Creating a virtual union catalog for Hong Kong dissertations and theses collections
Irene S.M. Wong and Yiu-On Li
Hong Kong Baptist University Library, Hong Kong, China

Abstract
Purpose – The aim of this paper is to discuss and analyze the efforts of Hong Kong Baptist University Library to create a virtual union catalog for dissertations and theses collections as a single search platform to retrieve the bibliographic records, abstracts and full-texts of Hong Kong postgraduate students’ theses and dissertations from seven university library online public access catalogues (OPACs).

Design/methodology/approach – The paper examines the use of advanced keyword searching functionality in Innovative Interfaces’ Millennium system and Microsoft active server pages scripts to develop an original search engine to cross search the OPACs of Hong Kong university libraries.

Findings – Library searching is more than simply doing a Google search. Libraries do not need to depend on vendors of their library integration systems exclusively for enhancement in OPAC search functionality. Librarians can also take the initiative to maximize the searching capabilities and functionalities of their existing library integration systems, and design new search engines to facilitate patrons to use the OPACs more efficiently and effectively.

Originality/value – The paper discusses the creation of a unique platform, which is a free and valuable reference tool for scholars and researchers around the world to cross search the Hong Kong postgraduate students’ theses and dissertations. It also provides insights into how librarians can maximize the functionalities of library integration systems to enhance cross searching on OPACs.

Keywords Theses, Academic libraries, Archives management, Online catalogues, Generation and dissemination of information, Hong Kong

Paper type Technical paper

Introduction
Teaching and research are two of the essential components of university education. Therefore, it is of the utmost importance to archive, preserve and disseminate the research works of the faculty, staff and postgraduate students at a university, and postgraduate students’ theses and dissertations become significant showcases of a university’s achievement and contribution in higher education. As a result, an increasing number of university libraries are investing more and more resources and efforts in creating, archiving and distributing the citations as well as the electronic theses and dissertations in institutional repositories, union catalogs and dissertation databases, etc. Yet, without a good search platform, full texts and even citations of these postgraduate students’ hard work will be inaccessible on the World Wide Web. In this regard, Hong Kong Baptist University Library has created the Dissertations and Theses Collections (DTC)[1], as a single search platform to bridge the gaps of different
library online public access catalogs (OPACs), and to facilitate the retrieval of Hong Kong students’ master’s and doctoral theses and dissertations.

**Theses and dissertation collections in Hong Kong**

Seven university libraries in Hong Kong[2] catalog their postgraduate students’ master’s and doctoral theses and dissertations (Hong Kong dissertations) in their Innovative Interfaces’ Millennium library integration systems[3]. Therefore, OPACs become the most reliable source of these Hong Kong dissertations.

To search the collections of the seven Hong Kong university libraries, library patrons can cross search the library OPACs via Z39.50[4, 5] which is developed as a physical centralized database for storing bibliographic records indexed in the selected OPACs. However, patrons cannot limit their searches to the collection subset of Hong Kong dissertations only. Therefore, patrons have to search each library OPAC individually if they want to search on all the theses and dissertations written by Hong Kong postgraduate students.

**Scope of DTC**

To simplify the tedious search process, the Hong Kong Baptist University Library created the Dissertations and Theses Collection (DTC) in 2006. DTC is an online reference tool to complement the existing bibliographies and individual library OPAC. It is a comprehensive and multidisciplinary virtual union catalog of over 36,000 Hong Kong dissertations (Table I). It enables patrons to cross search all Hong Kong dissertations dated back to 1941 in a single platform. Since three of these libraries also provide links to all or some of their electronic theses and dissertations in the bibliographic records, therefore, DTC becomes the only platform to:

- facilitate a one-stop broadcast or cross-searching on Hong Kong postgraduate students’ master’s and doctoral dissertations and theses online; and
- retrieve and access their abstracts and full texts if open access is available from the corresponding libraries. (Libraries providing full-text dissertations include Hong Kong University of Science & Technology Library, Lingnan University Library and University of Hong Kong Libraries.)

