

Buying bestsellers online: A case study in Search & Searchability

Trystan Upstill

Department of Computer Science
Australian National University
Canberra ACT 0200

Trystan.Upstill@cs.anu.edu.au

Nick Craswell

Enterprise Search Group
CSIRO Mathematical and Information Sciences
GPO Box 664, Canberra, ACT 2601

Nick.Craswell@csiro.au

David Hawking

Enterprise Search Group
CSIRO Mathematical and Information Sciences
GPO Box 664, Canberra, ACT 2601

David.Hawking@csiro.au

Abstract

A website's design directly affects how well search engines can crawl, match and rank its pages. For this reason, searchability is an important concern in site design. We study the interaction between search engines and Web sites by means of a case study of online bookstores and general-purpose search engines. The task modelled is that of finding web pages from which a book, described by its title, may be purchased.

We first compared the relative effectiveness of search engines in finding pages matching the criterion, regardless of bookstore. Then we compared the relative searchability of the bookstore websites by observing how many times each bookstore contributed useful answers to the search results.

Large differences in the performance of both search engines and bookstores were observed. Two of the search engines performed better than their peers, and one bookstore was far more searchable than all others. To further explore these differences we tabulate the total number of pages from each bookshop which are included in the search engine indexes.

We conclude with recommendations both to bookstores on how they may improve their Web presence, and to search engines on how they may improve their performance for product searches.

Keywords Information Retrieval

Additional Keywords Web search, evaluation, transactional search

**Proceedings of the 7th Australasian Document Computing Symposium,
Sydney, Australia, December 16, 2002.**

1 Introduction

The World-Wide Web is commonly used by consumers to research and purchase goods [7]. With this demand for goods and services online, product finding is becoming an increasingly important consideration for Web search engines. However, despite the prevalence of such tasks, the general problem of product, or *transactional* [11] search has not been thoroughly examined.

To begin to understand the difficulties involved in processing transactional searches we have conducted a study of the most popular Web bookstores and search engines. Our study examines product finding both from a search engine and from a bookstore point of view. This is necessary because any difficulties involved in Web product search may well be due to a mismatch between search engines and product providers.

Previous work has evaluated the service finding ability of TREC search systems [5] and Web search engines [6] on a set of apparently transactional queries extracted from natural language Web logs. In these studies the sole aim was to compare search engines on early precision; no information was available (or needed) about what resources were available to find and there was no opportunity to compare the *searchability* of online vendor sites. Searchability is concerned with *crawlability*, that is how easy it is to retrieve pages, and how well a site is matched and ranked by search engines.

Here, we study both search effectiveness and searchability with respect to a particular type of commodity (books) which is frequently sold over the Web. We measure the relative effectiveness of a selection of search engines in finding pages from which a book, specified only by its title, may be purchased. We also compare the relative searchability of a selection of online bookstores. This

is performed through an examination of the proportion of best-seller books for which a “buy-here” page from a bookseller appears in search engine listings.

2 Method

The process we followed to gather and evaluate our data is outlined in this section. We began by identifying our book query set. Following this we identified a set of search engines for querying. We sent our queries to the search engines and retrieved and recorded the first 1000 results. We then used popular directory listings to determine which bookstores would be included in the study. Finally we determined the correct books for our queries (by ISBN) and created a listing of correct bookstore URLs at which they could be found.

These steps are detailed below:

2.1 Query set

The query set was derived from the titles of the New York Times best-sellers for September 2002 [14]. We found 206 unique book titles in 9 categories. The book/category breakdown is outlined in the Appendix. We sent the titles to search engines as phrases (i.e. inside ‘ ‘ or marked as phrases in advanced searches). Since the book titles on the best-seller lists were fully capitalized, we changed their case to combinations of lower and upper-case letters. All words, apart from join words (such as “the”, “and” and “or”), begin with a capital.

2.2 Search engine set

Table 1: **Search Engines Properties.** The column labelled “Abr.” contains abbreviations used through the paper. “Used by” reports search services that use the search system. “Rank” reports the position in the Nielsen/NetRatings Search Engine Ratings for September 2002 [9].

