Printed books and the WWW

Paul Nieuwenhuysen
University Library
Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussel, Belgium
32 2 629 2436
Paul.Nieuwenhuysen@vub.ac.be

ABSTRACT
This contribution reports on an ongoing investigation and evaluation of freely available tools on the Internet and WWW to discover information about printed books or to find copies of a particular book for sale.
Introduction:
Various information needs exist related to books. The Internet and WWW provide many services to discover information.
Problem statements / aims:
Which information services are available free of charge for all users via the Internet and the WWW, which allow us to find information concerning various aspects of a particular printed book title? Furthermore, how well do these perform?
Methodology:
Information services that are relevant in the framework of this investigation have been identified. Their features and mechanisms have been compared. Some systems that looked promising have been evaluated in more detail by searching for information concerning particular, individual book titles, simulating normal usage.
Results / findings:
Internet services to find information about books can be categorized as dealing 1. with only one collection, 2. with more than one collection in a single, based on a single, merged database, 3. federated search services that exploit some of the other services simultaneously in one search action. The federated search systems offer in general the highest coverage. Therefore these are investigated in more detail.
Conclusion / value:
This investigation clarifies the state of the art of the usefulness of book search systems on the Internet and WWW for librarians as well as other users.

KEYWORDS: Internet, WWW, books, booksellers, search engines, multi-dealer database, federated searching, addall.com / usedaddall.com, bookfinder.com

1. INTRODUCTION / BACKGROUND
Various information needs exist related to books.
You may want to find / discover
- books with specific words in their title,
- books by a specific author,
- more detailed bibliographical information about a book that you know already,
- booksellers who have a specific, known book for sale,
- the price of a book that you want, as set by sellers,
- the value of a book that you own already,
- the shipping cost, if you buy the book from a particular seller
- ...

When such information needs occur, then various information sources on the Internet and WWW can be useful. Therefore knowledge about such sources can be useful
- for end users, as well as
- for librarians or other information professionals,
  - who buy books,
  - who serve their users by performing searches for books,
  - who propose databases to their users on their library WWW site or
  - who want to include one or several book search engines on their WWW site or in their own local system for federated searching through several targets in one action.

This brings us to the following problem statements and aims of our investigation.

2. PROBLEM STATEMENTS / AIMS
1. Which information services are available free of charge for all users, via the Internet and the WWW, which allow us to find information concerning various aspects of a particular printed book title? Can we distinguish various categories of services to get a better view on the possibilities?
2. How well do these perform when information is needed concerning a book, as specified above?

3. NOMENCLATURE
Various authors use different words and phrases / terms for very similar or identical concepts. The following words are used as synonyms in this text:

- Used book = second hand book
- Dealer = seller = shop = store
- Search engine = search system = search service = search tool
- Multi-dealer database = database obtained after merging smaller dealer databases = aggregator database = listing service = resale network = virtual mall
- Federated search = meta-search = multi-search = poly-search = cross-database search = gateway search = search network
- Target database = underlying database

On the other hand, a "book title" is not the same as a physical "book copy" or "copy of a book"; each "book title" corresponds with many "book copies" of that "book title".

4. FINDINGS

Identification and categorization of book search services

Many automated services / systems exist on the Internet and the WWW, which we can use to find information related to book titles and to more concrete, physical copies of books available for sale from some dealer. A reasonable way to categorize these is the following.

1. Basic databases
   Many basic databases allow us to search in the catalogue of one book publisher or bookseller or library or even in the full text of a collection of books.
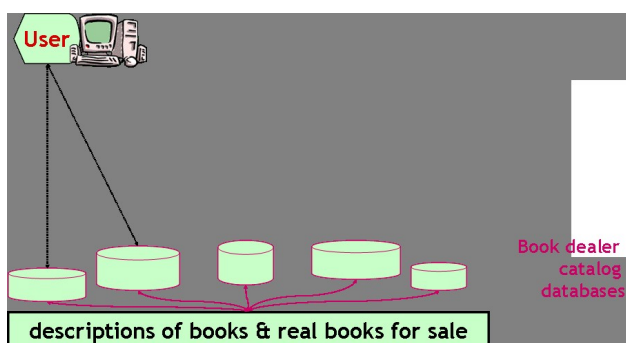   The scheme in Figure 1 focuses on the databases of book dealers.