<table>
<thead>
<tr>
<th>Library</th>
<th>Doctoral theses</th>
<th>Master’s theses</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Hong Kong Libraries (HKU)</td>
<td>2,846</td>
<td>10,959</td>
<td>13,805</td>
<td>37.9</td>
</tr>
<tr>
<td>Chinese University of Hong Kong Libraries (CUHK)</td>
<td>2,349</td>
<td>7,587</td>
<td>9,936</td>
<td>27.2</td>
</tr>
<tr>
<td>Hong Kong Polytechnic University Library (PolyU)</td>
<td>1,024</td>
<td>4,404</td>
<td>5,428</td>
<td>14.9</td>
</tr>
<tr>
<td>City University of Hong Kong Library (CityU)</td>
<td>690</td>
<td>2,551</td>
<td>3,241</td>
<td>8.9</td>
</tr>
<tr>
<td>Hong Kong University of Science &amp; Technology Library (HKUST)</td>
<td>966</td>
<td>2,114</td>
<td>3,080</td>
<td>8.4</td>
</tr>
<tr>
<td>Hong Kong Baptist University Library (HKBU)</td>
<td>207</td>
<td>646</td>
<td>853</td>
<td>2.3</td>
</tr>
<tr>
<td>Lingnan University Library (LU)</td>
<td>6</td>
<td>115</td>
<td>121</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>8,088</strong></td>
<td><strong>28,376</strong></td>
<td><strong>36,464</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

**Table I.**

Number of Hong Kong dissertations indexed in seven university library OPACs as of April 24, 2008
Search mechanism of DTC

All DTC records are created in real time by searching each of the seven university library OPACs through the AVS/ASP search engine that we have designed. The workflow of a search in DTC includes (Figure 1):

- users input their search terms and criteria on a “Search Interface” (Figure 2);
- an in-house designed “AVS/ASP Search Engine” to construct and execute the search query; and
- the “AVS/ASP Search Engine” creates a “Search Result List” of dissertation titles and hyperlinks to their bibliographic records indexed in the corresponding OPACs, whereas links to the abstracts and full texts, if available, will also be displayed (Figure 3).

AVS/ASP search engine

The AVS/ASP Search Engine is specifically designed to cross search and generate search results. It makes full use of the “advanced keyword searching” (AVS) search functionality of the Millennium system, as well as “Microsoft Active Server Page” (ASP) scripts. To execute a search, the AVS/ASP Search Engine performs the following four tasks:

Figure 1. DTC – AVS/ASP search engine program flow chart
(1) reads the search term(s) and criteria inputted by users;
(2) constructs an AVS search query and direct it to the library OPACs selected by the users;
(3) reads and retrieves the bibliographic records that fulfill all the search criteria; and
(4) displays the results on a “search result list”.

**Syntax of an AVS search query**
An AVS search query is actually a URL sent to a Millennium system (Innovative Interfaces (III) (2008)). For example, to retrieve all the Hong Kong dissertation collection in HKBU, the following AVS search query will be generated: http://hkbulib.
A search query is basically made up of the following three components:


2. Search terms from variable fields such as title, author, subject and keyword phrases; asterisk “*” used in the above example denotes a wildcard searching for all records.

3. Fixed fields which direct the search to a specific search scope such as subsets by material or collection types, or by a specific bibliographic location code. In the above example, “b = th++” refers to the bibliographic location of Hong Kong dissertations collection at HKBU Library.

Among the three components mentioned above, the “fixed fields” are the most decisive factor in successfully cross searching all the Hong Kong dissertations subsets in the seven university libraries. It is because no union catalog of Hong Kong dissertations...
subset is available, and university libraries adopt different cataloging practices to identify and differentiate the bibliographic records of theses and dissertations from other collections. Consequently, we have to devise three different types of AVS search queries to retrieve the appropriate bibliographic records from different university library OPACs.

