

Using AI to format Alma item descriptions

WRLC's First Artificial Intelligence Project



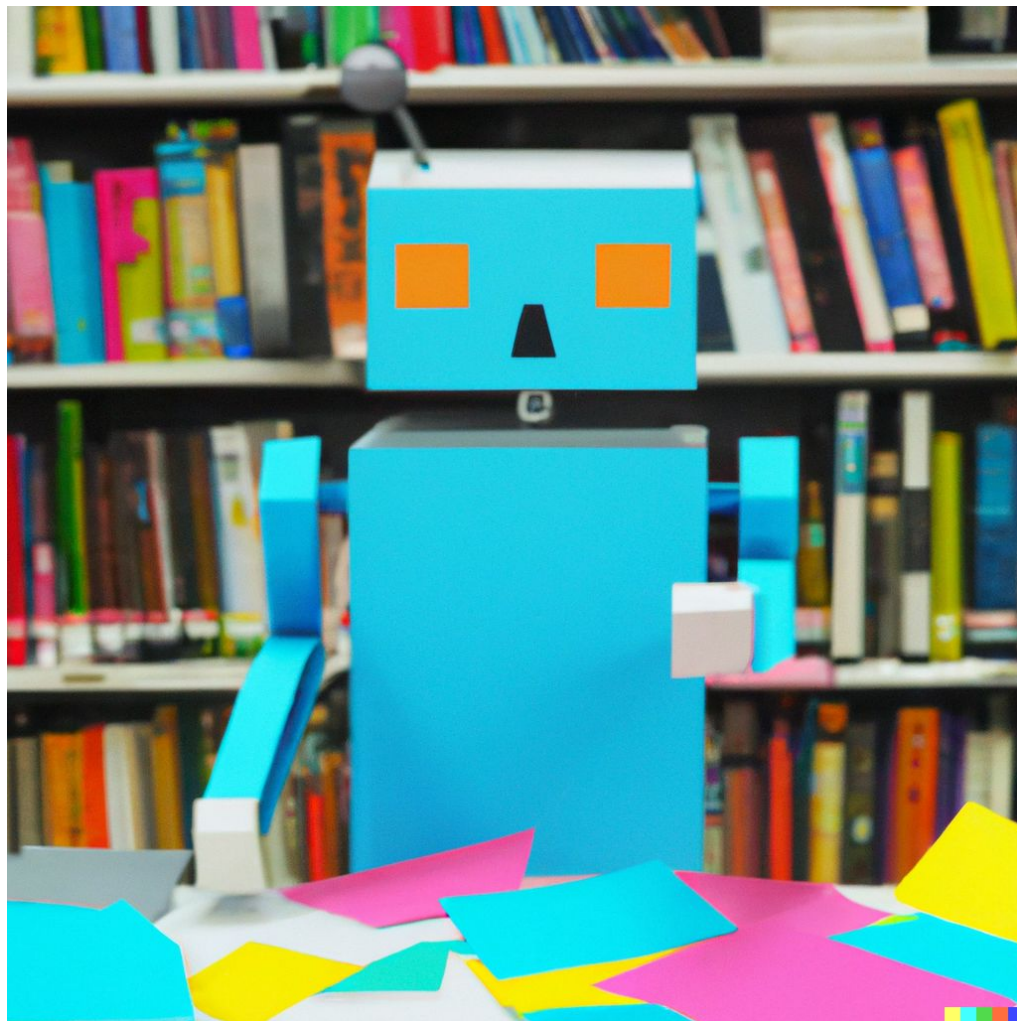


A Holy Grail of Shared Print

- Other tools like OpenRefine might work, but...
- AI is teachable and iterative
- This is the only way we'll all eventually get on the same page for shared print periodicals
- It's a project that can be done centrally to benefit all partners

Getting Started

Last year WRLC's recommendations for print periodicals were written by the Serial Holdings Subgroup and approved by the WRLC Metadata Committee.



