

Converting Millennium ILS Bibliographic records into Dublin- Core XML format for DSpace

PNC 2009 Annual Conference and Joint Meetings
Taipei, Taiwan

Alan Ng

Hong Kong University Libraries

Introduction

HKU Libraries

- established in 1912
- the oldest academic library in HK
- main library and 6 branches

HKU Libraries

- 2.84M total physical volumes
- 49K print periodical titles
- 80K electronic periodical titles
- 1.90M e-book

HKU Libraries

- Millennium ILS from Innovative Interface Inc.
- hosting the HKALL union catalog for 8 university libraries in HK

Institutional Repository

HKU Scholars Hub

- collects intellectual output of HKU for fulltext open access
- <http://hub.hku.hk/>

HKU Scholars Hub

- uses DSpace (version 1.5)
- OAI-compliant
- implements DCMI

HKU Scholars Hub

- 25300+ records (as of 2009 June)
- Articles
- Conference paper
- Postgraduate thesis and others
- 1.6M download (as of 2009 June)

HKU Scholars Hub

- some records originate from the OPAC
- HKU postgraduate thesis
- Digital editions from HKU Press
- Bibliographic MARC fields are mapped to DC XML data

MARC to DC mapping

001	identifier -- other
008	language
020	identifier -- isbn
022	identifier -- issn
050	subject -- lcc
092 a b	subject -- dcc
110 a	contributor -- author
245 a b	title
260 b	publisher
260 c	date -- issued
300 a b c	format -- extent
490 a	relation -- ispartofseries
5XX	description
650	subject -- lcsch
710 a b	contributor -- other
856 u	identifier

A record in OPAC

Start Over Export MARC Display Another Search

RECORD # b4200627 View Entire Collection Search

Limit results to available items

AUTHOR [Yuen, Wing-ka, Charlotte.](#)
[袁穎嘉.](#)

TITLE **Production of stop consonants by Cantonese-speaking cochlear implant users [electronic resource] / Charlotte, Yuen Wing Ka.**

IMPRINT 2007.
Permanent URL for this record=> <http://library.hku.hk/record=b4200627>

[Click to view the E-thesis via HKU Scholars Hub](#)

OCLC # ocn430281940

DESCRIPT. 32 p. : ill.

NOTE "A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2007."

BIBLIOG. Includes bibliographical references (p. 27-30).

NOTE Also available in print.
Mode of access: World Wide Web.
Thesis (B.Sc)--University of Hong Kong, 2007.

LC SUBJECT [Cochlear implants](#)

ALT AUTHOR [University of Hong Kong.](#)


APPEARS IN [Hub Collection. BSc \(Speech & Hearing\) Theses](#)

OTHER FORM [Yuen, Wing-ka, Charlotte. Production of stop consonants by Cantonese-speaking cochlear implant users Print](#)

ADD THIS
Extend your search
HKALL
WebBridge
ResearchPro
Google Scholar
Manage your Info
Get a Citation
Access Rights
Find out More

<http://library.hku.hk/record=b4200627>

Same record in Hub



Item Metadata (Short)

The Hub	
Advanced Search	
Fulltext Search SCIFUS	
Browse by Author	
Browse by Title	
Browse by Community	
Browse by Type	
Publisher Policies	
Help	
To submit to the Hub? Please contact us	

Title	Production of stop consonants by Cantonese-speaking cochlear implant users
Authors	Yuen, Wing-ka, Charlotte. University of Hong Kong.
Issue Date	2007
Publisher	University of Hong Kong (Pokfulam Road, Hong Kong)
Description	A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2007. Also available in print. Thesis (B.Sc)--University of Hong Kong, 2007.
Persistent Identifier	http://hub.hku.hk/handle/123456789/55513
Other Identifiers	http://lookup.lib.hku.hk/lookup/bib/B42006272
Appears in Collections:	Speech & Hearing Sciences: Theses

Files for this Item:

File	Size	Format	
ft.pdf	297.55 kB	Adobe PDF	View/Open

[Show Item Metadata \(Full\)](#)

<http://hub.hku.hk/handle/123456789/55513>

Automated batch processing

Incentives

- needs to convert 100+ records at a time
- tedious, easy to make mistake manually
- time consuming

Automated approach

- efficiency
- accuracy
- eliminate duplicated effort of data entry
- easier quality control of converted data

Perl programming

- free of charge
- easy to program
- powerful in handling plain text in MARC
- runs on any computer platform
- needs a persistent URL syntax to locate a particular record on OPAC

Perl programming

- reads in a list of bibliographic record numbers
- captures the MARC records on OPAC real time one by one via HTTP
- regards the returned HTML as plain text

MARC record as seen by human

agon Results

Start Over Export Regular Display Another Search

00000nam 2200000Ia 4500
ocn430281940
n d
er mn
090625s2007 xx a sbm 000 0 eng d
HUA|cHUA
|c\$1
HKU Hub
B.Sc.(Sp.&Hear.Sc.)07|bY13
6880-01|aYuen, Wing-ka, Charlotte.
Production of stop consonants by Cantonese-speaking
cochlear implant users|h[electronic resource] /|cCharlotte,
Yuen Wing Ka.
|c2007.
32 p. :|bill.
"A dissertation submitted in partial fulfilment of the
requirements for the Bachelor of Science (Speech and
Hearing Sciences), The University of Hong Kong, June 30,
2007."
Thesis (B.Sc)--University of Hong Kong, 2007.
Includes bibliographical references (p. 27-30).
Also available in print.
Mode of access: World Wide Web.
Cochlear implants
University of Hong Kong.
|tHub Collection. BSc (Speech & Hearing) Theses
Yuen, Wing-ka, Charlotte.|tProduction of stop consonants
by Cantonese-speaking cochlear implant users|cPrint
|uhttp://lookup.lib.hku.hk/lookup/bib/B42006272|x126
|zClick to view the E-thesis via HKU Scholars Hub
6100-01/\$1|a袁穎嘉.
iiky|econ
Speech and Hearing Sciences
y|cip|1126|dHKU Hub|eth|go|tt|s6178

