

Using ChatGPT to format Alma item descriptions:

a mostly successful experiment by
Aaron Krebeck



BOB'S BURGERS

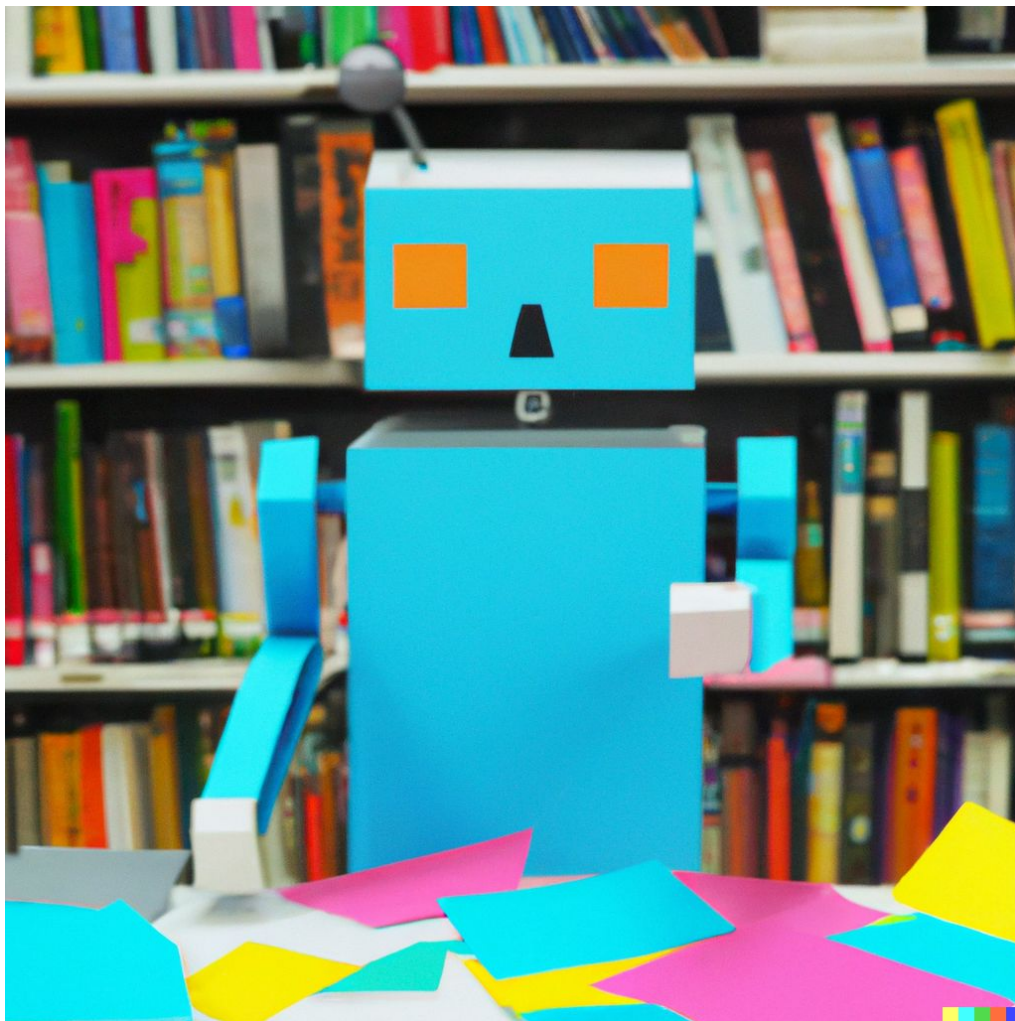


Getting Started

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

The purpose of these recommendations is to standardize serial holding and item descriptions in the WRLC Alma Institution Zones, and to facilitate the discovery of print periodicals in the WRLC instances of Primo VE.

Recommendations **did not include** any initiative to correct the millions of descriptions that were already in the catalog.