The Proof of Concept

- Slowly refined a query with ChatGPT
- [Item description templates](#) were very helpful
- Great opportunity to begin learning to use AI
- ChatGPT free tier allows 100 lines of input which is plenty for testing a query statement





[Guide to style: All description statements should begin with volume information if available like "v.101:". Then there should be number information if available like "no.3,". Then there should be issue information if available like "iss.7 ". Date information if available should be enclosed in parentheses like "(2022: Mar. 21)". Statements starting with a 4 digit date should be rewritten as "v.2 (1998)". There should not be more than one hyphen in a statement. So "v.33-36: (1976-1979)" should be reprinted as "v.33 (1976) - v.36 (1979)".] Reprint the following strings according to the style

v.20:1990

v.1-3 1978-1980

v.4-6 1981-1983

v.7-10 1984-1987

v.44-46 1987-1989

v.29 (1972)-v.32 (1975)

v.33-36 1976-1979

v.41-43 1984-1986

v. 53 1986

v.5 1981-1982

v.91:no.4-6 1983:Aug.-Dec.

v.44 1998

v.23 (1997)

1998:v.1

1998:v.2

1998:v.3



[Guide to style] **1** All description statements should begin with volume information if available like "v.101!" **2** Then there should be number information if available like "no.3." **3** Then there should be issue information if available like "iss.7" **4** Date information if available should be enclosed in parentheses like "(2022: Mar. 21)". Statements starting with a 4 digit date should

be rewritten as "v.2 (1998)". There should not be more than one hyphen in a statement. So "v.33-36. (1976-1979)" should be reprinted as "v.33 (1976) -v.36 (1979)".] Reprint the following strings according to the style

v.20:1990

v.1-3 1978-1980

v.4-6 1981-1983

v.7-10 1984-1987

v.44-46 1987-1989

v.29 (1972)-v.32 (1975)

v.33-36 1976-1979

v.41-43 1984-1986

v. 53 1986

v.5 1981-1982

v.91:no.4-6 1983:Aug.-Dec.

v.44 1998

v.23 (1997)

1998:v.1

1998:v.2

1998:v.3

ADDITIONAL CORRECTIONS AS NEEDED



[Guide to style: All description statements should begin with volume information if available like "v.101:". Then there should be number information if available like "no.3,". Then there should be issue information if available like "iss.7 ". Date information if available should be enclosed in parentheses like "(2022: Mar. 21)". Statements starting with a 4 digit date should be rewritten as "v.2 (1998)". There should not be more than one hyphen in a statement. So "v.33-36: (1976-1979)" should be reprinted as "v.33 (1976) - v.36 (1979)".] Reprint all strings in the description column of the attached csv and write to a new column



find all volume statements.csv



Finished working

Show work: ▾

The CSV file contains two columns: "Barcode" and "Description". The "Description" column appears to contain various forms of volume, issue, and year information. Let's proceed with reformatting these strings according to the rules you've provided.

Here are the rules summarized:

1. All description statements should begin with volume information if available (e.g., "v.101:").
2. Then there should be number information if available (e.g., "no.3").

2. There should be number information if available (e.g., "no. 7").
3. Then there should be issue information if available (e.g., "iss.7").
4. Date information should be enclosed in parentheses (e.g., "(2022: Mar. 21)").
5. Statements starting with a 4-digit date should be rewritten (e.g., "1998 v.2" becomes "v.2 (1998)").
6. There should not be more than one hyphen in a statement. So "v.33-36: (1976-1979)" should be reprinted as "v.33 (1976) - v.36 (1979)".

Please note that without explicit examples for each rule, some assumptions will be made in the implementation. If the results are not as expected, please provide examples for clarification. Let's start with the reformatting process.

Finished working

Show work 

The descriptions have been reformatted according to the rules provided and written into a new column named "New Description". Here's how the first few rows of the dataframe look after reformatting:

	Barcode	Description	New Description
0	21194001655113X	1950 v.12	v.12 (1950)
1	31194000000137X	v.4	v.4
2	31194000000822X	v.19 1977	v.19 (1977)

What it can fix...

Original Item Description	Reformatted Item Description
v.30 apr.-june 1994 pp.663-1350	v.30 (1994: Apr.-June) pp.663-1350
v.30, no.1 suppl. 1940	v.30: no.1 (1940) suppl.
Index v. 72 Jan.-June 1998	v.72 (1998: Jan.-June) Index

and what it can't.

