1-22-2012

Fast, but Accurate? Pitfalls of Batch Metadata Editing

Kathryn Lybarger
University of Kentucky, kathryn.lybarger@uky.edu

Repository Citation

Lybarger, Kathryn, "Fast, but Accurate? Pitfalls of Batch Metadata Editing" (2012). *Library Presentations*. 15.
https://uknowledge.uky.edu/libraries_present/15

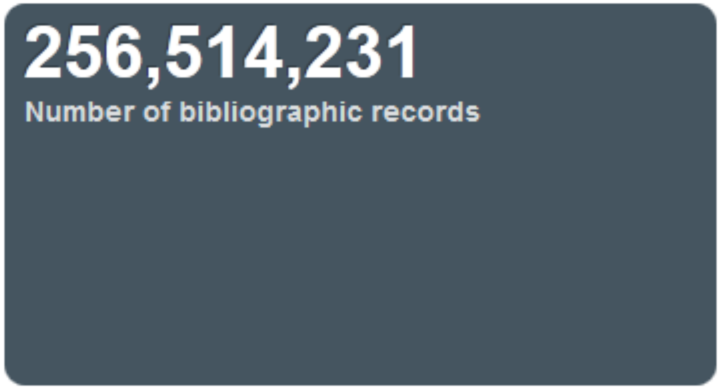
Kathryn Lybarger

ALA Midwinter
Cataloging & Classification Research Interest Group
January 22, 2012

Fast, but Accurate?
Pitfalls of Batch Metadata Editing

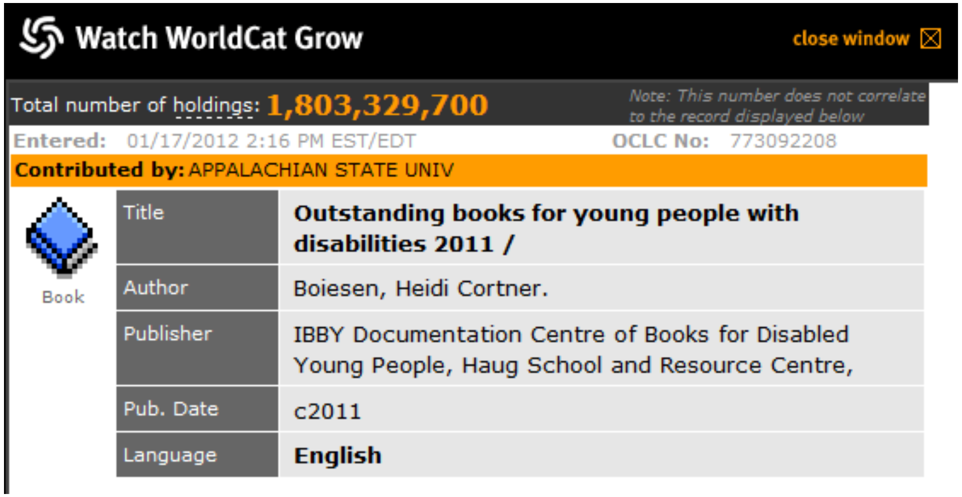
MARC

- A data format used to encode and share bibliographic data
- Developed in the 1960’s, still quite popular



256,514,231

Number of bibliographic records



Watch WorldCat Grow

Total number of holdings: 1,803,329,700

Note: This number does not correlate to the record displayed below

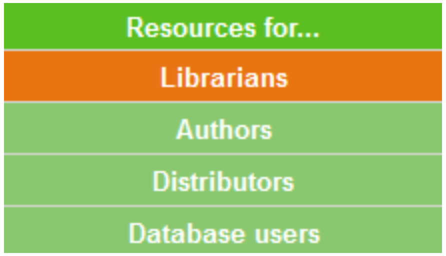
Entered: 01/17/2012 2:16 PM EST/EDT

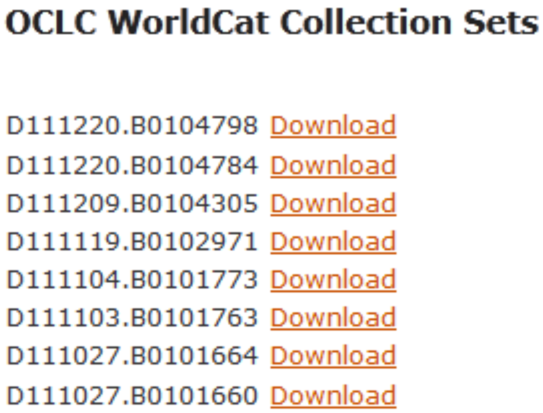
OCLC No: 773092208

Contributed by: APPALACHIAN STATE UNIV

Title	Outstanding books for young people with disabilities 2011 /
Author	Boiesen, Heidi Cortner.
Publisher	IBBY Documentation Centre of Books for Disabled Young People, Haug School and Resource Centre,
Pub. Date	c2011
Language	English