**Search scope/collection scoping**

Both CUHK and HKBU use collection scoping to differentiate their Hong Kong dissertations collections. So, in the AVS search query, we have to use “search scope” to search the designated “dissertation collection”, i.e. “searchscope = [a designated number]”, and an asterisk “*” as a search term to search the whole dissertation collection, e.g.

- a search query to search CUHK OPAC: http://library.cuhk.edu.hk/search/X?SEARCH = *&searchscope = 2&m = &b = &l = &SORT = D&Da = &Db = ; and
- a search query to search HKBU OPAC: http://hkbulib.hkbu.edu.hk/search/X?SEARCH = *&searchscope = 8&SORT = D&l = &m = &p = &b = th + + &Da = &Db =

**Bibliographic location code**

HKUST uses a unique “Bibliographic location code” in their bibliographic records of its Hong Kong dissertations. So we use bibliographic location code of “b = these” and an asterisk “*” in the search query, i.e., http://ustlib.ust.hk/search/X?SEARCH = *&searchscope = 0&SORT = D&l = &m = &p = &b = these&Da = &Db =

**Search terms**

HKU, CityU, PolyU, and LU use “author” or “subject” added entries to distinguish their Hong Kong dissertations. Therefore, we use keyword phrase or search terms derived from variable fields in the search query, e.g.

1. A search query to search Author added entry: HKU inputs “Hong Kong University Theses Online” in Author added entry (USMARC tag 773) to distinguish their Hong Kong dissertations collection, so we use “t:” as the title index tag in the search query as follows: http://library.hku.hk/search/X?SEARCH = t:(Hong%20Kong%20University%20Theses%20Online)&searchscope = 0&SORT = D&l = &m = &p = &b = &Da = &Db =

2. A search query to search Subject added entry: CityU, PolyU and LU use unique Subject added entries for their dissertation collections, so our search queries contain combined keyword phrases to search on the Subject added entries whereas “s:” or “d” are assigned as the subject index tags in Millennium system:

- A search query sent to CityU: http://lib.cityu.edu.hk/search/X?SEARCH = s:(dissertations%20and%20kong)%20and%20s:(city%20university%20or%20city%20polytechnic)&searchscope = 0&SORT = D&l = &m = &p = &b = &Da = &Db = . This is a combined keyword phrase search: “s:(dissertations and kong) and s:(city university or city polytechnic)” on the two following Subject added entries (USMARC tag 690):
Search criteria

To perform a precise search, patrons can further narrow down their searches with some additional search criteria such as year, language, sorting order, choice of library, degree types and department names. The last two search criteria are the most helpful and commonly used search strategies for library patrons. Yet, since the names and forms of degree types and department/school vary differently in these seven universities, then this problem becomes another challenge for us.

To limit a search by master’s and doctoral degrees

As a common practice, catalogers input “degree types” in the note fields, therefore, we need to identify all the doctoral degree titles used in these seven universities first (Table II). Then we construct a keyword phrase search on degree titles. For example, a search query to find all doctoral theses in CUHK will be: http://library.cuhk.edu.hk/search/X?SEARCH = (D.M.%20or%20Ed.D.%20or%20M.D.%20or%20Ph.D.)&searchscope = 2&m = &b = &l = &SORT = D&Da = &Db = . This search query will only retrieve records containing any of the doctoral degree titles of “DM, EdD, MD, PhD”.

