, the metadata included in the search terms (title, author, date etc.) and whether it was a copy-and-pasted citation were noted. If the search was not for a known item, it was noted whether the user typed in a phrase, a question or discrete query terms. For all queries the presence or absence of search operators (and whether they were used correctly) was noted, as was the presence or absence of typographical errors.

Finally, search results were checked in the relevant system (catalogue or combined search) and results encoded. For each query the number of search results was recorded, and for known-item searches the position at which the item first appeared in results (if it was on the first page, otherwise the search was deemed to have failed). Each zero-results search was recorded, along with whether a 'did you mean' spelling correction was offered. By encoding search results we can determine how successful users' search strategies were, and in turn how well the systems tested are meeting their needs.

Once searches were encoded they were analysed for any patterns of behaviour. As much of our analysis was of populations with three factors, log-linear analysis was used. While the chi-squared test is more common, it is poor for three-factor tests. The G^2 value produced from log-linear tests is comparable with χ^2. Log-linear analysis allows testing for more detailed effects, just as Tukey's HSD test is used with ANOVA, to determine the factors correlated with true differences, and those that are merely superficial effects caused by other factors. The results of our analysis are presented in the next section.

RESULTS

In this section we first present our analysis of our sample searches, which we classified according to a number of parameters, then we discuss the changes in search results seen for these queries. Finally, we examine characteristics of the global data to assess the validity of our findings.

Metadata	Mar. 12	May 12	Mar. 13	May 13
Title	43	46	46	34
Author	23	18	15	17
Date	7	6	5	9
Publisher	3	2	1	1
Source	2	1	5	2
Issue	4	4	9	1
Pages	1	0	4	1
Keyword	49	48	49	58

Table 1. Metadata types, per 100 sample searches

How users searched

The first information we extracted from the sample sets was the relative frequency of search term types from each of the 100 searches of each sample set (Table 1).

These results did not betray an immediate and certain difference in behaviour, and the application of log-linear and chi-squared tests did not suggest any statistical significance. An inspection of the title searches in more detail reveals a different picture, however:

Title	Mar. 12	May 12	Mar. 13	May 13
Exact	18	37	27	18
Approx.	25	9	19	16

Table 2. Title searches: approximate vs exact (sample sets)

The immediate impression is that between March and May 2012 there was a rise in exact title searches, while approximate title searches fell; whereas in 2013, the picture is of a fall for both types of title search.

Using a log-linear analysis on the three factors of title type (exact vs. approximate), month and year, gave a global significance at p=0.0023 (G^2=16.66, df=4). This tells us that there are marked differences in the data.

The next stage of the analysis is to consider pairs of factors (title vs month, title vs year, months vs year. Each can be compared without (more cautious) or with (more precise) counterbalancing for the third factor. An effect between month and title type, without counterbalancing, was confirmed with p=0.0158 (G^2=5.82,df=1), and counterbalanced tests consistently provided significance: month-year p=0.0058 (G^2=10.3,df=2); year-title p=0.009 (G^2=9.42, df=2); and month-title p=0.0006 (G^2=14.7, df=2). We can thus be confident that each factor (month, year and title) plays a role, and also that they interact, as each pair of factors does. The 2013 web scale interface corresponds to a relative drop in all title searching, and exact title searches see a particular decline.

We also analysed and categorised known-item searches (of which title searches are one form), to see if there were adjustments in behaviour across the different periods. A comparison with the combined (web-scale) search tool available in 2012 illuminates some of the differences.

	2012 Combo.		2012 Catalog.		2013	
Month	Mar.	May	Mar.	May	Mar	May
Book	13	12	39	30	21	20
Article	23	15	1	7	17	9
Journal	2	0	0	2	1	0
Other	0	0	3	5	7	5
Total	38	27	44	44	46	34

Table 3. Known item target types (sample sets)

As noted in the previous section, in 2012 there were two search tools: a catalogue-specific search and a combined web-scale search; while in 2013 the interface had moved to a web-scale system only. We see that the 2013 behaviour (right) lies between the two different types of

search from 2012. As we have in fact four factors (interface, year, month and type), and one variable is dependent, a careful two-step analysis using the log-linear test is required.