Figure 1: Scheme of search systems for books, showing only the basic databases produced by book sellers.

2. Merged databases
   Several searchable databases allow us to search through the catalogues of several booksellers or libraries or even through the full text of some collections of books in one action.

Each of these databases is produced and obtained by merging several existing smaller databases, for instance inventory databases that are managed and updated by individual book dealers. Such a system can include the contents from a few to more than 10000 individual dealer databases. These can be considered as virtual shopping malls.

"Resale networks allow book sellers to upload listings of their books into one central database. The book buyer may search this database to obtain a list of book sellers selling the book they need. Such networks are ideal for book sellers who want to reach a mass audience and perfect for book buyers wanting to search thousands of book stores using one interface." (Books online)

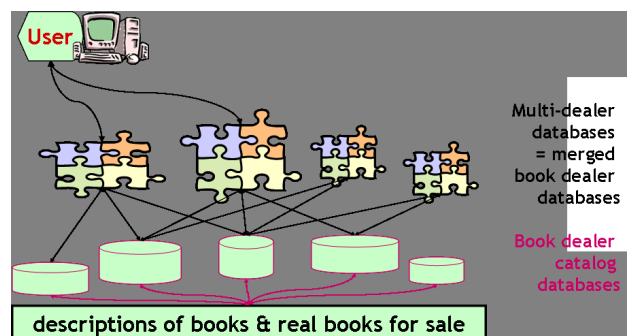Multi-dealer databases besides basic book databases are included in Figure 2.

Figure 2: Scheme of search systems for books, including multi-dealer book databases.

Examples of international rather than nationally oriented systems are
- http://www.abebooks.com/
- http://www.alibris.com/
- http://www.amazon.com/
  started as a book seller with a corresponding catalog database only and has evolved into Amazon Marketplace which is nowadays a bookseller plus a gateway to other booksellers, supported by one search system
- http://www.antiqbook.com/books/
- http://www.biblio.com/
  (or http://biblio.com/)
- http://www.choosebooks.com/
- http://www.tomfolio.com/

Some of these systems offer a potential buyer a link to the real bookshops that have the book for sale. Others allow a potential buyer to buy immediately on their site and communicate afterwards with the real dealer.

In recent years the full text search system Google Books receives a lot of attention and creates a lot of controversy. This is an interesting and powerful service due to
- its large size which is still growing fast
- the powerful full-text search features.

The system allows a user
- to identify book titles that may be interesting or relevant
- to see small fragments of a book
- to download the contents of some books which do not suffer from copyright restrictions

However, Google Books is not considered in the quantitative investigation reported here, as it is not primarily aimed at finding a book for sale. Instead it refers an interested user to some of the book seller databases that form the main subject of this paper.

3. Federated search systems
Federated search systems even allow us to search in one action through many of the systems mentioned above: databases as well as merged, multi-dealer databases. Analogous systems are available for library catalogues. These are gateway systems: they transmit your query to the real databases, they collect and merge the results, and show them finally in a useful way. Figure 3 shows this in the form of a general scheme.
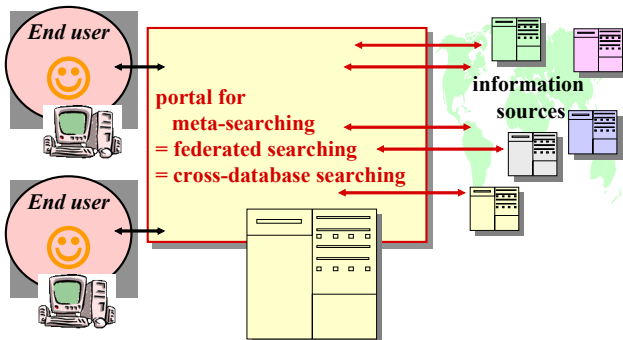

Figure 3: Scheme of a federated search system.

In Figure 4, federated search services are included besides the other book search services mentioned above
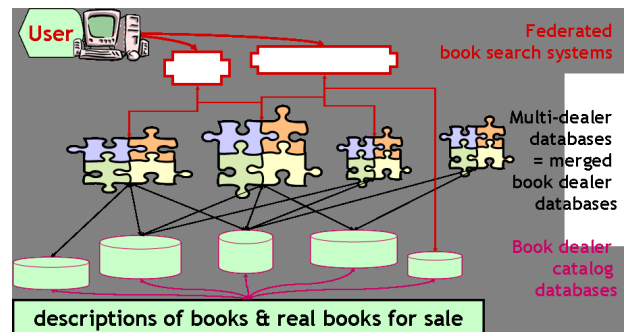

Figure 4: Scheme of search systems for books, including federated search system.