MARC record as seen by program

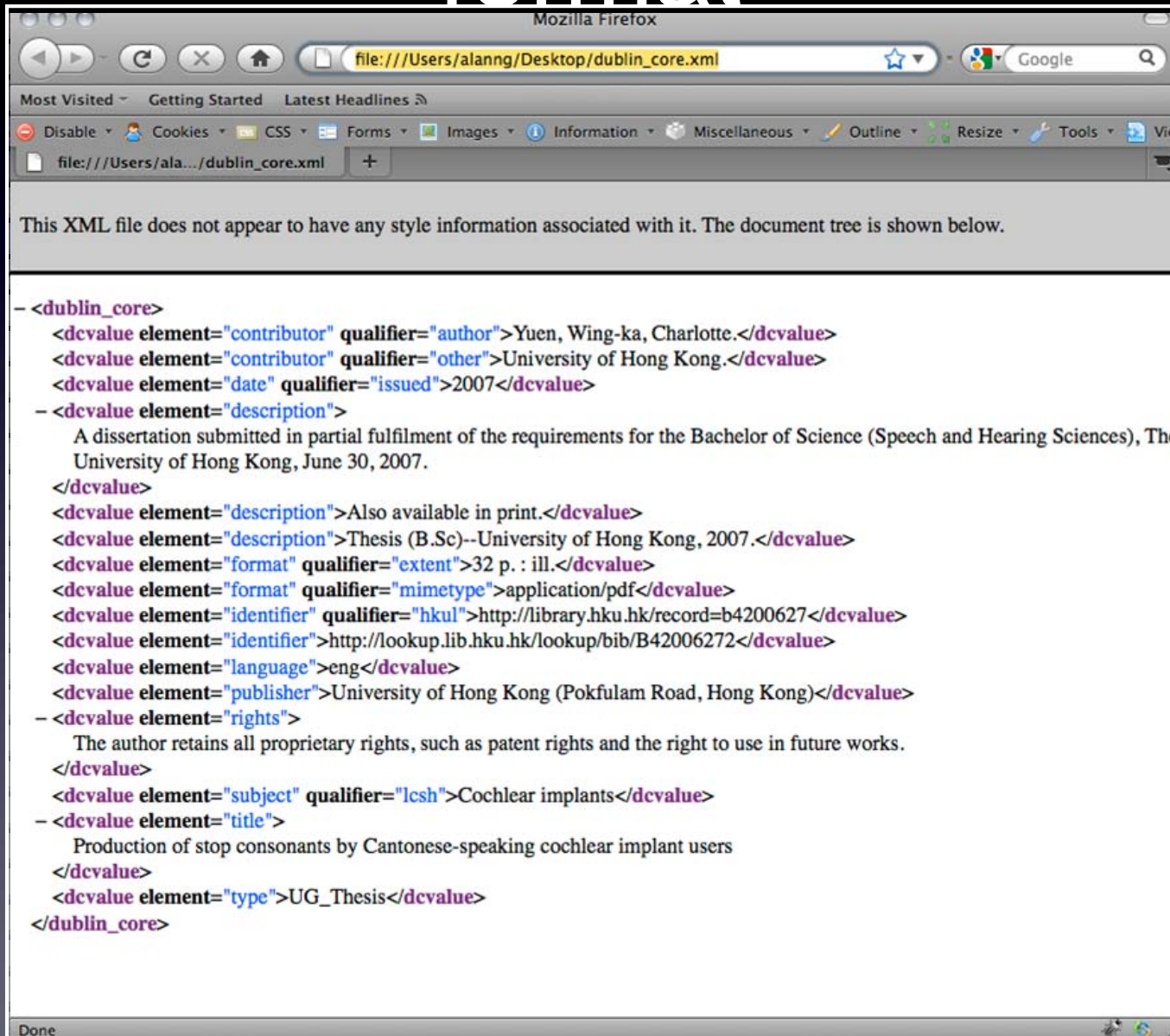
```
</div>
<div align="left"><pre>
LEADER 00000nam 2200000Ia 4500
001 ocn430281940
006 m d
007 cr mn
008 090625s2007 xx a sbm 000 0 eng d
040 HUA|CHUA
066 |c$1
092 HKU Hub
092 B.Sc.(Sp.&Hear.Sc.)07|bY13
100 1 |6880-01|aYuen, Wing-ka, Charlotte.
245 10 Production of stop consonants by Cantonese-speaking
cochlear implant users|h[electronic resource] /|cCharlotte,
Yuen Wing Ka.
260 |c2007.
300 32 p. :|bill.
500 &#34;A dissertation submitted in partial fulfilment of the
requirements for the Bachelor of Science (Speech and
Hearing Sciences), The University of Hong Kong, June 30,
2007.&#34;
502 Thesis (B.Sc)--University of Hong Kong, 2007.
504 Includes bibliographical references (p. 27-30).
530 Also available in print.
538 Mode of access: World Wide Web.
650 0 Cochlear implants
710 2 University of Hong Kong.
773 0 |tHub Collection. BSc (Speech & Hearing) Theses
776 1 Yuen, Wing-ka, Charlotte.|tProduction of stop consonants
by Cantonese-speaking cochlear implant users|cPrint
856 40 |uhttp://lookup.lib.hku.hk/lookup/bib/B42006272|x126
|zClick to view the E-thesis via HKU Scholars Hub
880 1 |6100-01/$1|a袁穎嘉.
993 01 |iiky|econ
995 Speech and Hearing Sciences
996 y|cip|l126|dHKU Hub|eth|go|tt|s6178
</pre></div>
<table width="100%" border="0" cellspacing="1">
</table>
<!-- END BIBSCREEN -->
<br />
<!-- START of footer.html -->

<!-- remove the scope info from the permanent link -->
<script type="text/javascript">
var bibnum = removeBibLinkScope();
</script>
<!-- remove the scope info from the permanent link -->
```