S.Engine	Abr.	Used by [10]	Rank
AltaVista [1]	AV	AltaVista	8
Fast [2]	FA	AllTheWeb	-
Google [4]	GO	Google	3
		AOL	4
		Netscape	9
		Yahoo	1
MSN Search [8] (based on Inktomi)	MS	MSN Search	2
		Looksmart	10
		HotBot	-
		Overture	6

Table 1 outlines the characteristics of the search engines examined in our study. Our studies span four popular Web search systems. These systems provide core search services for the 4 most popular search engines, and for 8 of the top 10 search services [9].

The query syntax submitted to each search engine is reported in the Appendix.

2.3 Bookstore set

The bookstore set was derived from the Google “*Shopping > Publications > Books > General*” [3] and Yahoo! “*Business and Economy > Shopping and Services > Books > Booksellers*” [13] directories. Bookstores were included if they sold the top bestseller in at least three of the nine categories. The books were found using internal search engines to search for both the title and the author of each book. The bookstores were only judged on the categories for which they stocked (or listed) the bestseller. The intuitive justification for this was that there may have been some specialised (e.g. fiction only) bookstores and we did not want to exclude them from our study. The full list of the 38 eligible bookstores and their salient properties is provided in Table 7 in the Appendix.

2.4 Correct answers

In our study correct answers must fulfill two criteria: the page must relate to the book whose title is given as the query, and the retrieved page must be transactional in nature.

When identifying book identities we deemed correct answers to be either hardcover or paperback editions of the book¹. A list of the queries and the ISBNs of the books judged as correct answers is available [12].

We define a *transactional page* to be a bookstore page from where you can buy a book. Browse pages, search results or genre listings are not judged as correct results. For many bookstores the correct answers were observed to have the hardcover or paperback ISBN in the URL (in many cases there were many correct URLs which were all observed to contain the ISBN). To cut down on manual judging for these bookstores, we performed automatic judging based on the presence or absence of the ISBN in the URL. For other bookstores we located unique product identifiers for each book and checked URLs for their presence.

3 Results

We begin by outlining our findings from a book retailers’ perspective - detailing which bookstore was the most searchable. We then report which search engine was best at finding books. Finally we examine the size of the index and link graph for each bookstore reported by each search engine.

¹Large print books and audio books were deemed to be incorrect answers.

Table 2: **Bookstore comparison.** This table includes all bookstores which had at least one success at 1000 for any search engine. *Powells* is included in the table due to the high number of results matched in search listings. The S@1000 breakdown column gives an indication of which search engine returned what correct books from which bookstores in its top 1000. The “Hostname Results” column reports the number of pages found for each bookstore by all search engines.

Bookstore	S@1	S@5	S@10	S@100	S@1000	S@1000 breakdown (AV:FA:GO:MS)	Hostname Results
Amazon	0.124	0.325	0.402	0.492	0.584	104:83:162:132	3903
Barnes and Noble	0.028	0.096	0.140	0.225	0.316	0:87:170:3	3603
Walmart	0.010	0.030	0.045	0.070	0.075	2:0:0:60	277
BookSite	0.000	0.004	0.005	0.013	0.013	0:0:0:11	52
ecampus	0.0	0.0	0.0	0.005	0.012	0:7:0:3	290
AllDirect	0.0	0.0	0.0	0.002	0.005	0:4:0:0	52
NetstoreUSA	0.0	0.0	0.0	0.001	0.010	0:8:0:0	261
Sam Weller’s Books	0.0	0.0	0.0	0.001	0.006	0:5:0:0	22
Books-A-Million	0.0	0.0	0.0	0.0	0.008	0:4:0:3	775
1BookStreet	0.0	0.0	0.0	0.0	0.006	0:5:0:0	17
Wordsworth.com	0.0	0.0	0.0	0.0	0.004	1:0:1:1	92
TextbookX.com	0.0	0.0	0.0	0.0	0.002	0:2:0:0	22
CodysBooks.com	0.0	0.0	0.0	0.0	0.002	0:2:0:0	78
Arthurs Books	0.0	0.0	0.0	0.0	0.003	0:1:0:0	3
Powells Bookstore	0.0	0.0	0.0	0.0	0.0	0:0:0:0	1031

3.1 Bookstore comparison

We first measured the book finding success rates at several cutoffs. The success rate measure is indicated by S@ n where n is the cutoff rank. S@10 measures how often at least one correct page is returned within the first 10 results, and S@1 corresponds to the probability that a right answer appears at rank 1 (cf. the “I’m feeling lucky” button on **Google**). We also evaluated S@5, S@100 and S@1000.