<table>
<thead>
<tr>
<th>University</th>
<th>Doctoral degree offered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese University of HK</td>
<td>DM, EdD, MD, PhD</td>
</tr>
<tr>
<td>City University of HK</td>
<td>PhD</td>
</tr>
<tr>
<td>HK Baptist University</td>
<td>PhD</td>
</tr>
<tr>
<td>HK Polytechnic University</td>
<td>DBA, PhD</td>
</tr>
<tr>
<td>HK Uni. of Science &amp; Technology</td>
<td>PhD</td>
</tr>
<tr>
<td>Lingnan University</td>
<td>PhD</td>
</tr>
<tr>
<td>University of HK</td>
<td>MD, PhD, SJD</td>
</tr>
</tbody>
</table>

Table II.

Doctoral degree types in seven Hong Kong universities as at August 31, 2006
On the other hand, a search query to retrieve master’s dissertations and theses is actually a search on all non-doctoral dissertation records which do not contain any of the above doctoral degree titles, i.e. NOT “DM, EdD, MD, PhD”, thus, a search query to retrieve masters’ dissertations will be: http://library.cuhk.edu.hk/search/X?SEARCH = *%20AND%20NOT%20(D.M.%20or%20Ed.D.%20or%20M.D.%20or%20Ph.D.)&searchscope = 2&x = &m = &b = &l = &SORT = D&Da = &Db =

To limit a search by department
A “Department” search enables users to retrieve dissertations submitted by students enrolled in a particular department and school. (Note that Department searching is not applicable to HKU records because the departments or schools in which the students enrolled are not noted in the bibliographic records.) However, the names of departments and schools of the same discipline may vary in different universities. For instance, the Department of Accounting and Accountancy as noted in the bibliographic records are slightly different in different university library OPACs (Table III).

To solve this problem, we use a keyword phrase search on the “Department Names”. In the above example, an AVS boolean search query containing both words of “Accountancy” and “Accounting” will be used: http://library.cuhk.edu.hk/search/X?SEARCH = a:(Accountancy%20or%20Accounting)&searchscope = 2&x = &m = &b = &l = &SORT = D&Da = &Db =

An interesting case arises in this “Department” limit search. The word “Law” appears in some department and school names, but “Law” is also a common Chinese last name. Therefore, in order to limit the search on “Law” as a department or school name but not a personal name, we formulate the search query as “Law near University”, e.g. http://lib.cityu.edu.hk/search/X?SEARCH = s:(dissertations%20and%20kong)%20and%20s:(city%20university%20or%20city%20polytechnic)%20AND%20a:(Law%20near%20university)&searchscope = 0&SORT = D&l = &m = &p = &b = &Da = &Db =

### Merits of the DTC

**A simple, user-friendly search form**
The cross-searching form with self-explanatory search operators and pull-down menus for additional search criteria of “department”, “degree”, “sorting” and “language” requires simple or minimal search strategies, skills and commands from users.

<table>
<thead>
<tr>
<th>Libraries</th>
<th>MARC TAG</th>
<th>Department names</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUHK</td>
<td>710</td>
<td>Chinese University of Hong Kong, bGraduate School, bDivision of Accountancy</td>
</tr>
<tr>
<td>CityU</td>
<td>710</td>
<td>City University of Hong Kong, bDept. of Accountancy</td>
</tr>
<tr>
<td>HKBU</td>
<td>710</td>
<td>Hong Kong Baptist University, bDept. of Accountancy and Law</td>
</tr>
<tr>
<td>PolyU</td>
<td>710</td>
<td>Hong Kong Polytechnic University, bSchool of Accounting and Finance</td>
</tr>
<tr>
<td>HKUST</td>
<td>690</td>
<td>Dissertations, Academic</td>
</tr>
</tbody>
</table>

**Table III.** Variant names of the Department/School of Accounting as displayed in the bibliographic records of five university library OPACs
A flexible and efficient cross-searching platform and effective gateway
There is a built-in smart search algorithm, which enables users to search any one or all the library OPACs simultaneously. Free choices of searching fields and display modes of result list are also available. Users do not need to search various information sources, such as the specialized bibliographies, OPACs, institutional repository or electronic theses and dissertations database of each university library for the theses and dissertations submitted by Hong Kong postgraduate students.

