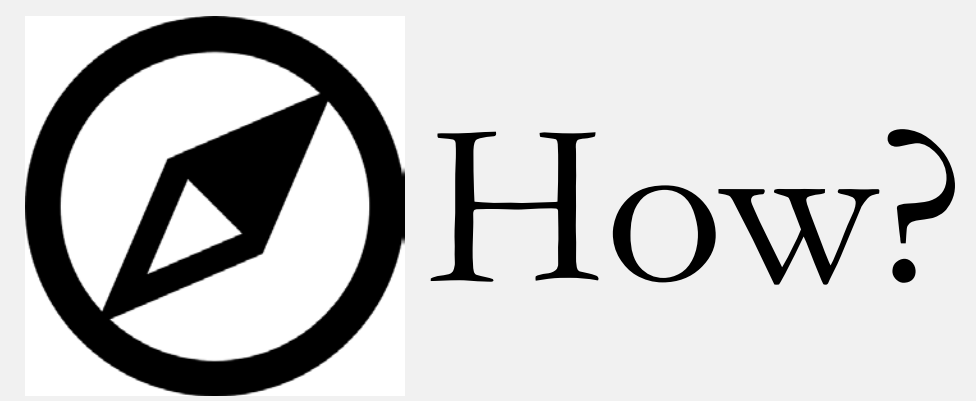


Creating Macros to Speed Up CJK Collection Development

A problem to solve

I spent too much time preparing lists of Japanese materials. My acquisitions colleagues can't read Japanese and need Romaji – I was essentially a professional copy and paster.

I decided inventing a macro to do this for me would be far more efficient.



With 5 different programs:

- Google Chrome
- Adblock Plus Plugin
- Pulover Macro Creator
- Auto Hot Key (AHK)
- Excel (and Visual Basic)



Time
 Manually: 6 min/book
 Macro: 40 seconds/book

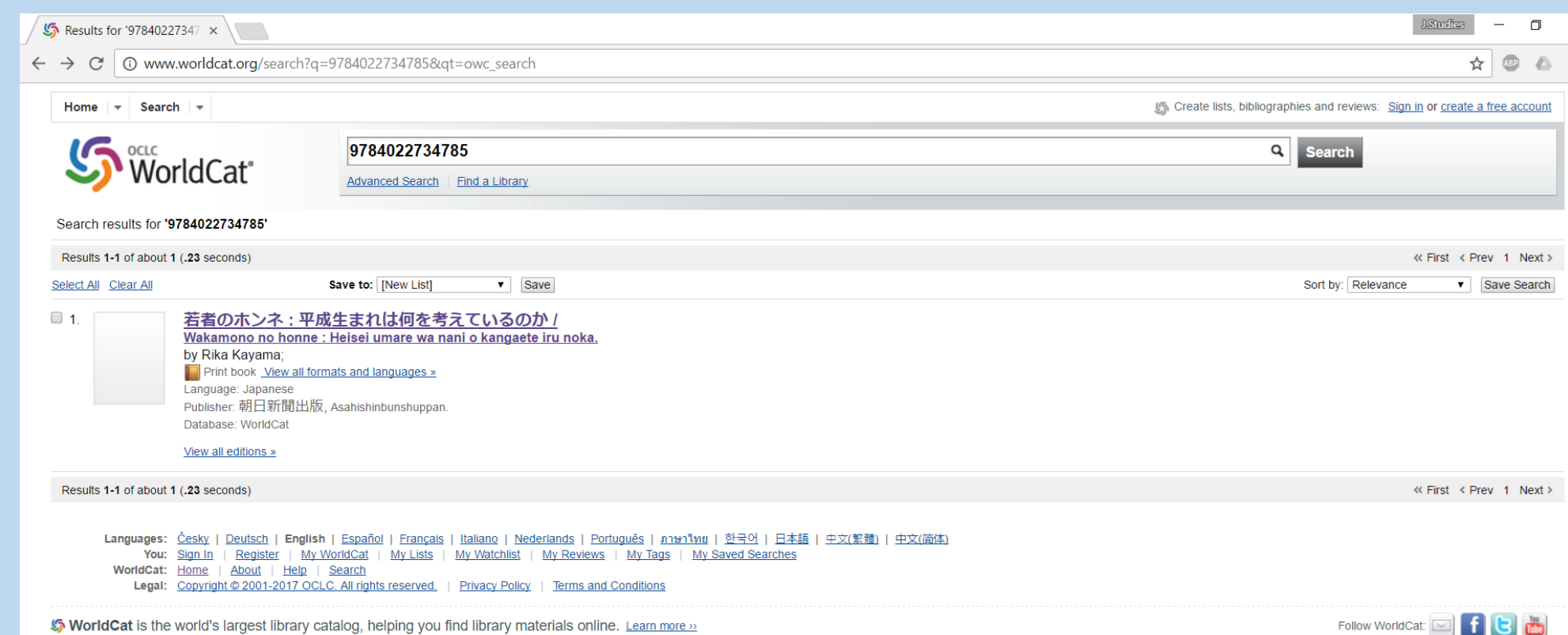
Health
 Numbness in hand from copy/paste