Table 2 contains the results for this experiment. Our observations are that:

- Of the 38 bookstores evaluated only 14 returned any correct answers within the top 1000 results by any of the search engines.
- Only 4 bookstores ever made it into the top 10 in any search engine: *Amazon*, *Barnes and Noble*, *Booksite* and *Walmart*
- *Amazon* was the most searchable bookstore in our evaluation, achieving the highest success rates.
- Only *Amazon* had correct results returned by every search engine.
- *Barnes and Noble* performed well on **Google** and **Fast**.
- *Walmart* performed well on the **MSN Search** search engine.

- The only search engine which returned results for many of the smaller bookstores was **Fast**.

3.2 Search engine comparison

Table 3: **Search Engine Success Rates.** The best result at each cutoff is highlighted.

search engine	Success Rates				
	@1	@5	@10	@100	@1000
AV	0.14	0.39	0.45	0.50	0.52
FA	0.00	0.02	0.05	0.18	0.52
GO	0.15	0.56	0.67	0.83	0.89
MS	0.36	0.57	0.65	0.72	0.73

Table 4: **Search Engine Precision.** Note that precision at 1 is equivalent to the success rate at 1. The precision at measures below 100 is less than 1/100 in all cases. The best result at each measure is highlighted.

search engine	Precision			
	@1	@5	@10	@ 100
AV	0.14	0.08	0.05	0.01
FA	0.00	0.00	0.01	0.00
GO	0.15	0.20	0.15	0.03
MS	0.36	0.13	0.08	0.01

Our observations on the success rates and precision reported in Table 3 and Table 4 were:

- **AltaVista's** (AV) performance was inferior to that of both **Google** and **MSN Search** at all cutoffs. **AltaVista** demonstrated around half the precision of **MSN Search**.
- **Fast** (FA) trailed well behind all other search engines, but provided a large number of correct answers between the 100th and 1000th position (success rate jumps from 0.18 to 0.52). The precision for **Fast** was very low.
- **Google** (GO) trailed **MSN Search** at S@1 but surpassed its performance from S@10 onwards. **Google** returned more correct answers in their top 5, 10 and 100 results than did **MSN Search**.
- **MSN Search** (MS) produced the strongest results at S@1 and S@5, but did not improve as much as other search engines when cutoffs were extended.

3.3 Search engine/bookstore URL coverage

The transactional pages for some bookstores may not be returned because they have never been crawled by a search engine. Table 5 lists the number of pages from each bookstore reported to be contained within each search engines' index.

From these results we observed that:

- *Amazon* had a consistently large search engine coverage (around three million on three out of four search engines). **Fast** covered an order of magnitude less documents from *Amazon* than did the other search engines, but still more than it did for any other bookstore.
- The coverage of *Barnes and Noble* varied widely. While the **MSN Search** coverage of *Barnes and Noble* was small it appeared to contain product pages, with three correct answers retrieved. Only 500 *Barnes and Noble* pages were covered by **AltaVista**. Over a million pages were covered by **Google**.
- A large number of *Walmart* pages were covered by **MSN Search**, whereas **Fast** and **Google** covered a relatively small number of pages.
- **Fast** did not have very large coverage of any one bookstore (maximum was 360,000). They tended to have a larger breadth of results, with larger crawls of lesser known bookstores.
- **AltaVista** had large coverage only of *Amazon*, *Walmart* and *Powells*. We could not expect to find book results in their small (sub 1000 page) crawls of other bookstores.

- *Powells* had large coverage (with three out of four search engines indexing over 40,000 pages) but never had any of their books returned in the top 1000 results for these search engines.

3.4 Search engine/bookstore link coverage

Links are used by search engines when discovering pages to crawl and when ranking pages. Only two of the evaluated search engines supported domain name link counts: **AltaVista** and **Fast**. Domain name link counts gave us the number of links to an entire domain name rather than just links to a single page. Using this information we could determine the link popularity of an entire bookstore.

Table 6 contains the results for this study. From these results we observed that:

- **Fast** observed a large number of links to *Amazon* but did not cover *Amazon* as comprehensively as other search engines.
- *Powells* bookstore had a large number of incoming links, but still performed poorly.
- **Fast** discovered more links to diverse hosts than did **AltaVista**. This could be attributed to the fact that **Fast** performed a deeper crawl and encountered a larger number of internal links.

