AI, Cataloging & Metadata

Sai Deng
University of Central Florida, sai.deng@ucf.edu

Part of the Cataloging and Metadata Commons

Find similar works at: https://stars.library.ucf.edu/ucfscholar
University of Central Florida Libraries http://library.ucf.edu

This Other Presentation is brought to you for free and open access by STARS. It has been accepted for inclusion in Faculty Scholarship and Creative Works by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation
Deng, Sai, "AI, Cataloging & Metadata" (2023). Faculty Scholarship and Creative Works. 1251. https://stars.library.ucf.edu/ucfscholar/1251
AI, CATALOGING & METADATA
Sai Deng, University of Central Florida Libraries
**AI IN A GLIMPSE**

- **Artificial intelligence (AI)**
  - AI is the theory and development of computer systems that are able to perform tasks that normally require human intelligence. (Schatsky)

- **AI Graph by Michael Mills (Neota Logic)**

- **AI is not a new concept**
  - John McCarthy coined the term in 1955.
  - More recent hot topic: Conversational AI, Chatbots
LITERATURE: AI, CATALOGING & METADATA

• The Cataloging & Metadata Field’s Response to AI

  • **Journal Article:** From ChatGPT to CatGPT: The Implications of Artificial Intelligence on Library Cataloging (Brzustowicz) DOI: 10.5860/ital.v42i3.16295
  
  • **Controversial AutoCat discussion**

  • **ALA Core Artificial Intelligence and Machine Learning in Libraries Interest Group**

  • **ALA Annual Programs,** e.g., Can OpenAI Be Taught to Catalog? (Heng, G., Holvoet, K. & Lindsey N., 2022) [https://scholarworks.calstate.edu/concern/presentations/sq87c223n](https://scholarworks.calstate.edu/concern/presentations/sq87c223n)

  • **National Archives and Records Administration** seeks to limit manual tagging for metadata automation (nextgov.com)

  • **Training in the Cataloging Department:** ChatGPT, AI “Creators” & Cataloging (Webinar)

  • **Questions worth exploration:** How can AI assist in data creation and enrichment, improve data abstracting, indexing, enhance data quality, cleaning, validation, reconciliation, as well as data transformation, interoperability and visualization? What are the opportunities and challenges?
A FEW SIMPLE EXPERIMENTS: CREATE A MARC RECORD

• Task 1: Create a MARC Record Using Bard and ChatGPT (3.5)
  • Create a marc record for Walks in the City of Canton by Guangdong People's Publishing House.
  • Bard’s response:

Here is a MARC record for Walks in the City of Canton by Guangdong People’s Publishing House:

```
Leader: 02000874 20231027 20231027 4500 MARC-8 eng d
001: 1234567890
003: ENG
005: 20231027143415.0
007: 
020: 9787218138879
035: 20231027143415.0
100: 1 Gray, John Henry.
245: 1 Walks in the City of Canton: a late Qing panoramic travelogue / by John Henry Gray; translated by Li Guoqing and Deng Sai
500: 1 Guangzhou (China) -- Description and travel.
560: 0 Guangzhou (China) -- History.
560: 0 Guangzhou (China) -- Geography.
```

(*)Followed by field explanation and notes, and also a disclaimer: Please note that this is just an example MARC record. Actual MARC records may vary depending on the specific information of the book.)
A FEW SIMPLE EXPERIMENTS:
CREATE A MARC RECORD (CONT.)

• Create a MARC record for Walks in the City of Canton by Guangdong People's Publishing House.

• ChatGPT’s response:

A MARC (Machine-Readable Cataloging) record for a book titled "Walks in the City of Canton" by "Guangdong People's Publishing House" might look something like this:

=LDR  00000nam a2200000i 4500
=001 0123456789  
=003 DLC  
=005 20231027000000.0  
=008 230427s2023    c          000 0 eng d
=020 \$a978-1-234567-89-0  
=040 \$aDLC$bc$erda$epn  
=100 1\$aGuangdong People's Publishing House  
=245 10\$aWalks in the City of Canton$eGuangdong People's Publishing House  
=250 \$aFirst edition  
=264 \$a2023  
=300 \$a200 pages  
=336 \$atext$2rdacdcontent  
=337 \$aunmediated$2rdamedia  
=338 \$avolume$2rdacarrier  
=546 \$aEnglish  
=651 \$aCanton (China)$vDescription and travel  
=700 1\$aGuangdong People's Publishing House  
=856 425\$Cover image$uhttps://example.com/cover.jpg  
=856 413\$Sample text$uhttps://example.com/sample.txt  
=856 425\$Publisher website$uhttps://www.example.com  
=852 \$81.2$8LIBRARY CODE
A FEW SIMPLE EXPERIMENTS:
CREATE A MARC RECORD