Why we need help from AI

- Our new policies state, “Legacy records should be corrected only as time and resources allow.” 😬
- 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 exactly what members of our consortia want from us

The Proof of Concept

- Slowly refined a query with ChatGPT
- “Guide to style” function
- [Item description templates](#) were very helpful
- ChatGPT free tier only allows 100 lines of input
- ~70 test strings to reformat 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

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)



Putting This Into Production

- July 6th code interpreter beta
- ChatGPT plus account required
- Alma Analytics query of description fields
- Script for batch changing item records in Alma



[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,").
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	1960 v.12	v.12 (1960)
1	31194000000137X	v.4	v.4
2	31194000000822X	v.19 1977	v.19 (1977)

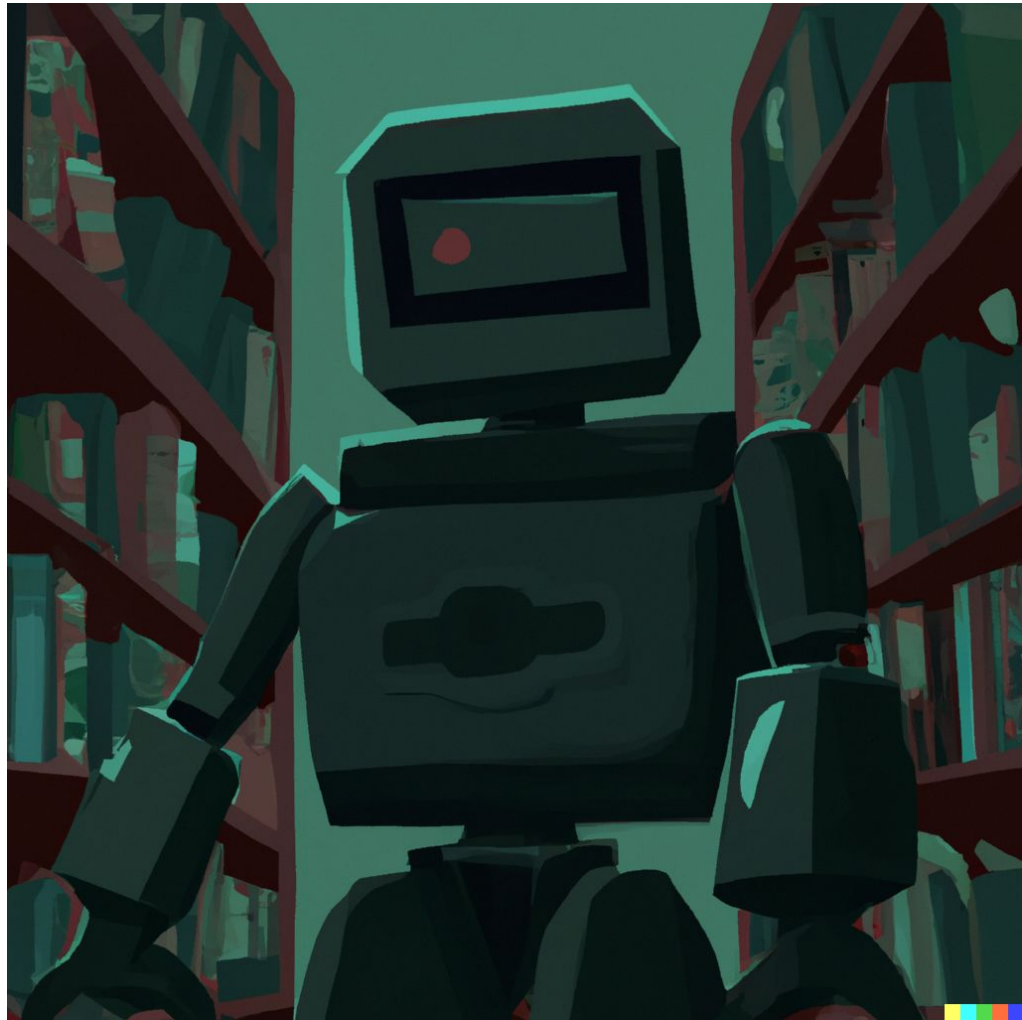


Putting This Into Production

- July 6th code interpreter beta
- ChatGPT plus account required
- Alma Analytics query of description fields
- [Script for batch changing item records in Alma](#)

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?



?



?