Cost-effective
It requires no additional server for central storage of metadata or other complex broadcasting search information retrieval protocol (e.g. Z39.50, OAI).

Simple workflow and efficient project administration
It creates links to, instead of harvesting and storing, records indexed in other OPACs. It also synchronizes with any record updates in other libraries simultaneously. Thus, no extra effort or cost is needed in copyright clearance, metadata loading, and database maintenance.

Comprehensive statistics
The program automatically logs the date, time, user IP, search terms entered by users, and numbers of search results from the corresponding OPACs. It generates comprehensive usage statistics (Figure 4) for search behavior and use pattern analysis. For instance, usage statistics from September 2006 to March 2008 (Table IV) indicate

<table>
<thead>
<tr>
<th>Year</th>
<th>Hits</th>
<th>On campus access (HKBU) (%)</th>
<th>Off campus access (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 2006 – June 2007</td>
<td>26,000</td>
<td>23</td>
<td>77</td>
</tr>
<tr>
<td>July 2007 – March 2008</td>
<td>23,876</td>
<td>14</td>
<td>86</td>
</tr>
</tbody>
</table>

Figure 4. DTC search statistical log

Table IV. DTC usage statistics
that use of DTC is encouraging and 77-86 percent searches were performed off-campus. All these statistics are useful in helping us to enhance our user education programs and search engine design.

A program for expansion and progress
We can revise the search engine easily and formulate different AVS search queries to include other libraries’ dissertation collection subsets to DTC and create new virtual union catalogs. Having used similar search strategies and algorithm, we have already created three new virtual union collections of different material types and subject categories[6]. They are: Law Library collections (LLC); Medical Library Collections (MLS); and Multimedia Collections (MTC).

Conclusion
Library searching is more than simply Google searching. Libraries do not need to be totally dependent on our library integration system vendors for enhancement in OPAC search functionality. Being the librarians, we can also take initiatives to maximize the searching capabilities and functionalities of our existing library integration system, and design new search engines to facilitate patrons to use our OPACs more efficiently and effectively. In fact, the Dissertations and Theses Collection (DTC) was created with just this mission in our mind.

DTC is a valuable reference tool to both regional and international intellectual communities. It is a free service for scholars and researchers around the world to trace the research work undertaken by postgraduate students in Hong Kong. It provides links to the full texts of some electronic dissertations and theses and thus facilitates information dissemination on the World Wide Web. According to our usage statistics from September 2006 to March 2008, DTC has an average of 2,625 hits every month, and nearly 80 percent of them are from off-campus access. We hope DTC can serve the purpose of knowledge access and dissemination and become a successful showcase of the research work of local postgraduate students and the achievements of higher education in Hong Kong.

Notes


5. Hong Kong Academic Library Link (HKALL) is an online union catalogue allowing students and staff of the eight partner institutions to search a mega collection of over 4 million titles of monographs held in the eight participating libraries of HKALL. HKALL web site, available at: hkall.hku.hk/ (accessed 28 April, 2008).


Reference

About the authors
Irene S.M. Wong is currently the Special Collections Librarian and Archivist at the Hong Kong Baptist University Library. She holds an M.LS (Master of Library Services) from Rutgers, State University of New Jersey. She has published several articles in journals. Her research interests include digitization, preservation and archival management. Irene S.M. Wong is the corresponding author and can be contacted at: ismwong@hkbu.edu.hk

Yiu-On Li is the Senior Assistant Librarian (Systems) at the Hong Kong Baptist University Library. He holds an MALIS (Master of Arts in Library and Information Studies) from the University of Wisconsin-Madison, an MSc (Computing) from the City University of Hong Kong, an MPhil., and BA (Philosophy) from the Chinese University of Hong Kong. His research interests include digital library, and application of computing and information technology in library services and management. Yiu-On Li can be contacted at: yoli@hkbu.edu.hk