4 Discussion

The best book finding search engines were **Google** and **MSN Search** and the most successful bookstore was *Amazon*. **MSN Search** provides the most correct answers at position 1. However, **Google** provides more correct answers in the top 5; potentially giving users more book buying options. The following sections discuss why we believe **Google** and **MSN Search** perform better than the other two search engines, and why *Amazon* performance is better than that of other bookstores.

While we believe the concepts from this experiment are relevant to general Web product searches, it is important to note that these experiments are not necessarily reproducible. The algorithms and indexes used in the evaluated search engines and bookstores are not known and may well change over time.

4.1 Search engines

In order to maximize the book finding ability of a Web search engine it seems that the crawler must perform deep crawls (i.e. crawl long URLs) of at least one bookstore. This is because all the examined bookstores bury product pages deep within their URL directory tree (generally as leaf nodes).

Table 5: **Search Engines URL Coverage.** Note that the totals in the right hand side column may contain duplicate URLs. (this occurs when the same URL is found by different search engines.) Also the totals include figures for unlisted bookstores.

Bookstore	AV	FA	GO	MS	TOTAL
amazon.com	3,675,723	358,376	3,620,000	2,838,819	10,492,918
barnesandnoble.com	521	192,792	1,240,000	2,822	1,436,135
walmart.com	89,243	1,076	10,500	916,162	1,016,981
netstoreusa.com	1,171	315,002	93,000	42,052	451,225
<i>powells.com</i>	39,397	111,977	65,900	6,204	223,478
textbookx.com	18	23,157	38,600	150	61,925
alldirect.com	24	26,278	7	27	26,336
ecampus.com	300	7,763	2,010	240	10,313
planetgold.com	18	8,361	774	18	9,171
booksamillion.com	22	5,860	54	865	6,801
cornwalldiscountbooks.com	1	5,423	2	1	5,427
wordsworth.com	735	228	2,290	1,271	4,524
booksite.com	93	169	1,190	290	1,742
codysbooks.com	74	1,308	238	57	1,677
arthursbooks.com	7	1,221	8	384	1,620
samwellers.com	7	278	5	8	298
total	3,809,257	1,087,000	5,081,161	3,810,094	13,787,512

Table 6: **Search Engines Link Coverage.** Note that the totals in the right hand side column may contain several links to the same URL.

Bookstore	AV	FA	TOTAL
amazon.com	12,408,441	25,955,858	38,364,299
<i>powells.com</i>	5,197,526	316,989	5,514,515
textbookx.com	3,456,068	28,453	3,484,521
barnesandnoble.com	234,137	784,088	1,018,225
walmart.com	14,783	267,008	281,791
booksite.com	4,927	113,729	118,656
booksamillion.com	34,137	79,351	113,488
ecampus.com	2,170	102,047	104,217
netstoreusa.com	10,548	91,867	102,415
1bookstreet.com	25,229	50,064	75,293
wordsworth.com	2,750	21,694	24,444
a1books.com	4,545	16,270	20,815
codysbooks.com	1,062	9,512	10,574
alldirect.com	614	6,508	7,122
arthursbooks.com	109	1,700	1,809
samwellers.com	106	208	314
total	21,463,332	28,160,545	49,623,877

While **Fast** appeared to index a much larger selection of bookstores, they appeared to not crawl as much of the *Amazon* bookstore as the other search engines. Considering the majority of correct hits for all search engines came from the *Amazon* bookstore this could have been one of the main reasons for the observed low effectiveness of **Fast** on this task.

To improve crawler coverage of popular bookstores it is also necessary to crawl dynamic URLs with `?` or `cgi-bin`, or even crawling many pages generated from a single script with different parameters. On the *Powells*, *Walmart* and *Barnes and Noble* bookstores, all product pages are created from a single script, with the book's ISBN as a parameter. As many slightly different URLs frequently contain exactly the same information it is also necessary to perform advanced equivalence (duplicate) detection. This is the case with product pages from the *Amazon* bookstore which includes referral identifiers in their URLs.

Finally, **MSN Search** and **AltaVista** use referral information when directing traffic to bookstores but **Google** and **Fast** do not. While not wanting to encourage preferential treatment of search engine partnered bookstores, using a referral id to gain revenue when people purchase books found through your search engine is a potential source of income for a search engine.