Between the two interfaces of 2012, testing for differences by month and target document type, p<0.0001 (G^2=63.94,*df*=8), gets a clearly significant result. Taking all data, the strongest pair-wise effect is between search tool and target type (G^2=50.58, p<0.001), even without balancing for the third factor (i.e., month). With counter-balancing, results for tool-vs-type (p<0.0001;G^2=61,72) and tool-vs-month (p=0.0154, G^2=12.36) both prove significant. Month-vs-type is not significant (p=0.0589).

Retesting for a weighted norm of 2012 vs. 2013—the most cautious year-to-year comparison—detects an inter-action between year (i.e. search tool) and document target (p=0.025; G^2=9.36), but no significance for other factors.

From these tests we can draw three conclusions: first, there was a significant difference in the target document types between the two search tools in 2012; second, there was a change in the types of information sought between March and May 2012; third, that there was also a change in target documents between 2012 and 2013.

The use of web-scale search,in either year is associated with a greater focus on journal articles, but also correlates with a fall in known-item searches across time, particularly on article items.

We also investigated the number of key terms used in each sample set. Known-item searches were consistently longer: the median number of terms for keyword searches was 2; for known items it was 6. To test for conformance with previous research, we compared the 2012 catalogue with the 2012 combined search. In the 2012 catalogue set, the mean number of terms was 4.347, and for the 2012 combined search 6.133. Due to both the non-normality and the discrete values of query length, we used a single-tailed Mann-Whitney, resulting in *U*=5886.5, *z*=-2.16, p=0.0154. The 2012 and 2013 combined search data were also compared, resulting in non-significance (p=0.485).

Thus far we can draw no clear conclusions about the differences in our data, however it is obvious that the May 2013 set shows different search behavior than all other groups. This sample set had markedly fewer title searches than the three other samples, and it also contained fewer instances of infrequent search metadata (e.g. issue and edition). However, these patterns were not proven to be statistically significant. Analysing the whole data of each log, rather than the 100 samples, demonstrates a clearer pattern:

Searches	Mar. 12	May 12	Mar. 13	May 13
Total	1975	2271	1681	1397
Default	1367	1144	1594	1368
Combined	264	448	Not Applicable	
Specific	344	679		

Table 4. Search interfaces used (whole logs)

In 2012 there was a higher rate of overall use, with regular use of alternative information sources such as EBSCOhost or Google Scholar, etc.. Examining query metadata is also instructive, as seen in Table 5, and Table 6 shows how these and other metadata were combined.

Metadata	Mar. 12	May 12	Mar. 13	May 13
Title	626 (31.7%)	534 (23.5%)	550 (32.7%)	265 (18.9%)
Author	259 (13.1%)	195 (8.6%)	174 (10.4%)	91 (6.5%)
Date	137 (6.9%)	102 (4.5%)	78 (4.6%)	42 (3.0%)

Table 5. Commonly used metadata types, whole log (%age of searches in brackets)

A log-linear analysis was again performed, with secondary chi-squared tests that normalised for variances in sample sizes. The global result produced p<0.0001 (G^2=92.8; *df*=13), indicating that there is significant interaction between factors. Both year-month and month-type interactions proved strongly significant, with p<0.0001 (G^2=44.62,*df*=5; G^2=39.36,*df*=8, respectively). The interaction between year and month was also significant, but only with p=0.033 (G^2=16.74,*df*=8).

These results indicate there are differences in the use of title, author and data metadata, both between March and May of each year, and between 2012 and 2013.

We now examine what those differences are, reporting specific results from testing actual and nominal values that are properly weighted by the relative frequency of metadata combinations, total searches and time period.