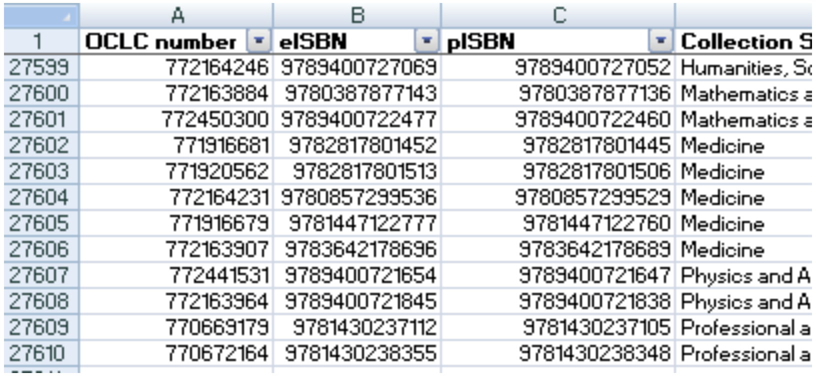
Vendors often provide MARC records



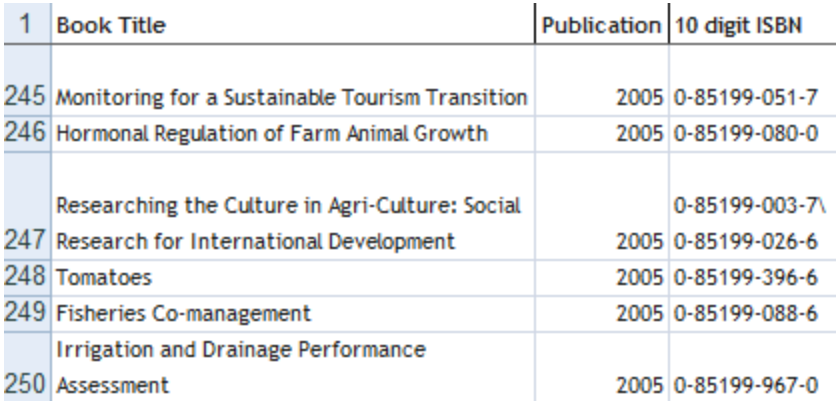


OCLC WorldCat Collection Sets

D111220.B0104798 Download
D111220.B0104784 Download
D111209.B0104305 Download
D111119.B0102971 Download
D111104.B0101773 Download
D111103.B0101763 Download
D111027.B0101664 Download
D111027.B0101660 Download

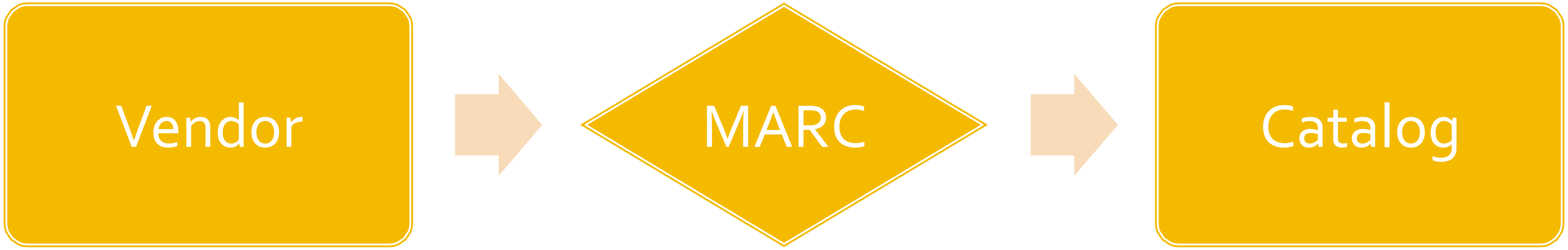


	A	B	C	
1	OCLC number	eISBN	pISBN	Collection S
27599	772164246	9789400727069	9789400727052	Humanities, Sc
27600	772163884	9780387877143	9780387877136	Mathematics a
27601	772450300	9789400722477	9789400722460	Mathematics a
27602	771916681	9782817801452	9782817801445	Medicine
27603	771920562	9782817801513	9782817801506	Medicine
27604	772164231	9780857299536	9780857299529	Medicine
27605	771916679	9781447122777	9781447122760	Medicine
27606	772163907	9783642178696	9783642178689	Medicine
27607	772441531	9789400721654	9789400721647	Physics and A
27608	772163964	9789400721845	9789400721838	Physics and A
27609	770669179	9781430237112	9781430237105	Professional a
27610	770672164	9781430238355	9781430238348	Professional a



1	Book Title	Publication	10 digit ISBN
245	Monitoring for a Sustainable Tourism Transition	2005	0-85199-051-7
246	Hormonal Regulation of Farm Animal Growth	2005	0-85199-080-0
247	Researching the Culture in Agri-Culture: Social Research for International Development	2005	0-85199-003-7\ 0-85199-026-6
248	Tomatoes	2005	0-85199-396-6
249	Fisheries Co-management	2005	0-85199-088-6
250	Irrigation and Drainage Performance Assessment	2005	0-85199-967-0

Batch loading


Vendor

MARC

Catalog

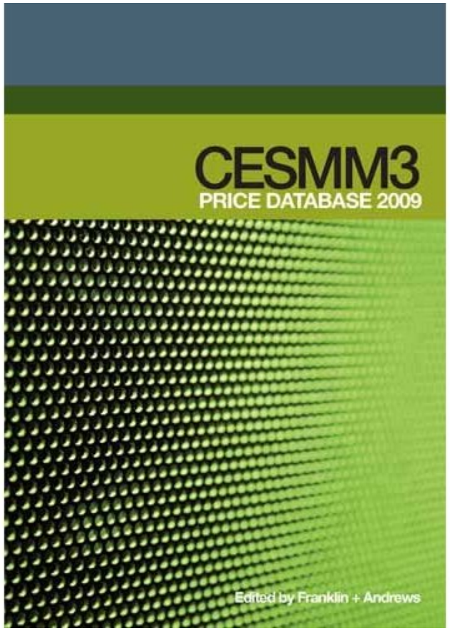
All done?

Not quite…

Records may be icky…

Title: CESMM3 price database 2009, edited by Franklin + Andrews

```
100 1_  Franklin.
245 10  CESMM3 price
database 2009 ‡h
[electronic resource] / ‡c
edited by Franklin and
Andrews.
500 __  Ebook.
516 __  Document.
538 __  PDF: Adobe PDF
700 1_  Andrews.
856 40  …
```