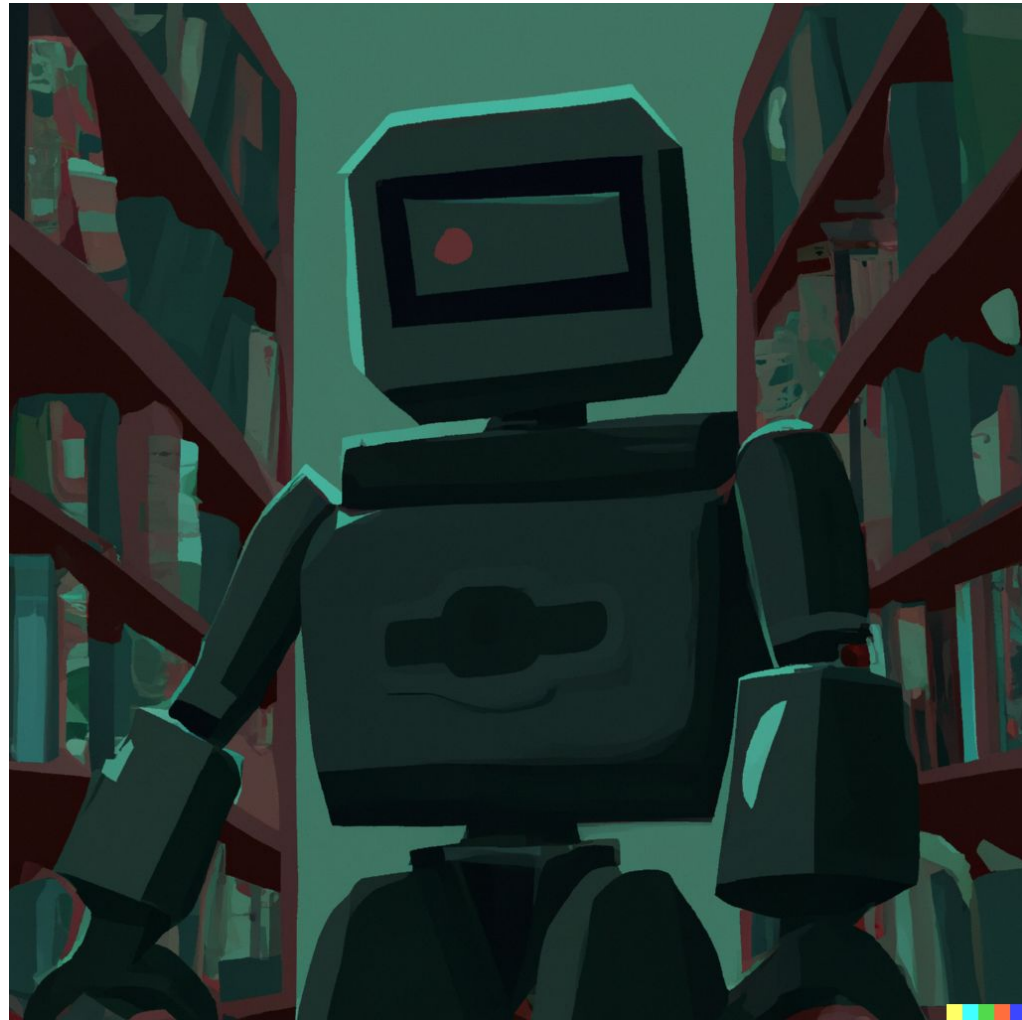
Original Item Description	Reformatted Item Description
v.39, pp. 1623-3606 Oct-Dec 1963	v.39 (1623-3606: Oct.-Dec.) pp. 1623-3606
January 1958 - November 1964 January 1958 - November 1964	(1958: Jan.-Jan.-Nov.)
v.28, nos. 1+12 1981	v.28: no.1 (1981)