Query	Mar. 12	May 12	Mar. 13	May 13
Citation (> 3)	96 (4.9%)	44 (1.9%)	53 (3.2%)	19 (1.4%)
Title and Author	27 (1.4%)	22 (1.0%)	14 (0.8%)	7 (0.5%)
Title, Date & Author	122 (6.2%)	93 (4.1%)	96 (5.7%)	46 (3.3%)
Author and Date	14 (0.7%)	36 (1.6%)	10 (0.6%)	15 (1.1%)
Title only	381 (19.3%)	375 (16.5%)	386 (23.0%)	189 (13.5%)

Table 6. Metadata combinations, whole logs (%age of searches in brackets)

Testing each query type separately using chi-squared, the distribution of citation searches between month and year is significant, (χ^2=45.32, p<0.0001, *df*=1), as are title, author and date (χ^2=6.79, p=0.0091, *df*=1), and indeed all metadata combinations. However, the direction varies: in May of each year, author and date searches are higher than in March, but title-and-author searches are lower.

A summary of the changes in metadata combinations, is that the web-scale search tool corresponds with an initially lower use of citation data, and a greater use of

titles. However, across time, the total volume of searching falls and. with the exception of author-and-date searches, the use of bibliographic data falls faster than with the catalogue search tool.

Search results users get

Search results are the feedback a search tool provides to the user on the efficacy of their queries. Given that the user interface to the search tools studied here varied only a small amount between 2012 and 2013, the vast majority of any change in search behaviour, particularly when looking at the May data sets, can reasonably be ascribed to the changing search results users saw. There are two obvious and expected differences between 2012 and 2013: we would expect web-scale search to produce a much larger number of search results (due to the greater number of items indexed), and we would expect web-scale search to produce search results that contained a wider range of resource types. Other impacts on search results also occur, as we will see below.

We reviewed the search results from the sample queries. First, we compared the total size of result sets:

	2012		2013	
Results	March	May	March	May
Median Hits	62	53	15167	19997
Mean Hits	586	793	1004788	695568
Null hits	20	16	7	2

Table 7. Number of hits per search (sample sets)

No significant differences are revealed within years, but between years results are consistently significant, using Mann-Whitney on matched months, with $p<0.001$; ($U>8000$, $z<-8.5$). Given the stark differences between groups, this result is unsurprising: web-scale search is producing markedly larger result sets.

A Chi-squared test, assuming the base null hypothesis of an even likelihood of null hits, produces $p<0.001$; $\chi^2 = 18.02$; $df=1$. Thus, web-scale search appears to be producing a lower rate of null result sets. Given these two marked differences, there are clear advantages to the web-scale search in terms of the number of matches found and the likelihood of a match.

Looking at the known-item matches within the sample sets teases out more detail.

	2012		2013	
Counts	March	May	March	May
#of KIs	44	44	46	34
Median	12	6	5404	4658
Mean	239	77	634452	584199
No hits	13	11	6	1
Hits but no KI	12	12	14	12

Table 8. Known-item search outcomes (sample sets)

This promisingly follows the overall data, and demonstrates that many of the null results follow known-item queries. Note that not all non-null results included the sought-for item, as noted in the final row. It is vital to note that in 2013, search results returned included a range of resource types, and that even searches for things that would not have been found in 2012 (such as electronic articles) returned results in 2013. This, and changes to the search terms used, led to the increase in result list sizes.

Though the mean count seems unlikely, weak searches across multiple catalogues or collections can easily result over a million matches. One query for a book review, comprised of many common words and complete bibliographic information, received over 33 million hits.

	2012		2013	
Position in results	March	May	March	May
1	16	19	14	13
2	7	2	7	7
3	1	1	0	0
4	1	1	1	6
5	0	0	0	0
Mean (if <5)	1.48	1.3	1.45	1.96
Median (if <5)	1	1	1	1.5

Table 9: Known-item search result ranking

To ensure that users who sought known items were still able to find them in the abundance of results, we examined where found known items (i.e. those that appeared within the first five results) fell in the results list. While the median rank was 1 for three of the four sample sets, in May 2013 it dropped back to 1.5. To ensure this was not a product of either sparse data or changing search technique, we also, finally, extracted all searches from the 2013 daily logs with two metadata items or more (in order to control the volume of searches sampled), and ran each query as per the 2012 and 2013 systems. These queries were generally well targeted, and matches occurred near the top of the list (if at all). While medians were close, at 1 and 2 for all matched queries, a paired Wilcoxon test provided $p=0.0001$; $z=-3.85$.