4.2 Bookstores

We split up our examination of bookstore performance into two parts: Coverage and Matching/Ranking. Coverage discusses how bookstores may maximise their crawlability. Matching/Ranking discusses reason why some bookstores' books may be returned in preference to others.

4.2.1 Coverage

When consulting Tables 5 and 6 we observed that the top three bookstores by URL coverage were also the top three bookstores by success rate. Having your bookstore featured extensively in search engine indexes appears to be essential for success. *Amazon* achieves high coverage in the indexes of all evaluated search engines.

It is important for a bookstore to have deep crawls indexed in as many search engines as possible. We found three potential reasons why bookstores were not crawled deeply.

1. The bookstores did not effectively convert incoming links into crawled pages.
2. The bookstores did not have sufficient external deep links directly to product pages.
3. The product pages were too deep in the bookstores site hierarchy.

Converting links into crawled product pages is an important consideration. Many bookstores that have a high link count are unable to achieve wide URL coverage. This may be due to bookstore sites appearing to search engines as being dynamically generated. When crawling a dynamic site a crawler may be confused by the site structure. This could be caused by dynamic information in the URL (such as question marks), or by the generation of a series of pages by a single script. In many of these cases a crawler will either simply ignore the link, or be unable to retrieve any meaningful information from it. This appears to occur on *Powells*, which has a large number of incoming links but less indexed pages than other well linked bookstores. The use of a dynamic-to-static web site convertor may go some way to addressing these coverage problems [15]. The site which managed to best convert incoming links to crawled pages was *NetstoreUSA*. In contrast to all other evaluated bookstores, *NetstoreUSA* had more pages indexed by the search engines than they had links. Upon further examination we found that *NetstoreUSA* have static-looking URLs, with a simple hierarchy of `shtml` pages.

To encourage a deep crawl that will cover all product pages it is necessary for bookstores to ensure they have both internal and external links directly to their product pages. To encourage user linking it is important to use meaningful, sensible and consistent URLs for products. While one can envisage a Web developer linking to a URL which has the form `foo.com/ISBN/` it may be less likely that they link directly to `foo.com/prod/prod.asp?prod=9283&source=09834`. Further deep linking may be encouraged through the use of incentive or partnership programs. If such a program is in place it is important to ensure partners are able to point directly to products and that all partners point to the same consistent URL for each product.

4.2.2 Matching/ranking performance

All of the evaluated search engines use some form of link and page content information when matching and ranking pages. It is important, from a bookstore perspective, that a search engine return relevant product pages in response to a query. However, many of the observed search results were browse and search pages. The *Powells* bookstore is a case in point. Despite having many links, reasonable coverage in search engine indexes and having results matched frequently, *Powells* transactional pages were never returned. This may indicate poor page content, site organisation and/or a lack of deep linking directly to products (as their referral program appears to be processed through their front page).

A potential method to alleviate these problems is to use a “robots.txt” to direct crawlers to ignore search and browse pages and only index product pages.

4.2.3 External factors and limitations

We encountered several interesting phenomena we could not explain with certainty.

Some search engines appear to favour certain bookstores over others. (e.g. **Google** and **Fast** have large indexes of *Barnes and Noble* while **MSN Search** and **AltaVista** have large indexes of *Walmart*.) The reasons for these preferences cannot be fully accounted for in this paper as we do not have access to the algorithms employed by the Web search engines. An example of this is the good performance of the *Walmart* bookstore in **MSN Search**. The results suggests that **MSN Search** has access to extra information for *Walmart* that is not available to the other search engines.

Further, we rely on each search engine returning fairly accurate size estimates. We assume that the figures reported are at the very least indicative of the relative coverage.

5 Conclusions

Throughout our experiments we have observed large differences in the performance of bookstores and search engines. The search engines’ precision at 10 varied from 0.00 to 0.20, while 24 of the evaluated bookstores did not appear in the top 1000 results for any of the evaluated search engines for any of the books.

Our results have illustrated the importance of a combined approach to improving product search. While search engines should endeavour to discover more product pages it is equally important for bookstores to build a suitable site structure that allows search engines to perform deep, thorough crawls. To improve transactional search effectiveness the search engines should perform deep crawls of provider sites and crawl their dynamic pages (especially those that are linked to by other sites). To improve searchability, bookstores should use short non-changing URLs and encourage deep linking directly to their product pages.