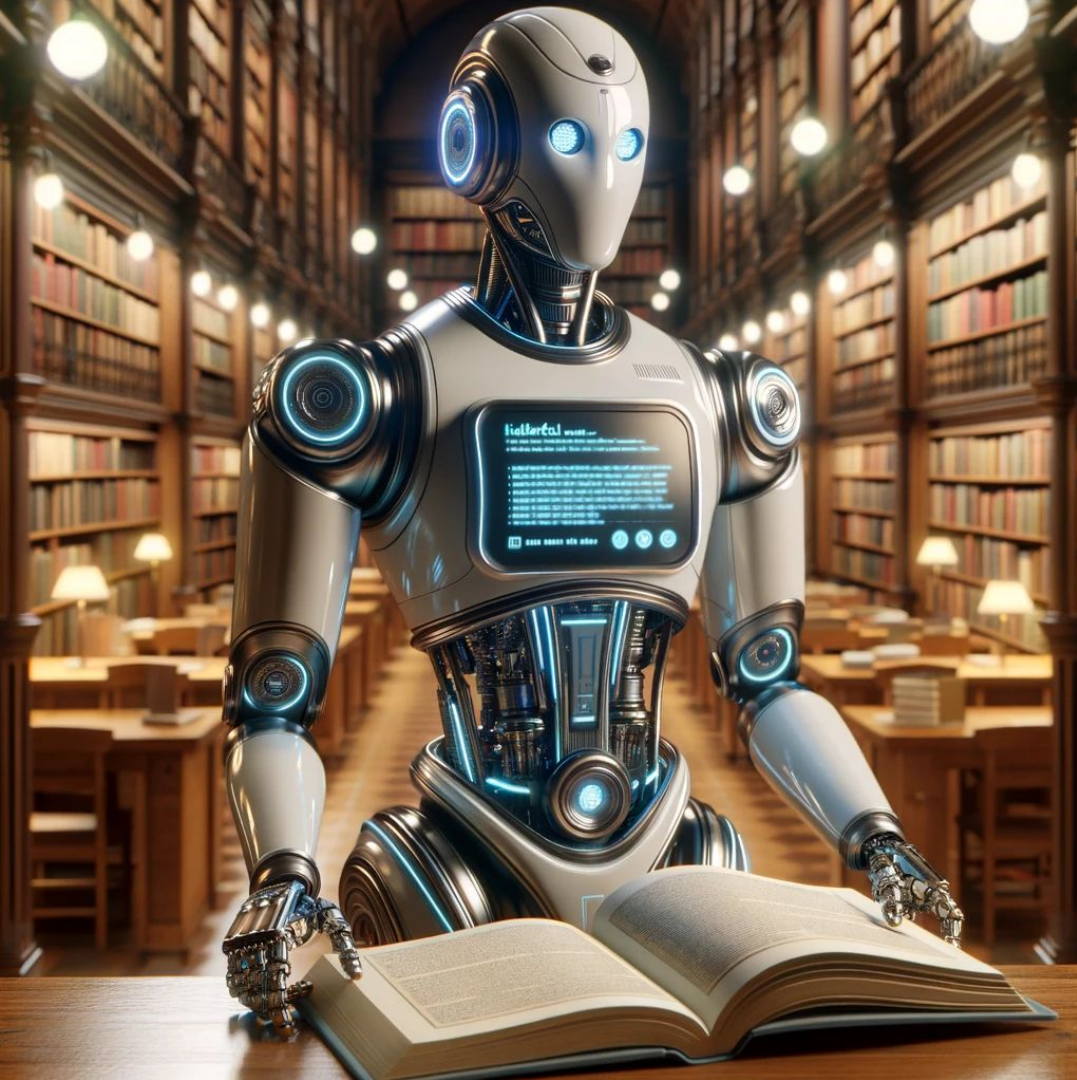
Checking our work...

<i>Original Description</i>	<i>Reformatted Description</i>	<i>Missing information</i>
v.92 pg.961-1944 1996	v.92 (1944) pg.961-1944	1996
Jan. 1981 - June 1982	(1981: Jan.-June)	1982
v.33:no.9,11 2020	v.33: no.9 (2020)	11
nos.3-5, 10	no.3-5	10

Next Steps...

- From normalized item descriptions, to 866s to coded fields
- National/international policies and templates for standardized 866s
- Or not!
- What other periodical problems can this solve?





Identifying Deaccessions

Potential Benefits of AI projects
and development of new
collection analysis tools

- AI for item descriptions
- New Alma collections analysis features
- Shiny with WorldCat
- [KNIME](#) Partnership

Q&A

Ask me some Qs and I'll try to A them

< Physical Item Editor

General

ENUM/CHRON

Notes

History

General Information

Barcode

30530005171343X

Generate

Material type

Issue

Provenance

Property of George Washington University

PO Line

Receiving date

08/30/2017

Enumeration A

22

Chronology I

1967

Description

v.22 (1967) suppl.

Generate

Pages

E:

Maximizing Storage

Materials in Shared Collections Facility in excess of current MOU requirements:

- Monographic titles in SCF 1 with possible excess copies: 124,553
- Serial titles in SCF 1 with possible excess copies: 9,427