The following are examples of gateways to search databases produced and maintained by book dealers:
- http://www.addall.com/ with http://used.addall.com/ (for used books)
- http://www.allbookstores.com/
- http://www.bookbutler.com/ (earlier http://nl.bookbutler.com/)
- http://www.bookfinder.com/
- http://www.bookfinder4u.com/
- http://www.bookpursuit.com/
- http://www.eurobuch.com/
- http://www.fetchbook.info/
- http://www.isbn.nu/
- http://www.marelibri.com/
- http://www.usedbooksearch.co.uk/ books.htm (for used books)
- http://www.vialibri.net/ (for rare books)

Some systems cover many types of shops, including but not only book shops:
- http://www.dealoz.com/ (was earlier http://www.campusi.com/)
- http://www.dealtime.com/
- http://www.epinions.com/
- http://www.priceminister.com/livres-bd

Besides the concrete URLs in the examples mentioned above, some systems operate also from another URL with a different coverage, user interface or language.

Most meta-search systems offer a large coverage, based on the underlying databases. This makes them attractive and efficient to find information about books that are scarce in the sense that no or only few copies can be found with a search system that offers a smaller coverage. They allow us to find
- books with specific words in their title,
- books by a specific author,

- more detailed bibliographical information about a book that you know already.

The systems to search bookseller databases also allow us to find
- booksellers that have a specific book for sale,
- the value of a book that you own already or that you consider to buy,
- the price of a book that you want, as set by each seller,
- the shipping cost, if you buy the book from a particular seller.

Figure 5 shows finally a scheme of the relations among the systems that allow you to search for books available from dealers.
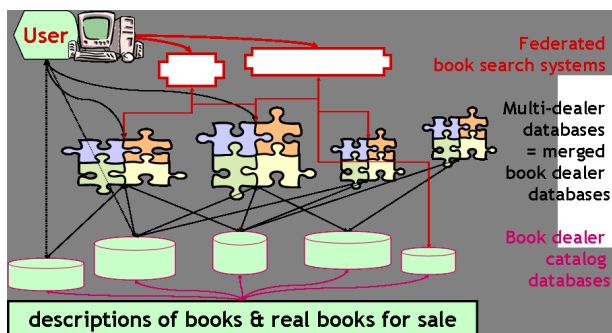

Figure 5: Scheme of search systems for books on the Internet / WWW.

Comparing the different types of search services, we note the following:
- Multi-dealer merged databases react faster than federated search systems, at least in most cases. This is caused / explained by the following facts: Merged databases do not need several simultaneous Internet connections and they do not have to merge raw preliminary results into the result that is shown to the user.
- On the other hand, meta-search systems offer a higher coverage in most cases. This is caused / explained as follows: Federated search systems can exploit not only basic databases but also several multi-dealer databases in one search action. In one search, they can cover more than 100 million couples of book & dealer (note: this is not the same as book titles).

For a user of a book search engine, it is not always clear in which of the categories mentioned above it fits best. Many users are not aware of the technical differences or they do not care. Even if they do care, then the help or

FAQ pages of the system do not always clearly and fully explain the workings of the system. Most systems are available free of charge for any user, but they are commercial systems in the sense that the producer/owner has to make profit to stay in business; so they are not public open systems that reveal all clearly how they work to interested users.

The terms meta-search or federated search system or service are sometimes used for both multi-dealer databases and strict federated search system (see for instance Neiger). This is confusing if one wants to understand and exploit the differences between the two types of search services.

Comparison of federated search services specialised in books

In spite of the popularity and usefulness of these systems on the Internet, a comparative evaluation of their features and performance is not readily available, as far as we know.

The texts, manual, help pages etc… on their WWW site does not allow a user to draw conclusions about the coverage of couples book & seller by the search services that form the subject of this work. More generally, a comparative evaluation of their performance these days, including features, user interface, ease of use, response times, coverage or relevant search results, etc… is not readily available, as far as we know.