5.1 Further work

In our experiments we only included results that were bookstore book purchase pages. An extension to our study would be to include other types of pages in search effectiveness measures, such as price comparison pages.

Another extension would be to examine how searchability changes according to book properties. It would be interesting to correlate the age of a

book with how well it is retrieved by the search engines. One would expect newer books to be less effectively retrieved, as there would be little anchor-text. In this way the current study may set a difficult task since many bestsellers may have only recently been published. However, approximately half the books are paperback versions for which hardcover versions, which have generally been released some time ago, are also considered relevant. Differences in searchability across book genres may also be of interest, we may expect computing books to be more effectively retrieved due to a larger amount of anchor-text.

References

- [1] AltaVista. AltaVista search engine, 2002. www.av.com.
- [2] FAST Search and Transfer, ASA. Personal communication, 2002. www.alltheweb.com.
- [3] Google. Google directory > shopping publications > books > general, September 2002. directory.google.com/Top/Shopping/Publications/Books/General.
- [4] Google. Google search engine, 2002. www.google.com.
- [5] David Hawking. Overview of the TREC-9 Web Track. In *Proceedings of TREC-9*, 2000. trec.nist.gov/pubs/trec9/.
- [6] David Hawking, Nick Craswell and Kathleen Griffiths. Which search engine is best at finding online services? In *WWW10 Poster Proceedings*, Hong Kong, 2001. www10.org/cdrom/posters/1089.pdf.
- [7] John B. Horrigan and Lee Rainie. PEW Internet & American life project: Getting serious online, March 2002. www.pewinternet.org/reports/reports.asp?Report=55&Section=ReportLevel11&Field=Level11ID&ID=241.
- [8] Inktomi. MSN search engine, 2002. search.msn.com.
- [9] Danny Sullivan. Nielsen//NetRatings search engine ratings. Web Site, September 2002. www.searchenginewatch.com/reports/netratings.html.
- [10] Danny Sullivan. Who powers whom? search providers chart. Web Site, September 2002. www.searchenginewatch.com/reports/alliances.html.
- [11] Bob Travis and Andrei Broder. Web search quality vs. informational relevance, www.infonortics.com/searchengines/sh01/slides-01/travis.html 2001.
- [12] Trystan Upstill. Full bookstore/ISBN query/result list. Web Site. cs.anu.edu.au/~Trystan.Upstill/pubs/results/bstore-rels-02.html.
- [13] Yahoo! Yahoo! business and economy > shopping and services > books > booksellers, September 2002. www.yahoo.com/Business_and_Economy/Shopping_and_Services/Books/Booksell%ers/.
- [14] New York Times. Bestsellers. Web Site, September 2002. www.nytimes.com/2002/09/01/books/bestseller/.
- [15] YourAmigo. Youramigo spider linker, 2002. www.youramigo.com/downloads/documents/SLWhite.pdf.

6 Appendix

6.1 Book Query Breakdown

- (27) Children's
- (15) Hardcover Advice
- (11) Hardcover Business
- (35) Hardcover Fiction
- (29) Hardcover Non-Fiction
- (15) Paperback Advice
- (07) Paperback Business
- (35) Paperback Fiction
- (32) Paperback Non-Fiction
- **(206) Total**

Note: duplicates were removed from the query set. (e.g. Stupid White Men was in both the Hardcover Business and Hardcover Non-Fiction sections, and so was only considered in the Hardcover Business category.)

6.2 Bookstores

Table 7 contains the list of evaluated bookstores.

6.3 Query Submission

- **AltaVista**
 - *General Queries*: Book title surrounded by quotation (“) marks.
 - *URL Coverage*: canonical domain name with “url:” parameter.
 - *Link Coverage*: canonical domain name with “link:” parameter.
 - *Timeframe*: General and Domain Restricted Queries submitted between 20/09/02 and 02/10/02 Link Coverage and URL Coverage experiments performed on the 09/10/02.
- **AllTheWeb (Fast)**
 - *General Queries*: Book title with exact phrase box ticked.
 - *URL Coverage*: Advanced search restricting to domain using “domain” textbox with canonical domain name.