Furthermore, though many results were at the top of the list or second, the difference quickly diverged, for example, an item coming at the 6th position in 2012 was the 16th hit in 2013. This suggests that recall has been boosted at the cost of precision: while results are found more frequently, they move further down the result list. Thus, the precision at a given number of results worsens markedly with web-scale search.

We also tested the relative rates of success and failure for different forms of metadata supplied for known-item searches, using the whole logs for March and May 2013.

	March		May	
Query	Success	Failure	Success	Failure
Citation	11	42	4	15
Title, Date & Author	5	9	4	4
Title & Author	49	47	27	19

Table 10. Search success rates for known items (whole logs)

This is reported in Table 10. A log-linear analysis resulted in p=0.001 (G^2=24.36;*df*=7). The pair-wise comparisons produced a correlation between success rates and metadata used, with p=0.0002 (G^2=21.86;*df*=4).

However, there was no change in effectiveness of specific metadata combinations (p>0.92). This clearly shows that while the relative rate of these searches did change, the relative success of each type of search did not alter over the period (indeed appears very likely to be identical). In other words, changing rates of success are due to changes in the metadata users supply in their search, not a consequence of improved effectiveness in search tactics.

Global data

One natural concern may be that our daily logs or sample sets are not representative of the time periods from which they are sampled. Monthly logs demonstrate that the difference in demand between the individual days we sampled (Table 4) is in fact consistent with differences between the months as a whole. In 2012, the total monthly searches were 68,742 in March and 60,527 in May; in 2013, these figures were 52,415 and 47,843.

This difference in usage cannot be explained by the same search in 2012 being repeated on different sources: an inspection of our day logs from that year indicates that this is less than 1% of searches, and that in no way can account for a 25-30% drop in search activity.

We also examined the query length for each of the four sets of data, plus the combined searches from 2012. Unsurprisingly, the combined interface attracted longer searches than the other sets (median = 4 vs. 3). When testing the four monthly sets no significant difference in query length was detected.

DISCUSSION

In HCI we know that users alter their behaviour in response to changes in system interface (Dix et al. 2004). This is certainly known to be true of information search interfaces (see for example (Ballard et al. 2011)).

In this paper we have presented user behaviours in three subtly different versions of both system and interface: the standard library catalogue as a default search on the library homepage; web-scale search as a non-default option on the library homepage; and web-scale search as default on the library homepage.

What users type into the search boxes for each of these systems was analysed at the beginning of term, and later in the term to determine whether users gained an understanding of the systems over time.

The overall picture is that metadata searches dropped markedly between March and May of both years, but also that there is a similar, if smaller, effect between years, more notably in the May data. The change between March and May 2013 is more pronounced than that of the year before. The main change is a drop in the complexity of known-item searches, using fewer metadata types. Some of this change is almost certainly due to changing student needs within a term, or between cohort years. However, the marked change seen in 2013 is most plausibly linked to the change in underlying system.

Further discussion of these changes will be divided into four topics: known item searching; keyword searching, user behaviour; and implications for system design.

Known-item searches

In both years there were more known-item searches, and more specific known-item searches, in March than in May. The March queries had a greater number of words and included more metadata types than May in both years. Moreover the greater the number of metadata types, the more likely the search was to fail—this replicates a result seen in a log analysis of a traditional catalogue, where longer searches were more likely to fail (Lau et al. 2006). Examining the March searches closely reveals that they sometimes comprise an entire citation; given users' known difficulties in creating citations (Aronsky et al. 2005; Yu et al. 2006), this is likely caused by copy-and-paste being used for searching.