Nevertheless, knowledge about their quality is interesting
- as for librarians as well as for end users who buy books,
- for librarians who serve their users by performing searches for books,
- for librarians who propose databases to their users, for instance on their library WWW site, or who want to include one or several book search engines in their own local system for federated searching through several targets in one action.

Therefore we investigate this more closely.

Which federated search engines for books are available and worthy to investigate?

McKinney (2004) used addall, bookfinder and usedbooksearch with (only) two scarce books, demonstrating the usefulness of the relatively new search services.

A list of multi-dealer databases, as well as a list of federated search systems is given for instance by the Independent Online Booksellers Association, by Bookshops Online and by Hanselaer.

More general and popular WWW directories give a list of book search services, but they make no clear difference between multi-dealer databases and federated search systems; see for instance Yahoo! and the DMOZ open directory project.

Harris produced another list of "Book-finder Services on the Internet".
Examples of some book search services that we discovered are mentioned above.

Most or all comparisons have not the simple outcome that one system has a larger coverage than another. Instead, the coverage of two systems overlaps in many cases, so that in some cases system 1 will perform better than 2 and in other cases the reverse will be true. So simple concepts like larger, bigger, better, superior make no sense.
Furthermore, the concept of overlap of coverage can have several meanings as clarified in the following text and the corresponding figures:

- Overlap of covered single-dealer databases:
  Multi-dealer databases have an overlapping coverage of single-dealer, basic databases. Furthermore, at a higher level, meta-search systems can search single-dealer databases as well as multi-dealer databases in one search action.
- Overlap of covered book titles:
  Each "title" stands for the corresponding set of real, physical, concrete copies/objects.
- Overlap of covered couples book & dealer:
  Each couple consists of one or several real, physical, concrete copies of a book title & a book dealer who offers this book for sale.
  A user may expect that a particular book copy is sold by only one dealer, or in other words, that each book copy appears only in one couple book & dealer. However, reality is more complicated, as a listing service that provides access to a multi-dealer book database can also function as a dealer directly with the user of the multi-dealer database inventory catalogue; in such a case, the listing service acts as an intermediary and not as a direct dealer, but that may not be clear for a user and may not matter to the user. So a book copy can appear in more than one couple book & dealer.
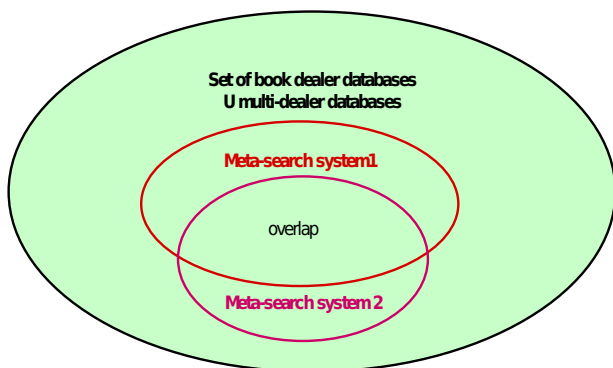

Figure 6: Meta-search systems for book databases have an overlapping coverage of book dealer databases
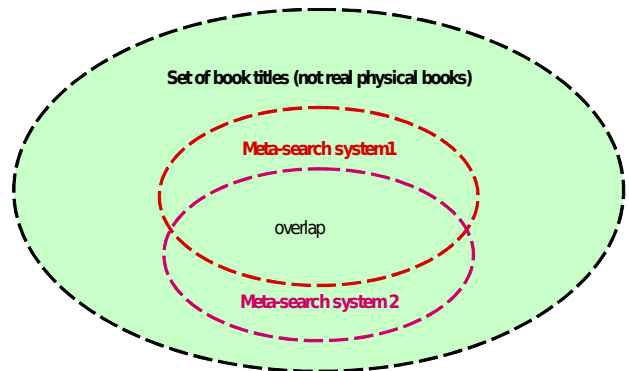

Figure 7: Meta-search systems for book databases have an overlapping coverage of book titles (while each book title can be available from one or more dealers).
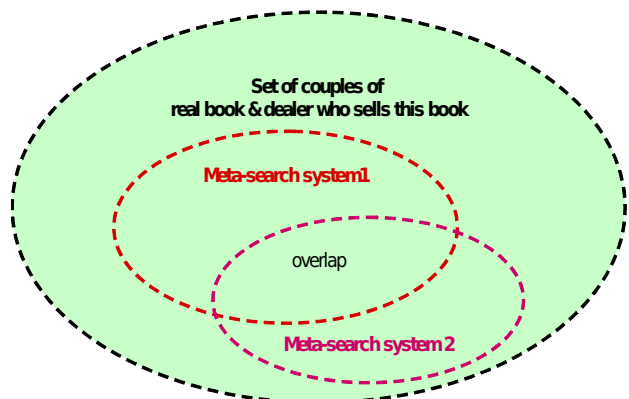

Figure 8: Meta-search systems for book databases have an overlapping coverage of couples book & dealer.