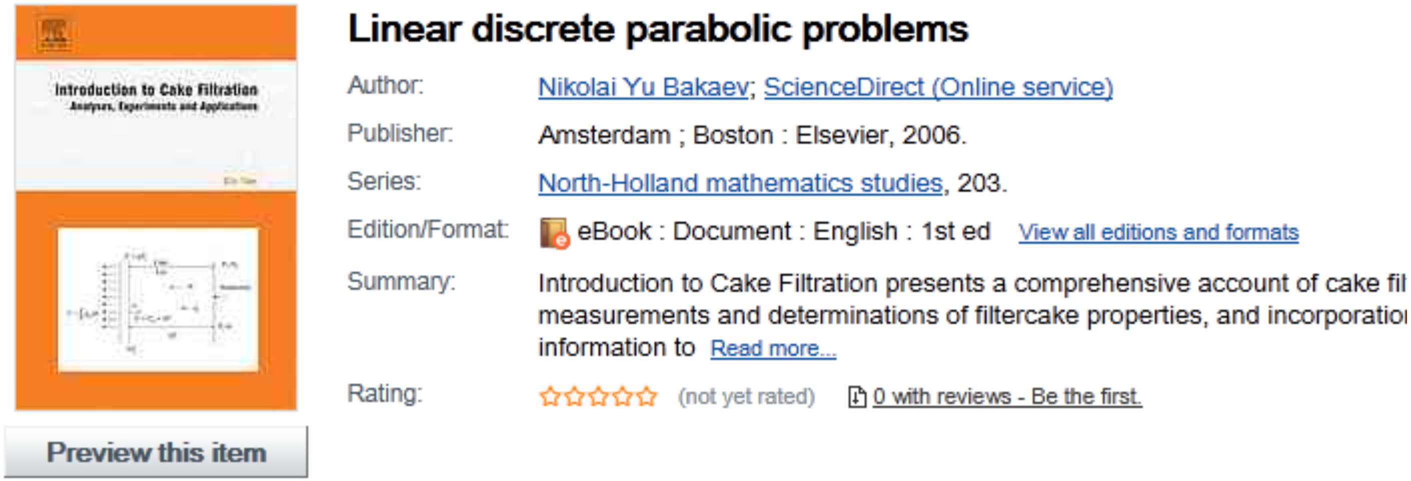


…but worse, non-functional!

- Data may be unhelpful, or misleading
- Links may not work

- This may change over time

A crazy mixed-up record (with 76 holdings)


Introduction to Cake Filtration
Analyses, Experiments and Applications
Preview this item

Linear discrete parabolic problems

Author:	Nikolai Yu Bakaev; ScienceDirect (Online service)
Publisher:	Amsterdam ; Boston : Elsevier, 2006.
Series:	North-Holland mathematics studies, 203.
Edition/Format:	eBook : Document : English : 1st ed View all editions and formats
Summary:	Introduction to Cake Filtration presents a comprehensive account of cake fil measurements and determinations of filtercake properties, and incorporatio information to Read more...
Rating:	☆☆☆☆☆ (not yet rated) 0 with reviews - Be the first.

- **From one book:**
  - Title
  - Author
  - Series
  - Subject headings
- **From another book:**
  - Notes
  - ISBN
  - Link to e-book

“Local” data (not local to you)

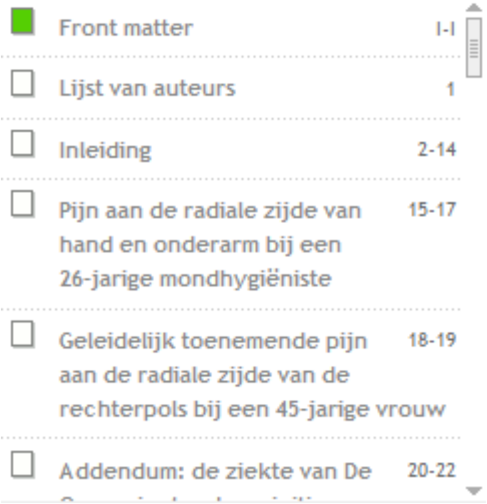
- Notes or link text:
  - “Restricted to <Not Your Institution>”

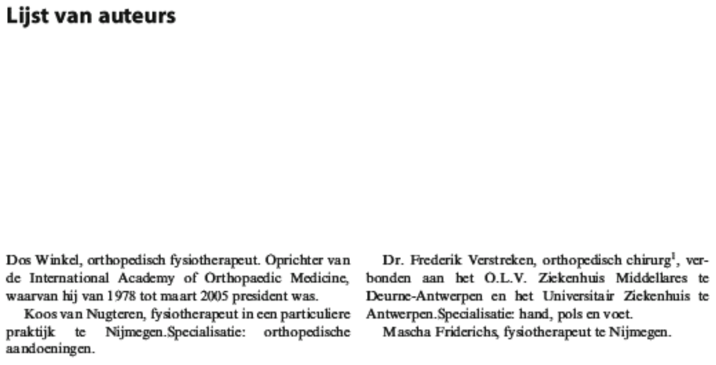
- Proxy prefixes from other locations
  - http://ezproxy.uky.edu/login?url=http://www…
- URLs that restrict access
  - http://www.uky.edu.ebook-vendor.com/…

URLs from other vendors

- Provider-neutral records may have URLs from multiple vendors
- An OCLC search for records with URLs from eblib, ebrary, ebscohost AND myilibrary returned over 25,000.
- Even if they are labeled, your patrons don't know which vendor you're using

Valid ebook … just not for you!





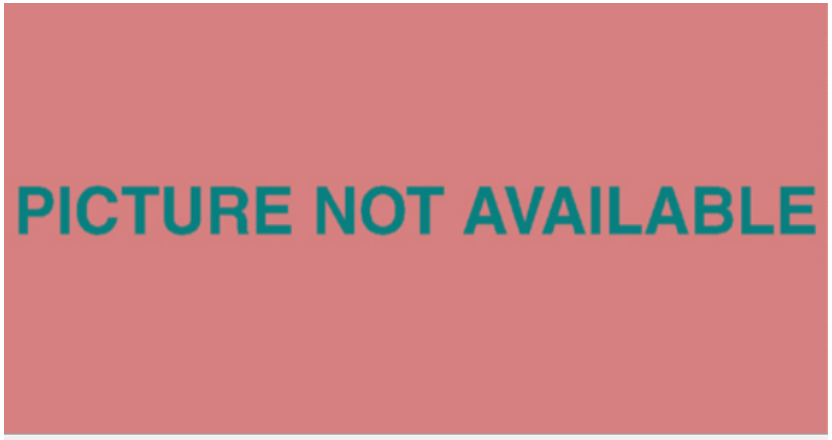
Lijst van auteurs

Dos Winkel, orthopedisch fysiotherapeut. Oprichter van de International Academy of Orthopaedic Medicine, waarvan hij van 1978 tot maart 2005 president was.