• Create a MARC record for Walks in the City of Canton by John Henry Gray

  Bard’s response:
  LDR 00000nam 22000000i 4500
  001 000000001 Fabricated identifier
  003 RDA RDA treated as agency
  005 20231102103844.0
  008 1919080220192019cnuuu 000 0eng d
  010 20190802 Fabricated LC Control Number
  020 9787545432363 -Where did you get this ISBN? 022 cn 035 $a (OCoLC)ocm1084185895 Fabricated LC Number
  040 $a DLC $c CNF0324680 Fabricated cataloging source
  050 $a 915.120471 Wrong LC Call Number
  082 0$a 910$b 915.120$c 915.1204$d 915.120471
  100 $a Gray, John Henry, $d 1837-1921.
  245 $a Walks in the City of Canton / $c by John Henry Gray.
  300 $a 162 p.; $c 22 cm.
  362 $a (Walks in historic China ; 2)
  500 $a A fascinating account of Canton in the early 1900s, by a British missionary who lived in the city for many years. Gray’s book is full of interesting details about the city's culture, history, and architecture.
  504 $a Translated from English.
  650 $a Canton (China) -- Description and travel.
  650 $a Canton (China) -- Social life and customs.

  Observations:
  • Bard generates a record based on some template with true and untrue info;
  • Record can be improved with modified prompts (supplying more information);
  • Can ask AI about field coding individually and supply correct info.

  Fabricated identifier
  Fabricated LC Control Number
  Wrong LC Call Number
  No geographic headings
  No series name
  Still no 700s for translators
A FEW SIMPLE EXPERIMENTS: CLASSIFY A BOOK

• Task 2: Classify a Book Using Bard and ChatGPT (3.5)
  • Construct a LC call number for Walks in the City of Canton by Guangdong People's Publishing House.
  • Bard’s response: DS796.C2 W35 2023 (*followed by some explanations of the call number)
    Used Title as cutter number Fabricated pub year
  • Rephrased prompt: Construct a LC call number for Walks in the City of Canton by John Henry Gray (publisher: Guangdong People's Publishing House)
  • Bard’s response: ...DS796.C2 G73 1875 2023 Included 2 years
  • Comments: knows some general principles, but generated a call number for a non-existing book
  • ChatGPT’s response to the initial prompt: DS730.G7 G73 2023 Class number problematic Fabricated pub year
  • With explanation of the components, also a note: Please note that the Cutter number "G73" is a hypothetical example. In practice, you would consult the Library of Congress classification schedules or work with a librarian to determine the appropriate Cutter number based on the author's name and ensure that it is unique within the library's catalog.

Manually Check Classification Web:

Cataloger supplied call number: DS797.32.G836 G66412 2019
A FEW SIMPLE EXPERIMENTS: ASSIGN THESAURUS TERMS

• Task 3: Assign Thesaurus Terms for An Article Using Bard and ChatGPT (3.5)
  • Find AGROVOC terms for the following article. Title: Connected Water Heater Load Shifting and Energy Efficiency Evaluation for the Southeast: Winter Laboratory Assessment, Final Report. Abstract: ...
  • Bard’s response: Generated 11 terms and 8 more specific terms
  • ChatGPT’s response: Generated 20 terms
  • Rephrased prompt: Find 3–8 AGROVOC terms for the following article. Title: ...
    • Comment: terms seem to be relevant, but 2 of the 6 terms are not AGROVOC terms
  • ChatGPT’s response: Load Shifting, Energy Efficiency, Heat Pump Water Heaters (HPWH), Electric Resistance Water Heaters (ERWH), Southeastern United States (*with explanation)
    • Comment: terms seem to be relevant, but 4 of the 5 terms are not AGROVOC terms
  • Rephrased prompt: AGROVOC Thesaurus can be accessed via https://agrovoc.fao.org/browse/agrovoc/en. Find 3–8 AGROVOC terms for the following article...
  • Bard’s response: I’m sorry. I’m not able to access the website(s) you’ve provided. The most common reasons the content may not be available to me are paywalls, login requirements or sensitive information, but there are other reasons that I may not be able to access a site.
  • ChatGPT’s response 1: Water heaters, Energy efficiency, Heat pump technology, Electric resistance, Grid-connected devices, Energy consumption, Load shifting, Southeastern United States
    • Comment: terms seem to be relevant, but 6 of the 8 terms are not AGROVOC terms
  • ChatGPT’s response 2: I couldn’t directly access the AGROVOC Thesaurus to fetch specific terms, but I can suggest some relevant terms based on the provided article: Water Heater, Energy Efficiency, Load shifting, Heat Pump Water Heater (HPWH), Electric Resistance Water Heater (ERWH), Grid-Connected, Southeastern United States, Energy Research
A FEW SIMPLE EXPERIMENTS: ASSIGN THESAURUS TERMS (CONT.)