Moreover, what is interesting and relevant in practice is the set of results obtained after a search; however, the coverage and overlap of underlying databases may be not the only factor that determines this outcome; also the way in which these databases are exploited can shape the result. For instance factors can be

- the algorithms and protocols for communication with each underlying database, and for merging of preliminary result sets,
- the internet connections, and so on.

In view of these considerations and the complexity observed, a more or less "blind", automated, computerised test was not developed. Instead this investigation is made by executing searches like normal users exploit these systems and by interpreting the retrieved entries that are shown on the computer display

as well as possible, knowing more or less how the systems work and being aware of the complications mentioned above.

Of course, for this investigation we did not use only one single system, like most users do, but several or at least two search systems were compared almost simultaneously, one after the other.

When a first inspection of unrefined results show that probably less than about 9 couples book & dealer would be revealed, then the results are inspected more closely:

- Results from multi-dealer databases are expanded to reveal all couples of book & dealer.
- Duplicate results for couples of book & dealer are intellectually united / deduplicated.
- The finally obtained unique couples book & dealer are noted and counted.
- Then another search system is tested.
- To avoid possible bias, the order of search systems should be random.

This investigation is carried out since 2007. This minimizes the influence of a temporary failure or malfunctioning on the results.

The results obtained over a period of about a year in the period 2007-2008 have been published earlier, but only quite briefly, due to space limitations in the printed conference proceedings (Nieuwenhuysen, 2008).

Book titles chosen for the investigation were published mainly after 1940, in several countries and languages.

Most were in the subject domain of arts and humanities, which is interesting for older books, as they do not get outdated as fast as books in other domains like science, technology and medicine.

Book titles were included that have been published by famous publishers for a wide audience, but also books published in fewer copies, for instance at the occasion of a temporary exhibition, as well as auction catalogs that are harder to find and thus more interesting in a comparison of search services.

Working with federated search engines revealed that this is more complicated than directly and simply using single-dealer book databases or even multi-dealer databases. Thus applying those systems and certainly a comparative quantitative investigation like this one becomes more time consuming than may be expected. The following gives various observations:

- A meta-search engine may search through dealer databases 1, 2, etc… as well as through multi-dealer databases A, B, etc… in one action. Searching each of those underlying target databases can give results. If the federated search service presents these entries to the user without merging/deduplication of very similar book descriptions, then each of the entries obtained from a multi-dealer database may further lead to one OR MORE THAN ONE couples of book & dealer.

In other words, a meta-search engine can give a result list in a first stage, which contains already a number of relevant entries, but in a second stage, each of these relevant entries can eventually lead to more than one couple book & dealer.

- Furthermore, the contents covered by single-dealer databases 1, 2, etc… and multi-dealer databases A, B, etc… overlap in many cases, as database 1, 2 etc… can by definition be included in database A, B, etc. So searching all these targets in a single action may lead to several entries in the unrefined displayed results that we can call duplicate, in the sense that they refer to only one and the same couple book & dealer.

Summarizing the two complications mentioned above: A federated search service does not always simply and directly provide a result set that consists only of couples of a concrete book & dealer who offers this book for sale. In other words, not all federated search systems implement full deduplication or merging of very similar results or this action is not working ideally. This difficulty is important in the sense that it makes normal usage of such a system rather cumbersome and it does hinder a quantitative investigation. This behavior is not exceptional or remarkable; it is encountered with many federated search systems for many kinds of databases.

Some meta-search sites offer not just one but two search systems: one for new books and one for used or rare books. This may be inconvenient, if the user/searcher is not interested in the distinction between new and used books, as two systems should be used to find both new as well as used book copies. Furthermore, the difference between normal and "rare" will not be clear in many cases. Examples of systems that take this approach are the two related systems addall.com and used.addall.com, and www.dealoz.com (earlier campusi.com) which offers a separate system for "rare" books.