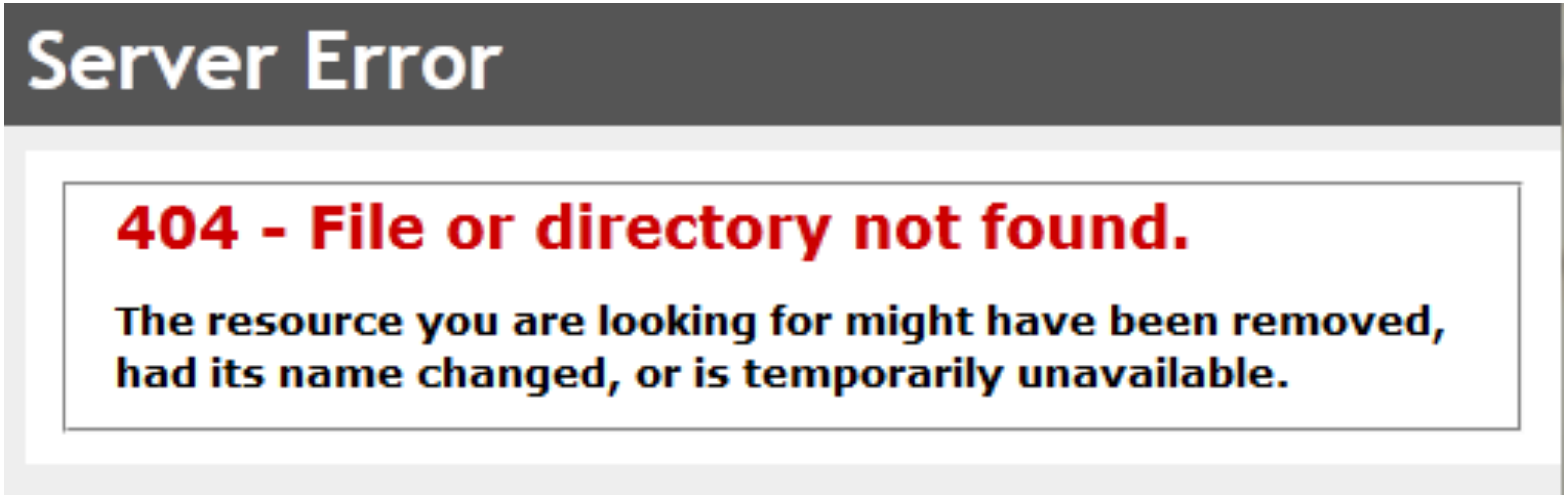
Koos van Nugteren, fysiotherapeut in een particuliere praktijk te Nijmegen.Specialisatie: orthopedische aandoeningen.

Dr. Frederik Verstreken, orthopedisch chirurg[1], verbonden aan het O.L.V. Ziekenhuis Middellares te Deurne-Antwerpen en het Universitair Ziekenhuis te Antwerpen.Specialisatie: hand, pols en voet.

Mascha Friderichs, fysiotherapeut te Nijmegen.



URLs that point nowhere

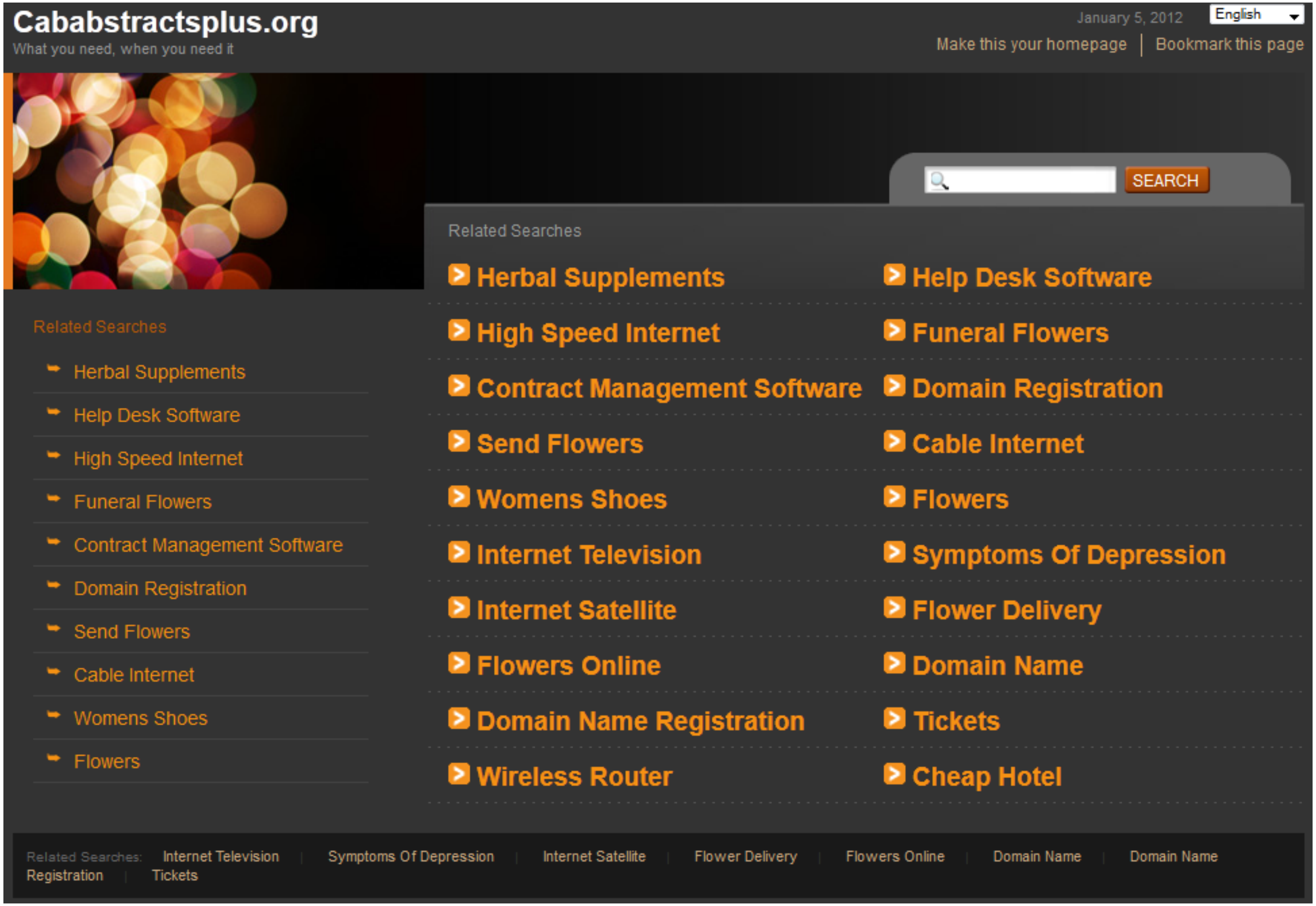


Server Error

404 - File or directory not found.

The resource you are looking for might have been removed, had its name changed, or is temporarily unavailable.

URLs that point somewhere new!