Known-item searches failed in one of two ways: returning no results at all, or returning a multiplicity of results in which the item sought was not obviously available. The former was the only way in which catalogue searches failed (again, echoing (Lau et al. 2006)). With the introduction of web scale discovery, however, we started to see the latter problem, confirming the user concerns noted in earlier studies of these systems (Gross et al. 2011). By May of both years, users were more parsimonious in their approach to known-item searching, performing fewer searches and providing less metadata.

From our data, in 2012 there was a more modest reduction in known-item searches than in 2013. Furthermore, the metadata used in May 2013 was much more limited in scope, as users increasingly searched only for title metadata. Only author-and-date searches demonstrated a rise between March and May either year.

Keyword searches

The proportion of searches in the default search box that were topic or keyword searches grew notably, particularly between May 2012 and May 2013. This at first seems incomprehensible, as the interface is broadly the same. This could be construed as a failing on the part of combined search, especially in light of earlier findings that users had difficulty finding known items in web scale systems (Gross et al. 2011). However in the context of our earlier work on searching different systems (McKay et al. 2011), it can be seen as positive evidence of users' understanding of the change to web-scale search. Subject

searching in traditional library catalogues is notoriously error prone, in part due to the controlled vocabularies employed in catalogues and in part simply due to limited inventory (Borgman 1996). As such, searchers typically avoid the library catalogue until they are seeking a known item (Ooi 2008; McKay 2011). In contrast, keyword searching is common in other types of interface (Nicholas et al. 2005; Jansen et al. 2006).

Insights into searcher behavior

In isolation, users' search strategies appear naïve, as they always have: in our data, it is notable that the users initially attempt sophisticated strategies for finding known items, but, especially in 2013, they revert to short, simple, keyword queries. These short queries have been seen in nearly every kind of online search (Mahoui et al. 2001; Nicholas et al. 2005; Huntington et al. 2006; Jansen et al. 2006; Lau et al. 2006; Lown et al. 2012), and has been interpreted near universally as poor quality searching (Meadow et al. 2012) that could be remedied by more training, better interface support (Gerwe et al. 2000) or assistance from a librarian (Mann 2008). Similarly the increased use of keyword searching could be seen as evidence of poor quality searching and lack of understanding of the system (Meadow et al. 2012).

We would argue that our data shows the opposite of what has been assumed in the past: users respond, and respond effectively, to the system's behaviour. The first key to this is the increase in keyword searching in 2013: far from showing users are adopting simple search strategies, we believe that it implies recognition that the system has changed, and utilization of the new functionality available. Our own earlier work on the interface shown in Figure 1 (McKay et al. 2011) demonstrated that users were more likely to keyword search in systems where keyword searches returned some results (in that study Google Scholar and EBSCO); our study demonstrates that web-scale search also returns results in response to keyword searching. Essentially, users have learned that the system is more 'Google-like' than the catalogue and are using it as such. While some would argue that this is encouraging lazy searching (Brabazon 2006), to present a stricter alternative would merely drive users to more convenient resources, even if they were of lower quality (Connaway et al. 2011).

The shortening of queries that appears so regressive in fact also shows users responding to the system. Longer queries with more metadata types (or, worse, citations) are less likely to return the known items users seek, while shorter queries more often work. Users are at least reluctant, if even able, to use advanced search techniques (Hock 2008) or Boolean searching (Bandos et al. 2002) (the next logical step after entering more information). As "higher quality" searching produces lower quality results, users revert to tried-but-true shorter queries. This is not simply a naïve response, but the product of experience of searching by users who have a better (albeit implicit) understanding of the system they are using than they are typically given credit for.

While those with search formulation expertise can identify the weakness in this strategy, it is not the product of a refusal or reluctance to learn that is characterised in much of the literature (c.f. (Brabazon 2006; Mann 2008; Meadow et al. 2012)). Instead, user responses are tailored to immediate experience. It is quite possible that a citation search is, for example, initially attempted as a more advanced technique, fails, and is then discontinued. No assistance is given by the system to identify possible improvements.