Of course the alternative approach in which the user/searcher does not have to choose in advance between new and used books or between new and rare books gives in many cases in the search results new as well as used books more or less mixed or separated on the display by the computer system. In that case, a user who is only interested in for instance new books may find this inconvenient. For instance bookfinder.com takes this approach.

In the settings of the preferences, the user can in some cases choose the destination country (for instance in the campusi system). However in some cases only few countries are included in this option. More surprisingly, it turned out that not only the price estimated for shipping depends on this setting, but also the number and type of found shops/dealers.

In bookfinder.com, in every search action, the user must choose the language of the book from a pick list. This takes time and is not user friendly. Furthermore, a restriction seems to be that books in many languages cannot be found, as it is obligatory to indicate the language of the book, while available choices are limited to English, French, German, Italian; at the end of 2007 Dutch has been added. So for instance Latin and Spanish are not included.

When English was chosen, even a book with a title in Dutch for instance could be found.

However, in the case of a book that exists in three versions (in English, French, Dutch) the English version was found without problem when selecting English as language, but it was not clear how to find the Dutch version.

Searching by ISBN does not work identically in all systems. For instance: searching with an ISBN that contains spaces was not accepted by bookfinder.com, but was well interpreted by used.addall.com.

The following reports on the comparative quantitative evaluation of several federated searching systems.

On the one hand, most search systems give many results in the case of well-known, popular books. On the other hand they give no results in the case of many rare or "obscure" books. Therefore comparisons of search systems make sense mainly in the intermediate cases in which there exist only few couples of book & dealer, so that perhaps significant differences between the search systems can be investigated and revealed.

To start with, four systems were compared.
Two occur in most published lists of book search services, such as DMOZ, Hanselaer, IOBA, Yahoo!:
- http: //www.bookfinder.com/ for both new and used books
- http: //www.addall.com/ with http: //used.addall.com/ for used books.

Many federated search systems publish on their WWW site the book databases that they target and when we compare this coverage of these search systems with the coverage of competing systems then they turn out to be big players in this field. This agrees with the comparison of how many databases are covered by a few meta-search engines as shown in a table by Hanselaer.
Two additional systems were included in comparative testing, which were known by us as functioning well, but which were NOT listed by Hanselaer, IOBA and Yahoo!:
- http://www.campusi.com/ which leads more recently in 2009 to http://www.dealoz.com/ and which is a general service to support comparative shopping, not dedicated to books only
- http://nl.bookbutler.com/ which leads more recently in 2009 to http://www.bookbutler.com/

Federated search systems react slower than "real" database search engines in most cases. Nevertheless the tests showed that response times were in the order of a few seconds only, which should be acceptable for most users.

In a series of tests, 14 test cases showed a number of retrieved couples book & dealer, which was small enough to allow and motivate a more detailed quantitative comparison. In the majority of these cases (13 of 14) either addall / used.addall or bookfinder gave more couples book & dealer than bookbutler or campusi. This is in accordance with the fact that they are included in the directories of book search services mentioned above.

Of the search systems investigated, only the campusi system for rare books referred occasionally to a book that was for sale through the famous and popular Internet auction system Ebay, by coincidence at the time of testing. This occurred in 3 of the 14 test cases. In 1 of the test cases this yielded the only couple book & dealer. So this feature can be seen as a plus.

Subsequent tests were executed only with the federated search services addall / used.addall and bookfinder, because the test outcome outlined above indicates that these find more couples of book & dealer.