- Find 3-8 FAST terms for the following article. Title: Connected Water Heater Load Shifting and Energy Efficiency Evaluation for the Southeast: Winter Laboratory Assessment, Final Report. Abstract: ...

- Bard’s response: Grid-connected heat pump water heater (HPWH), Load shifting, Critical peak (CP) event, CTA-2045-A standard, Southeastern United States
  - Comment: These terms are not FAST terms

  - Comment: terms seem to be relevant, but 2 of these terms are not FAST terms

- Rephrased prompt: FAST can be accessed via https://fast.oclc.org/searchfast/. Find 3-8 FAST terms for the following article. Title:....

- Bard’s response: I've found the following FAST terms for the article: Heat pump water heaters, Electric resistance water heaters, Load shifting, Critical peak pricing, Demand response, Smart grid, Energy efficiency, Southeastern United States. Websites reviewed: fast.oclc.org/searchfast/
  - Comment: These terms are not FAST; only one is a synonym of a FAST term

- ChatGPT’s response: I'm sorry, but I cannot access external websites or databases, including the FAST authority file. However, I can suggest some potential FAST (Faceted Application of Subject Terminology) terms based on the information provided in the article: Water heaters, Energy Efficiency, Heat pump water heaters (HPWH), Grid-connected appliances, Load shifting, Southeastern United States, Control command designs, Energy consumption patterns.

- All the above testing are done on ChatGPT 3.5 and Bard Experiment.
ChatGPT 4.0:

- **ChatGPT 4.0**: small monthly fee;
  - It can now process up to 25,000 words of text from the user;
  - It has a character limit of 8,192 characters (*for generated content);
  - It allows users to upload an article and ask for a summary;
  - It seeks to access external websites or databases via “plugins” in the near future;
  - It’s trained on a dataset of articles that have already been assigned LCSH terms; can generate more accurate LCSH terms
OTHER EXPERIMENTS:
ARTICLE REVIEW, TRANSLATION CHECK

• Check Article or Text for Errors
  • Please check to see if the following article has errors: “Chinese Studies in the West: The Chinese Record” Series...
  • ChatGPT’s response:
    • In the sentence "The colonizing countries of the West, in seeking new opportunities, became interested in knowing China and its people to help govern their 'new territories,'" it seems that the phrase "help govern their 'new territories'" is not clear or may be misinterpreted. It could be rephrased for clarity.
    • In the sentence "In The Chinese as They Are: Their Moral, Social, and Literary Character (1841), George Tradescant Lay introduced China to the West as he understood it," the phrase "as he understood it" is unnecessary and could be omitted...
  • Comment: Part of the comments make sense; need to break the article into small sections to get more detailed responses

  • Please check the following article's grammar: “Chinese Studies in the West: The Chinese Record”...
  • ChatGPT's response:
    • In the second paragraph, after "researched the emergence," add a comma for better readability: "researched the emergence, phenomenon, and essence of Orientalism in the West."
    • In the third paragraph, remove the space between "Sino-" and "Western." The correct term is "Sinowestern."
  • Comment: Part of the comments make sense; it could be useful for article review

• Write A Summary for A Book or An Article
  • Comment: can be used to get more ideas; cannot depend on the outcome

• Check Manually Translated Text Against AI-Translated Text
  • Comment: good for checking some individual terms, not the sentence flow; can spot translation errors/negligence
This talk was originally prepared for the UCF Libraries’ AI Interest Group meeting, and the Bing Chat Enterprise test section was added for the Cataloging departmental meeting

- Protected data; data not used for training; has three modes: Balanced, Creative, Precise

Access Bing Chat Enterprise at UCF:

- Start at the public Bing Chat
- Click “sign in” at the top-right
- Select “work or school” for the type of account
- Type your full UCF email and click NEXT
- Log in with your NID and NID password.
  (Note: you may need to alter your SafeSearch settings away from “Strict”)
- Above the box where you would type your question, you will see “Your personal and company data are protected in this chat” – this is how you know you are in BCE.