Cababstractsplus.org
What you need, when you need it

January 5, 2012 English

Make this your homepage | Bookmark this page

SEARCH

Related Searches

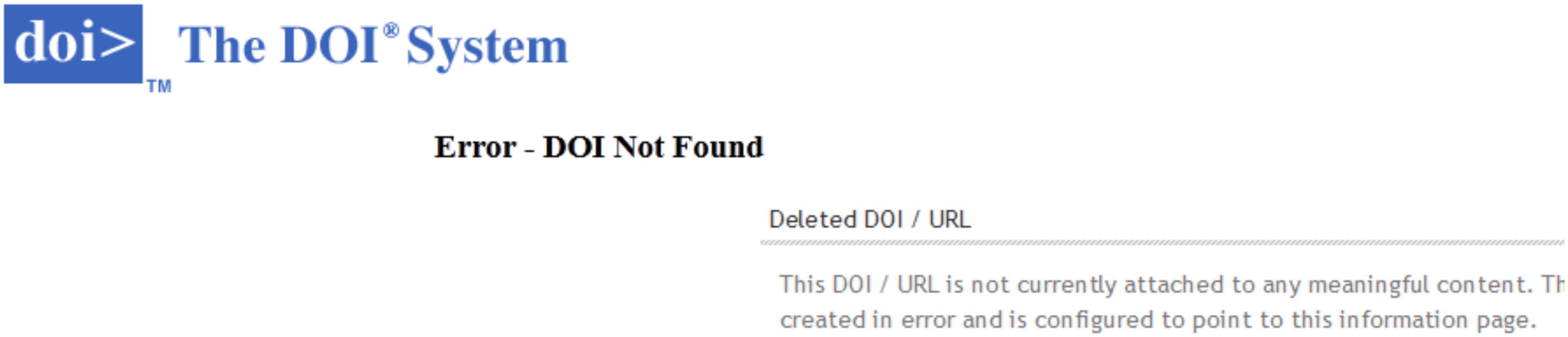
- Herbal Supplements
- Help Desk Software
- High Speed Internet
- Funeral Flowers
- Contract Management Software
- Domain Registration
- Send Flowers
- Cable Internet
- Womens Shoes
- Flowers

Related Searches

- Herbal Supplements
- High Speed Internet
- Contract Management Software
- Send Flowers
- Womens Shoes
- Internet Television
- Internet Satellite
- Flowers Online
- Domain Name Registration
- Wireless Router
- Help Desk Software
- Funeral Flowers
- Domain Registration
- Cable Internet
- Flowers
- Symptoms Of Depression
- Flower Delivery
- Domain Name
- Tickets
- Cheap Hotel

Related Searches: Internet Television | Symptoms Of Depression | Internet Satellite | Flower Delivery | Flowers Online | Domain Name | Domain Name Registration | Tickets

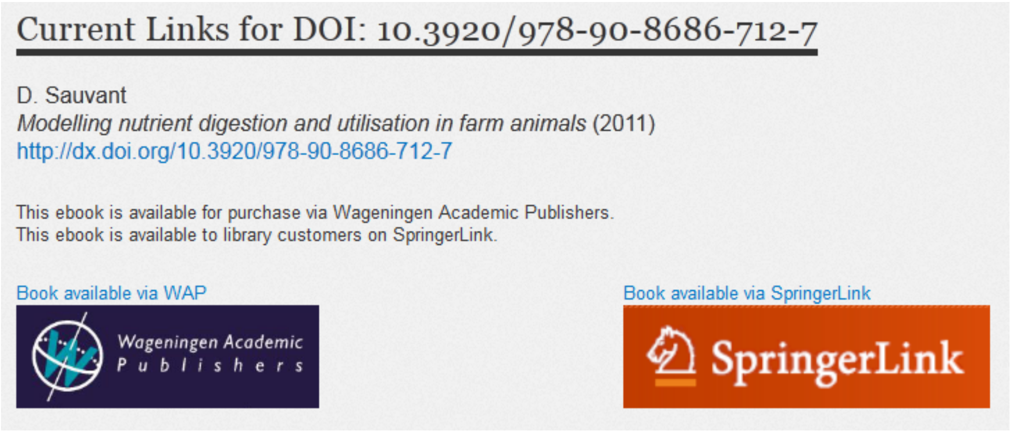
DOI troubles



Error - DOI Not Found

Deleted DOI / URL

This DOI / URL is not currently attached to any meaningful content. Th
created in error and is configured to point to this information page.



Current Links for DOI: 10.3920/978-90-8686-712-7

D. Sauvant
Modelling nutrient digestion and utilisation in farm animals (2011)
http://dx.doi.org/10.3920/978-90-8686-712-7

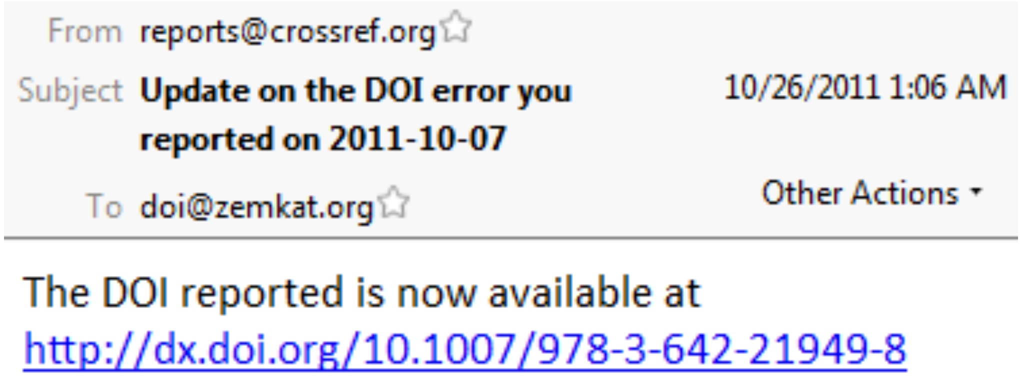
This ebook is available for purchase via Wageningen Academic Publishers.
This ebook is available to library customers on SpringerLink.

Book available via WAP
Wageningen Academic Publishers
Book available via SpringerLink
SpringerLink

Some DOI troubles can be fixed

- Bonus: When they fix it for you, it is fixed for everyone!

- In the meantime, you can use the direct link.