For most book titles either no result is found or quite many, so that in these cases the search systems offer equal value. However, with all the book titles used for testing, 45 books fell between those extremes of untraceable and abundant books. With these book titles we revealed differences in the couples book & dealer that were retrieved by the investigated search engines.
We can count a search system as winner after an search with a single book title, when that system finds for that specific book title more couples of book & dealer than the other, competing search system(s). The number of times that a system could be called a winner in this way was:
- used.addall.com   22
- bookfinder.com   7
Besides these cases, no winner came out in 16 other comparisons.
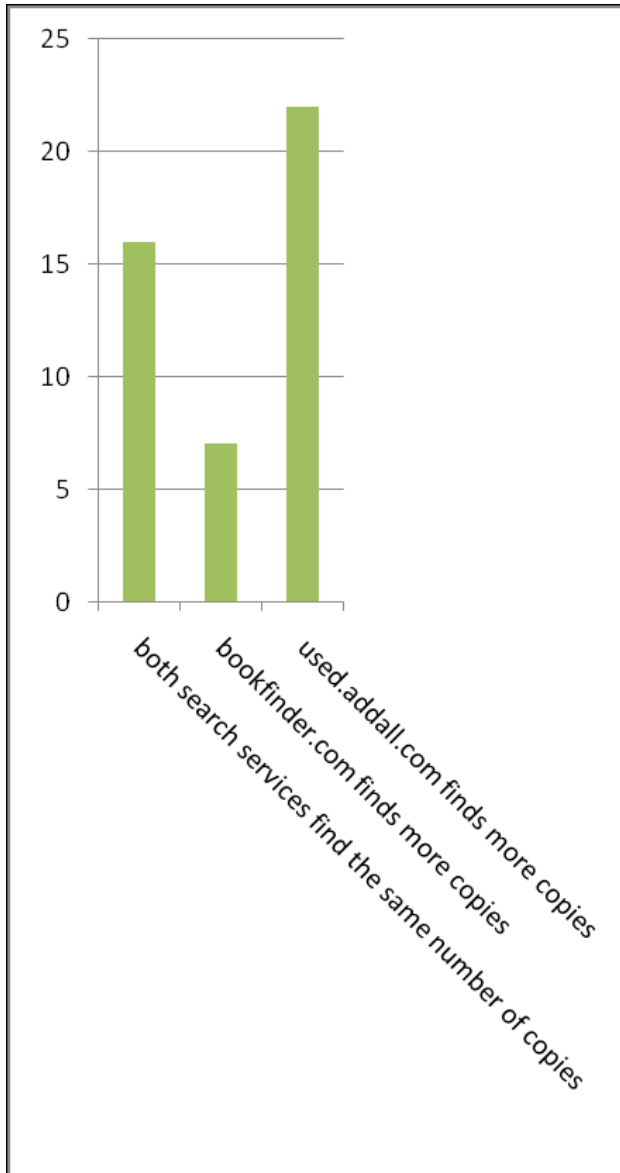This is shown on Chart 1.

Chart 1: Comparison of the performance of federated search engines to find a copy of a particular rare book title offered for sale by some dealer.
In this view, what counts is the number of cases in which one search system has found more copies of the book than the other system.

We can also count a search system as a winner after a search for a single book title, in a more restrictive way, namely when that system finds at least one dealer (occurrence) for that book title, while the other system finds NO dealer at all. The number of times that a system could be called a strict winner in this way was:

- used.addall.com   7
- bookfinder.com   4

Besides these cases, no winner came out in 34 other comparisons as both search systems found a copy of a book title.
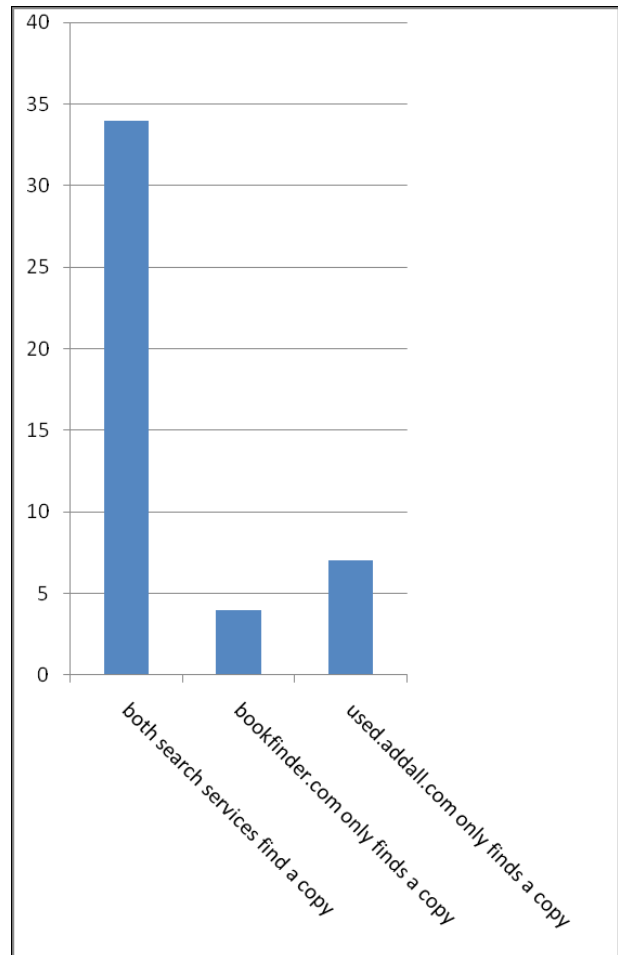This is shown in Chart 2.