Accuracy
 Removes human error, auto fixes common problems

Bilingual Data in WorldCat Source Code

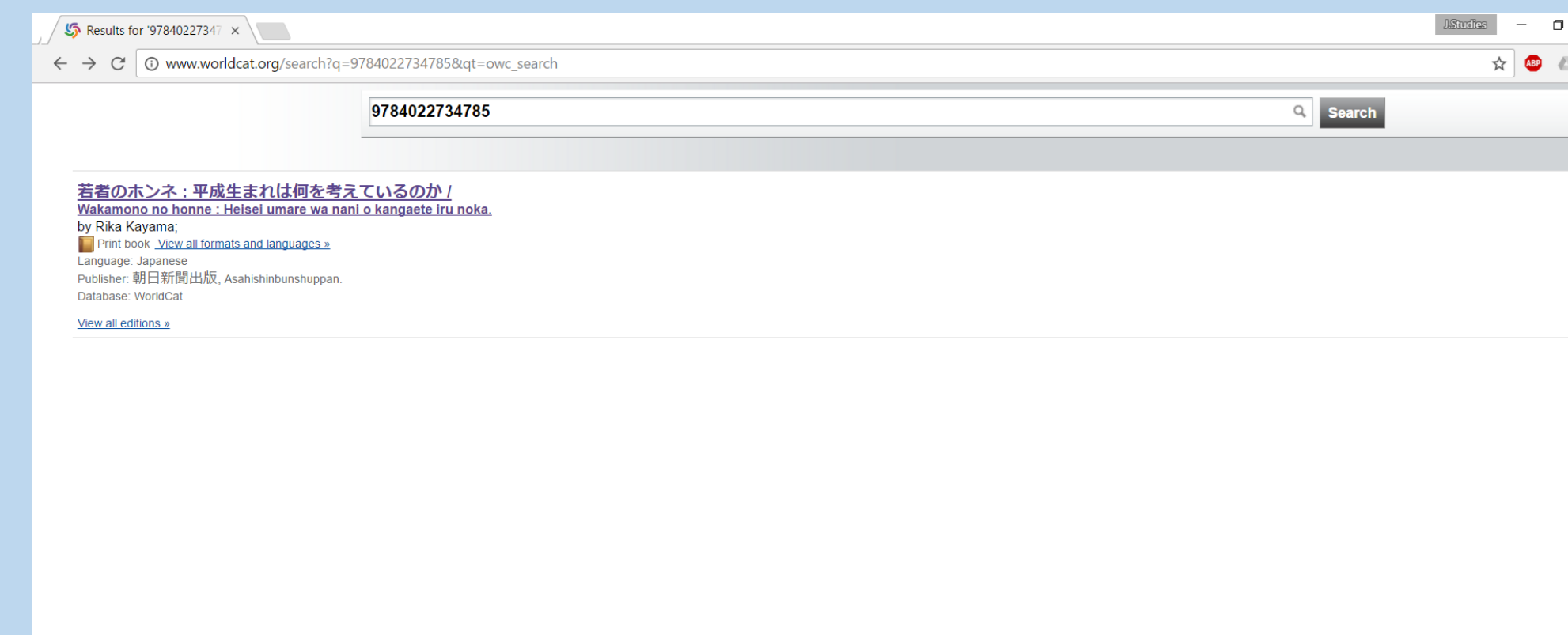
Making WorldCat macro friendly

Regular WorldCat



Variable contextual elements make data inconsistent for macros

Adblock suppressing page elements



No Contextual page elements = Identical set of key strokes to access data

A unique Chrome profile for collection development helps customize the browser for a CD workflow.

Identifying unique strings of data

Unique text string Vernacular Romanized

```
<td id="bib-publisher-cell"><span class=vernacular lang="ja">朝日新聞出版,</span> Tōkyō-to Chūō-ku : Asahishinbunshuppan, 2015.</td>
```

not ALA-LC compliant
 can be fixed in Excel automatically

ISBNs, OCLC#, Author data, etc – all have unique text strings in the source code, a ctrl+f search locates them quickly.

Text strings copied / pasted into Excel data columns

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	!!! Macro's in this spreadsheet only clean data up to row 250.												
2	Rush	Inform Faculty	Collection Location	ISBN #	OCLC #	Author (Eng)	Author (Jpn)	Title (Eng)	Title (Jpn)	Series	Publisher (Eng)	Publisher (Jpn)	PubDate
3				4000222236	793575339	Norihisa Suzuki	鈴木範久	Uchimura Kanzō no	内村鑑三の人と思想		Iwanami Shoten	岩波書店	2012

Pulover Macro Creator used to make an AutoHotKey macro to move data from Chrome to Excel

Additional macros in Excel clean text strings

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	!!! Macro's in this spreadsheet only clean data up to row 250.												
2	Rush	Inform Faculty	Collection Location	ISBN #	OCLC #	Author (Eng)	Author (Jpn)	Title (Eng)	Title (Jpn)	Series	Publisher (Eng)	Publisher (Jpn)	PubDate
3				4000222236	793575339	Norihisa Suzuki	鈴木範久	Uchimura Kanzō no	内村鑑三の人と思想		Iwanami Shoten	岩波書店	2012

Through a series of 'find and replace' commands just the human readable data is left.

Challenges

Diacritics

Composed vs. Decomposed

ō vs. ō

They look identical to us, but the computer sees:

ChrW(333)

vs .

ChrW(175) & ChrW(111)

Very New Items

New books often don't appear in WorldCat. But a second macro can be designed to search Japanese websites.

Opportunities

Not Limited to CJK

This macro should work with any non roman script language in world cat, but further testing is needed to see if it's robust enough.

What Next?

Implementation

How to make this macro easier for others to adopt. If more people are able to use it than it will be possible to figure out problems for efficiently.