From reports@crossref.org

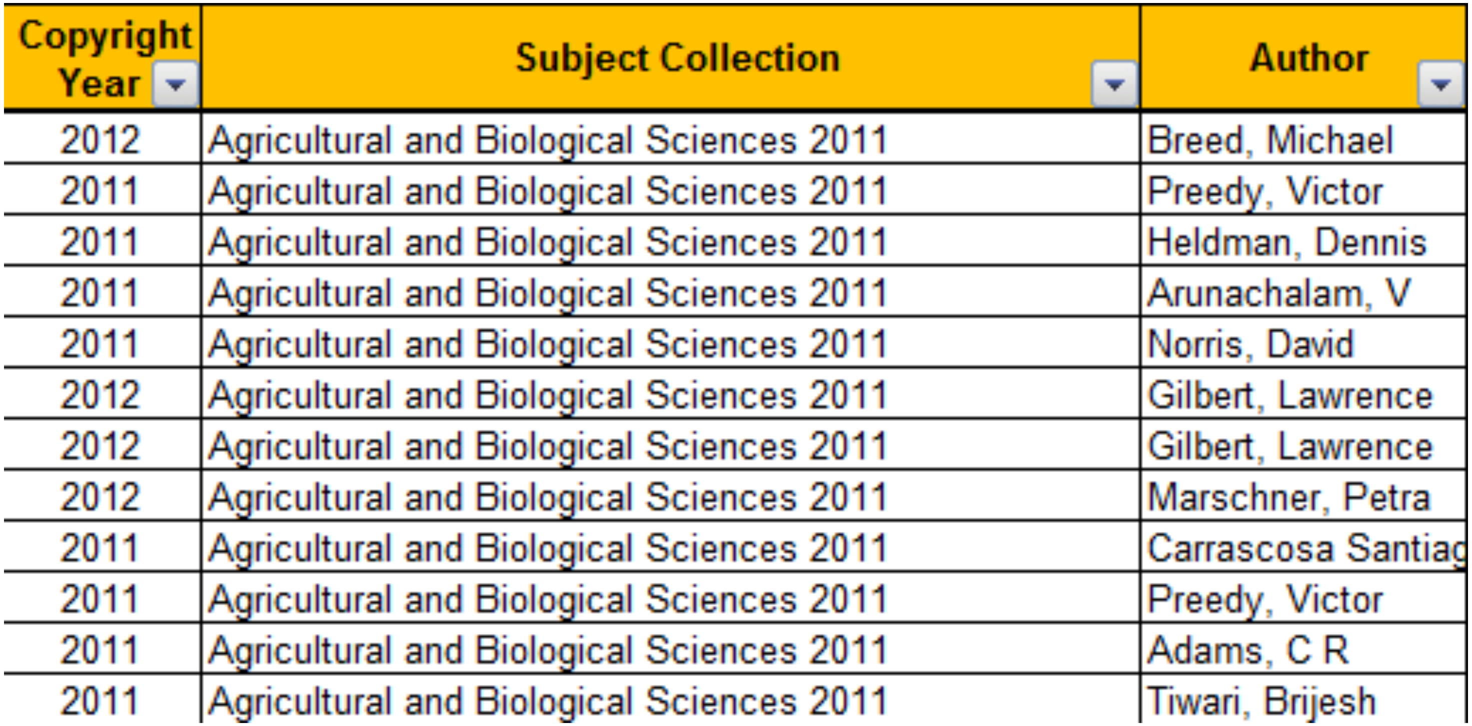
Subject **Update on the DOI error you reported on 2011-10-07** 10/26/2011 1:06 AM

To doi@zemkat.org

The DOI reported is now available at http://dx.doi.org/10.1007/978-3-642-21949-8

- … unless the book is not actually there.

Books may not be available yet (or ever)



Copyright Year	Subject Collection	Author
2012	Agricultural and Biological Sciences 2011	Breed, Michael
2011	Agricultural and Biological Sciences 2011	Preedy, Victor
2011	Agricultural and Biological Sciences 2011	Heldman, Dennis
2011	Agricultural and Biological Sciences 2011	Arunachalam, V
2011	Agricultural and Biological Sciences 2011	Norris, David
2012	Agricultural and Biological Sciences 2011	Gilbert, Lawrence
2012	Agricultural and Biological Sciences 2011	Gilbert, Lawrence
2012	Agricultural and Biological Sciences 2011	Marschner, Petra
2011	Agricultural and Biological Sciences 2011	Carrascosa Santiag
2011	Agricultural and Biological Sciences 2011	Preedy, Victor
2011	Agricultural and Biological Sciences 2011	Adams, C R
2011	Agricultural and Biological Sciences 2011	Tiwari, Brijesh

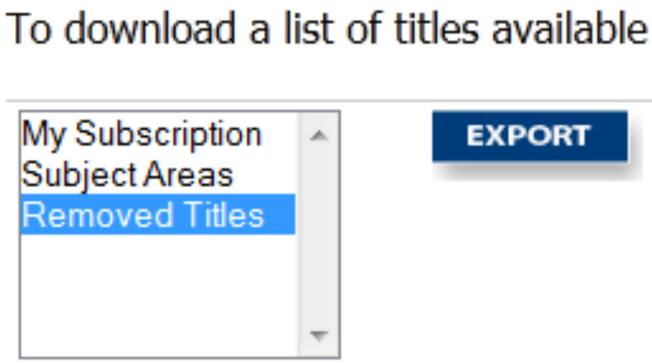
“Slippage”

- Some ebooks on a frontlist may never appear on the site
- Individual ebooks may just disappear

Lists may be available…

- But not forthcoming.