Perl programming

- extracts the essential MARC fields using Regular Expression
- constructs the DC fields according to the mapping table
- converts 100+ records in a couple of minutes

Converted record in DC XML format



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
- <dublin_core>
  <dcvalue element="contributor" qualifier="author">Yuen, Wing-ka, Charlotte.</dcvalue>
  <dcvalue element="contributor" qualifier="other">University of Hong Kong.</dcvalue>
  <dcvalue element="date" qualifier="issued">2007</dcvalue>
  - <dcvalue element="description">
    A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2007.
  </dcvalue>
  <dcvalue element="description">Also available in print.</dcvalue>
  <dcvalue element="description">Thesis (B.Sc)--University of Hong Kong, 2007.</dcvalue>
  <dcvalue element="format" qualifier="extent">32 p. : ill.</dcvalue>
  <dcvalue element="format" qualifier="mimetype">application/pdf</dcvalue>
  <dcvalue element="identifier" qualifier="hku">http://library.hku.hk/record=b4200627</dcvalue>
  <dcvalue element="identifier">http://lookup.lib.hku.hk/lookup/bib/B42006272</dcvalue>
  <dcvalue element="language">eng</dcvalue>
  <dcvalue element="publisher">University of Hong Kong (Pokfulam Road, Hong Kong)</dcvalue>
  - <dcvalue element="rights">
    The author retains all proprietary rights, such as patent rights and the right to use in future works.
  </dcvalue>
  <dcvalue element="subject" qualifier="lch">Cochlear implants</dcvalue>
  - <dcvalue element="title">
    Production of stop consonants by Cantonese-speaking cochlear implant users
  </dcvalue>
  <dcvalue element="type">UG_Thesis</dcvalue>
</dublin_core>
```

Done

Running Perl program

- runs natively on Unix, Linux and Mac OS X
- needs Perl interpreter on Windows
- download ActivePerl
- <http://www.activestate.com/activeperl/>

Running the program on Mac OS X

```
Terminal — telnet — 98x45
-bash-3.2$
-bash-3.2$
-bash-3.2$
-bash-3.2$ ./Bib2DC.pl
This program will convert INNOPAC Bib record into Dublic Core format for DSpace
Usage:
  1) Bib2DC.pl
     - display this help

  2) Bib2DC.pl [bibnum list file] [bib collection type] [content type]
     e.g. Bib2DC.pl sample_list.txt bsc 01

List of bib collection type:
  bsc - BSc Theses
  cupem - CUPEM Student Reports
  ssrc - Social Sciences Research Centre
  chp - Community Health Projects
  hkup - Digital Editions from HKU Press
  gen - general MARC record

List of bib content type:
  01 - PG_Thesis
  02 - PG_ThesisAbstract
  03 - PG_ThesisTOC
  04 - UG_Thesis
  05 - Article
  06 - ConferencePaper
  07 - Text
  08 - Postprint
  09 - Abstract
  10 - MovingImage
  11 - Presentation
  12 - Image
  13 - Patent
  14 - NewspaperArticle
  15 - Book
  16 - Journal
  17 - Preprint
  18 - BookChapter
  19 - LearningObject
  20 - StudentProject
  21 - FindingAid

-bash-3.2$
-bash-3.2$
```


Demo

Recap

Recap

- uses existing MARC records for DSpace
- uses Perl program for fast batch converting
- retrieves MARC in real time via HTTP
- works with any OPAC with persistent URL
- source codes is free for sharing

Q & A

Thank You !!

My contact : lbalan@hku.hk