Chart 2: Comparison of the performance of federated search engines to find a copy of a particular rare book title offered for sale by some dealer.
In this view, what counts is the number of cases in which one search system has found at least one copy of the book, while the other system found none.

In conclusion, used.addall.com performed best in this quantitative investigation.

All the numbers given above as the outcome of our test cases are not more than that. In other words, they should not be generalized as a reflection of the potential of the investigated search engines for EVERY book. Indeed, the quantitative results from an investigation like this depend on the books chosen in the test cases. A search engine may perform better for one type of book (older versus more recent, new versus used, subject domain, language, value and price) since the target databases are different; some of the dealers and thus the target dealer databases

are specialized in a particular type of books. Therefore we have used various types of books, as outlined above, but choosing a random sample is not straightforward; even defining what would be random seems impossible as the population discussed is quite heterogeneous.

Having discovered some flaw in a system, like problems related to searching on ISBN or language, this can of course be taken into account to favor one system or another in subsequent tests. As the purpose of this work was not to shine a bright light on a particular system, but to inform users objectively, such a way of proceeding has been avoided here as far as possible. This is mentioned here to underline that some other test may well come up with other numbers, even when the investigator does not want to favor any investigated service.

5. CONCLUSION / VALUE

This investigation clarifies to some extent the state of the art of book search services on the Internet and WWW for librarians as well as other users.

For an abundant book title, most of the systems that search bookseller databases will satisfy the user in most cases. However, in the case of a scarce book title, it is wise to choose a federated search system with a wide coverage. Further, more than one search system can serve as complementary search tools for several reasons:

- Their way of working and user interface is far from identical, even though their aims are quite similar.
- Their coverage is overlapping but has unique parts, so that a book that cannot be found by one system can perhaps be retrieved by the other system.

6. REFERENCES

Bookshops online
Sources [online]
Available from: http: //www.bookshopsonline.com /bookshops /page.asp?tmplt=htm&wbpg=resources [accessed 2008]

DMOZ Open Directory Project
Top: Home: Consumer Information: Price Comparisons: Books [online]
Available from: http: //www.dmoz.org /Home/Consumer_Information /Price_Comparisons/Books/ [accessed 2008]

Hanselaer
Neuzen bij de "bouquinistes" aan de oevers van de Atlantische en Stille Oceaan. Het internet als hulpmiddel bij acquisitie. [online]
Available from: http: //www.slideshare.net /VVBAD /3-hanselaer/ [accessed 2008]

Harris, Roger
Book-finder Services on the Internet [online]

Available from: http: //www.nationalfinder.com /bfs /index.htm [accessed 2008]

IOBA (Independent Online Booksellers Association)
Book related links. [online]
Available from: http: //www.ioba.org/links.html [accessed 2008]

McKinney, Bruce
Search engines: a growing force in books. [online]
http: //www.americanaexchange.com /NewAE /aemonthly /printarticle.asp?from=a&id=132 Published 2004 [accessed 2004, 2008]

Neiger, Michael A.
The book finder. [online]
Available from: http: //therucksack.tripod.com /bookfinder.htm [accessed 2008]

Nieuwenhuysen, Paul
Internet federated search engines for bookseller databases: a comparative evaluation.
In Intelligence, Innovation and Library Services, Proceedings of the Fourth Shanghai International Library Forum = SILF2008 = Shanghai Library, Shanghai, China, October 20-22, 2008. Shanghai Scientific and Technological Literature Publishing House, 2008. 371 pp. ISBN 978-7-5439-3671-3 http://www.sstlp.com pp. 340-348.

Yahoo!
Directory > Business and Economy > Shopping and Services > Books > Book Search Services > Price Comparisons [online]
http://dir.yahoo.com/Business_and_Economy/Shopping_and_Services/Books/Book_Search_Services/Price_Comparisons/ [accessed 2008]