- You may have to periodically dig several levels deep on the website to get them:

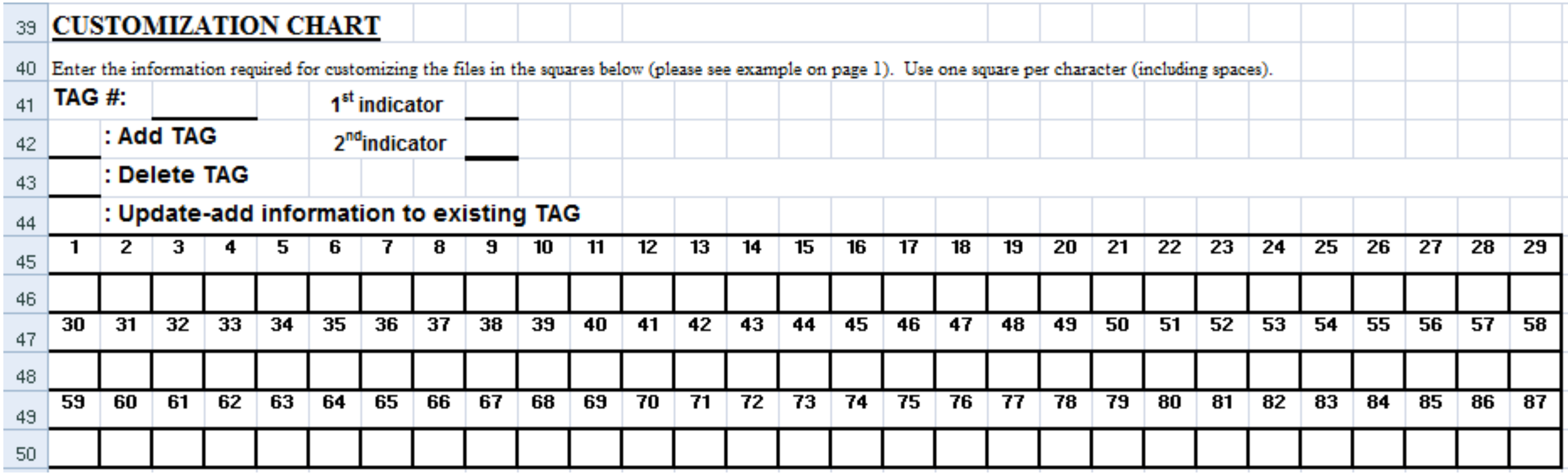


To download a list of titles available on

Solutions?

- Use provider-neutral records when you can
- Edit MARC records to conform with local standards
- Verify access to all titles (periodically)
- Communicate with other catalogers

Vendors may do some editing



CUSTOMIZATION CHART

Enter the information required for customizing the files in the squares below (please see example on page 1). Use one square per character (including spaces).

TAG #: ______ 1st indicator ____

____ : Add TAG 2nd indicator ____

____ : Delete TAG

____ : Update-add information to existing TAG

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58
59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87

- But how do you predict what you will need?

MarcEdit

- Developed by Terry Reese at Oregon State
- MARC editing in a friendly yet powerful text editor
- Z39.50 client
- (Binary editor!)

Version control

- Maintain previous versions of files efficiently
  - No need for `fileFeb12-FINAL6.mrk.bak`
  - Undo to any previous version
- Mercurial (Hg):
  - Free, lightweight, cross-platform
  - Easy to set up and remove repositories

- Command line, GUI (TortoiseHG, SourceTree)

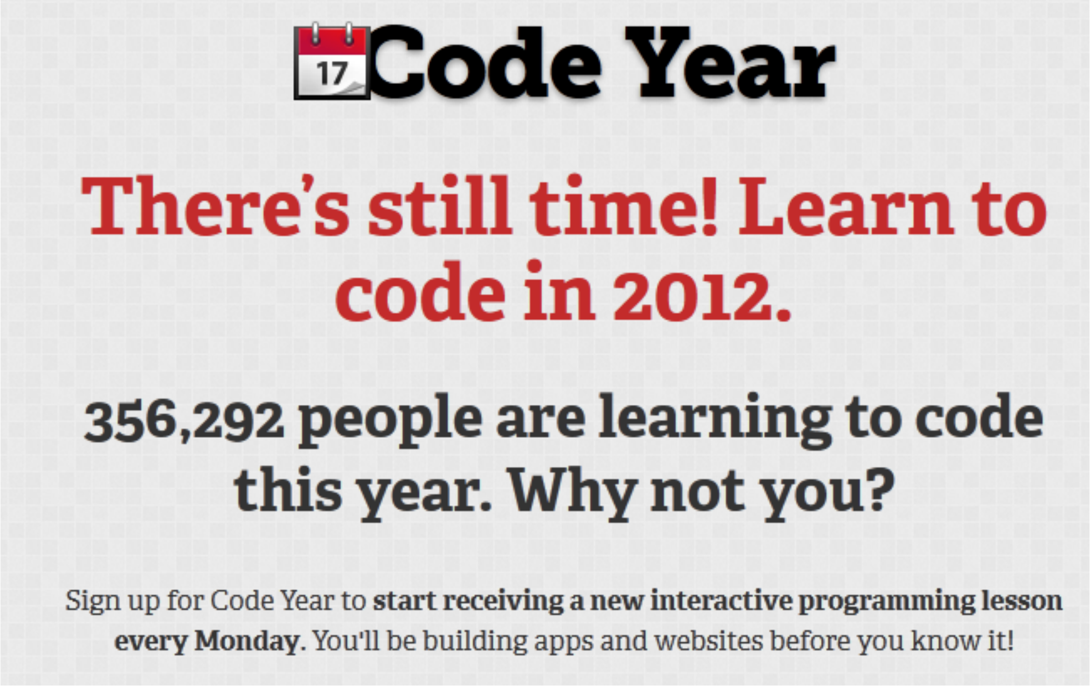
Automation

- MarcEdit Macros
  - Visual Basic, Visual Basic.NET

- .mrk format is text, so you can process with your favorite programming language
- Don’t have a favorite language (yet)?

#catcode #libcodeyear

- From **CodeAcademy.com**:



17 **Code Year**

There's still time! Learn to code in 2012.

356,292 people are learning to code this year. Why not you?

Sign up for Code Year to **start receiving a new interactive programming lesson every Monday**. You'll be building apps and websites before you know it!

Text processing tools

Cygwin (unix) tools: grep, vim, vimdiff, sort, wc (and the list goes on)

```
grep ^=856 ebooks.mrk
```

```
=856  40$u http://dx.doi.org/10.1007/978-1-4419-9934-4
=856  40$u http://dx.doi.org/10.1007/978-1-4302-3513-2
=856  40$u http://public.eblib.com/EBLPublic/PublicVie...
=856  40$u http://dx.doi.org/10.1007/978-0-85729-661-0
=856  40$u http://dx.doi.org/10.1007/978-3-8349-6217-1
```

My automation (bash, PHP, mysql)

- `new_ebsco.sh`
- Profile for each vendor answers:
  - What lines should I add/delete?
  - What does a valid URL look like?
  - How can I tell if the ebook is live?
- (Check logs for problems)
- `pull.sh <filename>`

Generic link checkers may not be effective

- Ebook errors can be valid web pages, and errors don’t mean you should give up!