Implications for system design

This study shows that users are considerably more savvy in their approach to search than they have been considered to be. In both 2012 and 2013 our data demonstrates changes in search behaviour in response to learning about the default search system. The differences in behaviour reflect the alteration in system operation between the years, and in each year adjusments in user searches reduce the chances of search failure.

The question, then, is how best to further support users in their quest for information? The key to answering this question perhaps lies in the concept of "interface permissiveness" as described by Thimbleby in 2001 (Thimbleby 2001): systems should, insofar as is possible, allow users to approach a task in the way that suits them best. The author uses the example of a vending machine: users should be able to either put in money or choose their item first, not only in a needlessly set order.

Our data shows that web-scale search is more permissive than the traditional catalogue for keyword searches. Due to the vast number of items and the considerable vocabulary these search tools cover, users will almost invariably find something with a keyword search. This stands in stark contrast to library catalogue systems, which regularly fail users who do not know exactly what they are looking for (Borgman 1996; Lau et al. 2006). This permissiveness may go too far, however: it has been anecdotally reported in previous usability studies of web-scale search that searchers have been discouraged by the high number of results (Gibson et al. 2009; Gross et al. 2011). Part of the issue is that some users clearly had a specific type of resource in mind when they searched: not being able to identify, for example, scholarly articles easily has been flagged as a problem in previous studies (Gross et al. 2011; Slaven et al. 2011). Excellent relevance ranking can ameliorate some of the problem of "too many" results: Google is not perceived as having this problem (Khoo et al. 2012).

Essentially, having too many results turns a searching problem into a browsing problem: this is to be expected given that keyword searching typically occurs early in the information seeking process, when searching and browsing are heavily interleaved (Kuhlthau 1999). There are a number of ways to help users at this stage in the process, such as good support for query refinement (Khoo et al. 2012), support for interleaving results browsing with search (Bates 1993), and support for results browsing—perhaps in the form of facets (facet use is known to rise when web-scale search is introduced (Ballard et al. 2011)). How best to support users browsing web-scale search results, and whether the current facet

options in the search studied here are effective, remain questions for further research.

In contrast to the improved keyword searching permissiveness offered by web-scale search, known-item searching is still very limited. Common user behaviours, such as pasting in citations and providing more than one type of metadata (also seen in our earlier work (McKay et al. 2011)) are not well supported by web-scale search, leaving users to type in fewer words, and consequently get a larger number of results. This problem was identified in an early usability study of web-scale search: users struggled to find known items (Slaven et al. 2011). This finding is reflected in another study, which showed users going out of their way to use an old catalogue three months after the implementation of a web-scale search as default (Way 2010). All library systems—including web-scale search—need to better support users' attempts to create more precise queries, whether by entering greater number of metadata types or by copying and pasting citations.

CONCLUSIONS

In this paper we have presented an analysis of queries in a library search box before and after the implementation of a web scale discovery platform, at the beginning of an academic term and near the end. Our study demonstrates that users' behaviour changed both between the beginning of term and the end of term each year, and also in response to the implementation of web-scale search.

In both years users created shorter queries and used fewer metadata types in the later part of the term. We posit that while their response to experience of the system was conservative (using more cautious forms of familiar methods, rather than developing new skills), it nonetheless shows adaptation and learning.

Users' behaviour in 2013 similarly demonstrated a burgeoning understanding of the change in underlying search system: the proportion of keyword searches increased dramatically. This is almost certainly due to the higher likelihood of actually obtaining results in web-scale search.

This study, then, suggests that users are considerably more savvy and aware of effective search techniques than has been believed in the past. In light of this, it behoves us to better support users not just in supporting their native query strategies (e.g. pasting citations into the search box, and searching over multiple metadata types without search operators), but also in evaluating the large number of search results returned by web-scale search.

Faceted search is one option for supporting users in narrowing search results; interactive query reformulation is another. Which strategies, if any, are most effective remains a question for further research. Similarly, how best to redesign underlying search systems to accommodate more complex searches without search operators remains an open question.