Table 7: **Bookstores included in our evaluation.** This table reports whether the bookstore contained ISBNs in its internal URLs (URL), whether they used question marks in internal URLs (Dyn.), whether they were a derivative of another site (Deriv.) and how many of the 9 book categories they matched (Cat.). A “*” next to the “Deriv.” column indicates that the site was a *booksense.com* derivative, while a “+” indicates that the bookstore was a *booksite.com* derivative. A partial in the “Dyn” column indicates that the site was dynamic but didn’t “look” dynamic (it didn’t have a “?” with parameters following the URL). Note that this table includes all 38 evaluated bookstores. All other tables only consider *Powells* and the 14 bookstores that were returned by search engines at least once.

Bookstore	Core URL	Deriv.	Dyn.	URL	Cat.
1BookStreet	1bookstreet.com	N	Y	ISBN	9
A1Books	a1books.com	N	Y	ISBN	9
AllDirect	alldirect.com	N	Y	ISBN	9
Amazon	amazon.com	N	Partial	ISBN	9
Americana Books	americanabooks.com	N	Y	-	7
Arthurs Books	arthursbooks.com	N	Y	ISBN	4
Barnes and Noble	barnesandnoble.com	N	Y	ISBN	9
BookWorks	bookworksaptos.com	Y*	Y	ISBN	9
BookSite	booksite.com	Y+	Y	ISBN	9
Changing Hands	changinghands.com	Y*	Y	ISBN	9
ecampus	ecampus.com	N	Y	ISBN	9
NetstoreUSA	netstoreusa.com	N	Partial	ISBN	9
Planet Gold	planetgold.com	N	Y	-	9
TextbookX.com	textbookx.com	N	Y	ISBN	9
VStore	vstore.com	N	Y	ISBN	3
Sam Weller’s Books	samwellers.com	N	Y	ISBN	9
All Textbooks 4 Less	alltextbooks4less.com	N	Y	ISBN	9
The Book Shop	bookshopmorris.com	Y*	Y	ISBN	9
Cornwall Discount Books	cornwalldiscountbooks.com	N	Y	-	8
A Lot of Books	alotofbooks.com	N	Y	-	3
HearthFire Books	hearthfirebooks.com	Y*	Y	ISBN	9
Walmart	walmart.com	N	Y	-	9
Wordsworth.com	wordsworth.com	N	Y	ISBN	9
Powells	powells.com	N	Y	-	9
BiggerBooks.com	biggerbooks.com	N	Y	ISBN	9
That Bookstore in Blytheville	tbib.com	Y*	Y	ISBN	9
StrandBooks.com	strandbooks.com	N	Y	ISBN	7
St. Marks Bookshop	stmarksbookshop.com	Y*	Y	ISBN	9
RJ Julia	rjulia.com	N	Y	ISBN	9
Paulina Springs Book Company	paulinasprings.com	Y*	Y	ISBN	9
Books-A-Million	booksamillion.com	N	Y	ISBN	9
CodysBooks.com	codysbooks.com	Y*	Y	ISBN	9
The Concord Bookshop	concordbookshop.com	Y*	Y	ISBN	9
Dartmouth Bookshop	dartbook.com	Y*	Y	ISBN	9
GoodEnough Books	goodenoughbooks.com	Y*	Y	ISBN	9
MediaPlay.com	mediaplay.com	N	Y	-	9
Northshire Bookstore	northshire.com	N	Y	ISBN	8

- *Link Coverage:* Advanced search using Word Filter with “Must Include” in the preceding drop down box, canonical domain name in middle text box and “in the link to URL” in the final drop down box.
- *Timeframe:* General and Domain Restricted Queries submitted between 20/09/02 and 02/10/02 Link Coverage and URL Coverage experiments performed on the 09/10/02.

- **Google**

- *General Queries:* Book title surrounded by quotation (“) marks.
- *URL Coverage:* Search for the non-presence of a non-existing word (-adsljflkjlkjdfkjasdlfj0982739547asdhkas) and using canonical domain name with “host:” parameter.
- *Link Coverage:* Not available.
- *Timeframe:* General and Domain Restricted Queries submitted between 20/09/02 and 02/10/02 Link Coverage and URL Coverage experiments performed on the 09/10/02.

- **MSN Search (Inktomi)**

- *General Queries:* Advanced search with book title as an “exact phrase box”.
- *URL Coverage:* Advanced search using the domain name as the query, and restricting domain using “domain” textbox with canonical domain name.
- *Link Coverage:* Not available.
- *Timeframe:* General and Domain Restricted Queries submitted between 20/09/02 and 02/10/02 Link Coverage and URL Coverage experiments performed on the 09/10/02.