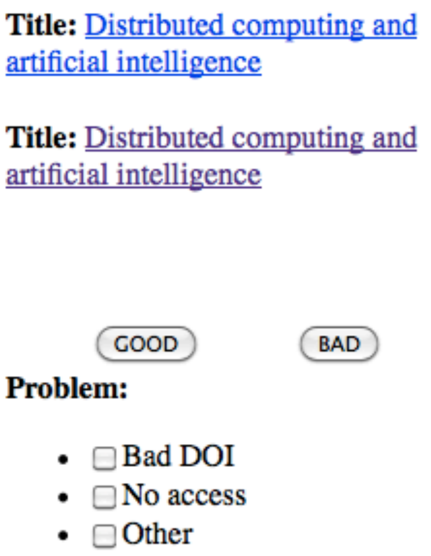
- **`HTTP/1.1 200 OK`**
  - Full text ebook
  - Web site form to buy the book

- **`HTTP/1.1 404 Not Found`**
  - No such page on server
  - Broken DOI (that you should report)

Effective link checking (my method)

- Database holds a list of links to be checked
- Script checks each according to site profile (pausing 10 seconds between each link):
  - Is it a PDF?
  - Does it contain the phrase "This is not part of your subscription"?
  - Can you click through to fulltext chapters?

More thorough link checking

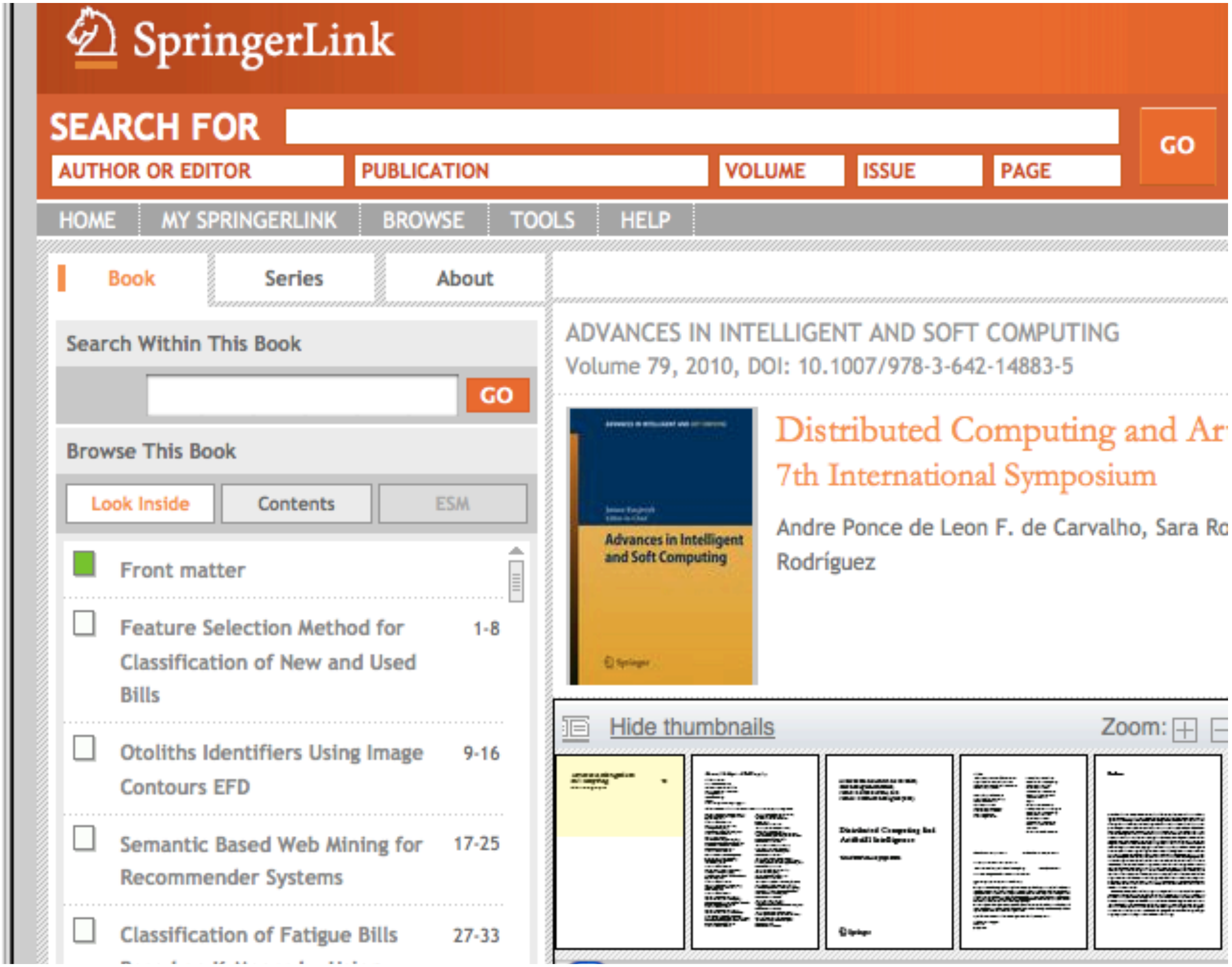


Title: Distributed computing and artificial intelligence

Title: Distributed computing and artificial intelligence

Problem:

- □ Bad DOI
- □ No access
- □ Other



🟩 Front matter

□ Feature Selection Method for Classification of New and Used Bills 1-8

□ Otoliths Identifiers Using Image Contours EFD 9-16

□ Semantic Based Web Mining for Recommender Systems 17-25

□ Classification of Fatigue Bills 27-33

ADVANCES IN INTELLIGENT AND SOFT COMPUTING

Volume 79, 2010, DOI: 10.1007/978-3-642-14883-5

Distributed Computing and Ar

7th International Symposium

Andre Ponce de Leon F. de Carvalho, Sara Ro Rodríguez

Communicate

- Dead links are in catalogs everywhere … how to let people know?
- Let vendors know if you find them in the wild!
- A blog / database for “zombie e-books” ?

Any questions?

Links

- MarcEdit http://people.oregonstate.edu/~reeset/marcedit/html/index.php
- Mercurial http://mercurial.selenic.com/
- Code Academy http://www.codeacademy.com
- Cygwin http://www.cygwin.com