REFERENCES

Agosto, D. E. Bounded Rationality and Satisficing in Young People's Web Based Decision Making. JASIST 53, 1, (2002) 16-27.

Aronsky, D., Ransom, J. and Robinson, K. Accuracy of References in Five Biomedical Informatics Journals. J Am Med Inform Assn 12, 2, (2005) 225-228.

Ballard, T. and Blaine, A. User search-limiting behavior in online catalogs: Comparing classic catalog use to search behavior in next-generation catalogs. New Library World 112, 5, (2011) 261-273.

Bandos, J. A. and Resnick, M. L. Understanding Query Formation in the Use of Internet Search Engines. Proc. HFES 46th Annual Meeting, Human Factors and Ergonomics Society. (2002).

Bates, M. J. The design of browsing and berrypicking techniques for the online search interface. Online Inform Rev 13, 5, (1993) 407-424.

Bell, S. J. The Infodiet: How Libraries Can Offer an Appetizing Alternative to Google. Chronicle of Higher Education 50, 24, (2004) 15.

Borgman, C. L. Why are online catalogs *still* hard to use? JASIST 47, 7, (1996) 493-503.

Brabazon, T. The Google Effect: Googling, Blogging, Wikis and the Flattening of Expertise. Libri 56, (2006) 157-167.

Buchanan, G., Cunningham, S. J., Blandford, A., Rimmer, J. and Warwick, C. Information Seeking by Humanities Scholars. Proc. ECDL 2005, Springer. (2005), 218-229.

Buchanan, G. and McKay, D. In the Bookshop: Examining Popular Search Strategies. Proc. JCDL 11, ACM. (2011).

Cockrell, B. J. and Jayne, E. A. How do I find an article? Insights from a web usability study. J Acad Libr 28, 3, (2002) 122-132.

Connaway, L. S., Dickey, T. J. and Radford, M. L. "If it is too inconvenient I'm not going after it:" Convenience as a critical factor in information-seeking behaviors. Lib Inform Sci Res 33, 3, (2011) 179-190.

Cooper, M. D. Usage patterns of a web-based library catalog. JASIST 52, 2, (2001) 137-148.

Dix, A., Finlay, J. and Abowd, G. Human-computer interaction, Pearson/Prentice Hall (2004).

Fast, K. V. and Campbell, D. G. "I still like Google" University student perceptions of searching OPACs and the web. ASIST Proceedings 41, 1, (2004) 138-146.

Foster, A. K. and MacDonald, J. B. A Tale of Two Discoveries: Comparing the Usability of Summon and EBSCO Discovery Service. Journal of Web Librarianship 7, 1, (2013) 1-19.

Fried Foster, N. and Gibbons, S. Studying Students: The Undergraduate Research Project at the University of Rochester. Rochester, NY, Association of College and Research Libraries (2007).

George, C., Bright, A., Hurlbert, T., Linke, E. C., St. Clair, G. and Stein, J. Scholarly use of information: Graduate students' information seeking behavior. Inform Res 11, 4, (2006).

Gerwe, P. and Viles, C. L. User effort in query construction and interface selection. Proc. DL'00, ACM. (2000).

Gibson, I., Goddard, L. and Gordon, S. One box to search them all: Implementing federated search at an academic library. Libr Hi Tech 27, 1, (2009) 118-133.

Gore, G. Undergraduates prefer federated searching to searching databases individually. EBLIP 3, 3, (2008) 61-63.

Griffiths, J. R. and Brophy, P. Student Searching Behavior and the Web: Use of Academic Resources and Google. Libr Trends 53, 4, (2005) 539-554.

Gross, J. and Sheridan, L. Web scale discovery: the user experience. New Library World 112, 5, (2011) 236-247.

Hane, P. J. The Truth About Federated Searching. Information Today 20, 10, (2003) 24.

Hock, R. The Elements of Advanced Search. (cover story). Online 32, 4, (2008) 14-19.

Huntington, P., Nicholas, D. and Jamali, H. R. Employing Log Metrics to Evaluate Search Behaviour and Success: Case Study BBC Search Engine. J Inform Sci 33, 5, (2006) 584-597.

Jansen, B. J. and Spink, A. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. Inform Process Manag 42, 1, (2006) 248-263.

Jansen, B. J., Spink, A. and Saracevic, T. Real life, real users, and real needs: a study and analysis of user queries on the web. Inform Process Manag 36, 2, (2000) 207-227.

Jones, S., Cunningham, S. J. and McNab, R. An Analysis of Usage of a Digital Library. Proc. ECDL 98, Springer. (1998), 261-277.

Khoo, M. and Hall, C. What would 'Google' Do? Users' Mental Models of a Digital Library Search Engine. Proc. TPDL 12, Springer Verlag. (2012).

Kuhlthau, C. C. Inside the Search Process: Information Seeking from the User's Perspective. JASIST 42, 5, (1999) 361-371.

Lau, E. P. and Goh, D. H.-L. In search of query patterns: A case study of a university OPAC. Inform Process Manag 42, 5, (2006) 1316-1329.

Lown, C., Sierra, T. and Boyer, J. How Users Search the Library from a Single Search Box. Coll Res Libr, (2012).

Mahoui, M. and Cunningham, S. J. Search Behavior in a Research Oriented Digital Library. Proc. ECDL 2001, Springer. (2001), 13-24.

Mann, T. The Peloponnesian War and the Future of Reference, Cataloging, and Scholarship in Research Libraries. J Libr Metadata 8, 1, (2008) 53-100.

Marchionini, G., Dwiggins, S., Katz, A. and Lin, X. Information Seeking in Full-Text End-User-Oriented Search Systems: The Roles of Domain and Search Expertise. Libr Inform Sci Res 15, 1, (1993) 35-69.

McKay, D. Gotta keep 'em separated: Why the single search box may not be right for libraries. Proc. CHINZ '11, ACM. (2011), 109-112.

McKay, D. and Buchanan, G. One of these things is not like the others: how users search different information resources. (2011), to appear.

Meadow, K. and Meadow, J. Search Query Quality and Web-Scale Discovery: A Qualitative and Quantitative Analysis. College & Undergraduate Libraries 19, 2-4, (2012) 163-175.

Nicholas, D., Huntington, P. and Watkinson, D. Scholarly Journal Usage: The Results of Deep Log Analysis. J Doc 61, 2, (2005) 248-280.

Nordlie, R. "User Revealment"— A Comparison of Initial Queries and Ensuing Question Development in Online Searching and in Human Reference Interactions. Proc. SIGIR '99, ACM Press. (1999), 11-18.

Ooi, K. How Adult Fiction Readers Select Fiction Books in Public Libraries: A Study of Information Seeking in Context. School of Information Management. Victoria University of Wellington (2008).

Slaven, C., Ewers, B. and Vollmerhause, K. From "I hate it" to "it's my new best friend!" : Making heads or tails of client feedback to improve our new Quick Find discovery service. Proc. Info Online '11, ALIA. (2011).

Swanson, T. A. and Green, J. Why We Are Not Google: Lessons from a Library Web site Usability Study. J Acad Libr 37, 3, (2011) 222-229.

Talja, S. and Maula, H. Reasons for the use and non-use of electronic journals and databases: A domain analytic study in four scholarly disciplines. J Doc 59, 6, (2003) 673-691.

Thimbleby, H. Permissive user interfaces. IJCHI 54, 3, (2001) 333-350.

Wallace, P. M. How do patrons search the online catalog when no one's looking? Transaction log analysis and implications for bibliographic instruction and system design. RQ 33, 2, (1993) 239-253.

Way, D. The Impact of Web-scale Discovery on the Use of a Library Collection. Serials Rev 36, 4, (2010) 214-220.

Yu, F., Sullivan, J. and Woodall, L. What Can Students' Bibliographies Tell Us?- Evidence Based Information Skills Teaching for Engineering Students. EBLIP 1, 2, (2006).