Task 1: Create a MARC Record Using Bing Chat Enterprise

- Create a MARC record for Walks in the City of Canton by Guangdong People's Publishing House
  - Bing Chat Enterprise’s Response: (*More creative style)
  - …the MARC record should be based on the original work, not the translation…
  - Learn more: 1. stars.library.ucf.edu; 2. books.google.com; 3. onlinebooks.library.upenn.edu; 4. mp.weixin.qq.com
Create a MARC record for Walks in the City of Canton by Guangdong People's Publishing House

Response (*More Balanced Style)

LDR 00000nam a2200000 a 4500
001 12345678
005 2023111513153.0
008 231115s2023    xxu           000 0 eng d
020 $a 978-7-5086-0000-0
040 $a DLC $b eng $e rda $c DLC $d DLC
100 1 $a Guangdong People's Publishing House.
245 10 $a Walks in the City of Canton / $c Guangdong People's Publishing House.
264 1 $a [Place of publication not identified] : $b [Publisher not identified], $c 2023.
300 $a 200 pages : $b illustrations ; $c 23 cm
336 $a text $b txt $2 rdacontent
337 $a unmediated $b n $2 rdamedia
338 $a volume $b nc $2 rdacarrier
500 $a Includes bibliographical references.

I used the ISBN from the second result, which is a digitized version of the original book by John Henry Gray. I used the Chinese title from the first result, which is a Wikisource page that contains a description of the city of Canton from the Chinese Repository. I did not find any information about the Chinese translation of the book, so I assumed it was the same as the original title. Learn more:
1. archive.org 2. en.wikisource.org 3. bing.com 4. academia.edu 5. bing.com

Responses to my further questions/requests:

=245 10$a 广州漫步 / $c John Henry Gray 著 ; 李国庆, 邓赛 译.
=264 1\$a 广州 : $b 广东人民出版社, $c 2019.
=300 1\$a 320 pages : $b illustrations ; $c 24 cm
=336 1\$a text $b txt $2 rdacontent
=337 1\$a unmediated $b n $2 rdamedia
=338 1\$a volume $b nc $2 rdacarrier
=700 1\$a 李国庆, $e translator.  No English equivalent for translators’ names
=700 1\$a 邓赛, $e translator.
=830 0\$a 广东人民出版社文化丛书 Wrong series name; no English equivalent
• Task 2: Classify a Book Using Bing Chat Enterprise
  • Construct a LC call number for Walks in the City of Canton by Guangdong People's Publishing House.
  • Bing Chat Enterprise’s Response:
    • DS796.C3 G73 2019
    • I used the following steps to create the LC call number:
      • I searched for the original book by John Henry Gray on WorldCat and found that it has the LC call number DS796.C3 G73.
      • I checked the [LC classification outline] and found that DS is for Asia, DS701-799 is for China, and DS796 is for local history and description of China. C3 is for Guangzhou (Canton), and G73 is for Gray as the author’s name.
      • I added the year of publication of the Chinese translation, which is 2019, to the end of the LC call number. This is to distinguish it from the original edition, which was published in 1875.
  • Retried later:

Check Its Claims:
From OCLC Connexion:

From Classification Web:

*Can use Bing Chat Enterprise as a reference tool, but cannot adopt its suggestions without checking.
Task 3: Assign Thesaurus Terms for An Article

- Find 3-8 AGROVOC terms for the following article. Title: Connected Water Heater Load Shifting and Energy Efficiency Evaluation for the Southeast: Winter Laboratory Assessment, Final Report. Abstract: ...

- Bing Chat Enterprise’s Response:
  - water heaters
  - energy efficiency
  - load management
  - heat pumps
  - electric power
  - laboratory experimentation
  - Southeastern United States of America

- You can browse the AGROVOC online to find more terms and definitions related to your topic.

- Comments: 3 of the 7 terms are not in AGROVOC

Chat with AI:

- water heaters, load management, Southeastern United States of America are not in AGROVOC. Where did you find them?

- You are right, water heaters, load management, and Southeastern United States of America are not in AGROVOC. I found them by using a web search tool that returns results from Bing. I used these terms as examples of how to search for relevant information related to your article...

- Find only AGROVOC terms for the above article.

- energy efficiency
  - load management
  - heat pumps
  - electric power
  - laboratory experimentation
  - Southeastern United States of America

- Comments: 2 of the 6 terms are still not in AGROVOC
• Find 3-8 FAST terms for the following article. Title: Connected Water Heater Load Shifting and Energy Efficiency Evaluation for the Southeast: Winter Laboratory Assessment, Final Report. Abstract: ...

• Response:
  • Water heaters
  • Energy efficiency
  • Load management (Electric utilities)
  • Heat pumps
  • Electric power consumption
  • Laboratories
  • Southeastern States
  • Comments: 1 of the 7 terms are not in FAST

Rephrased prompts:
Find only FAST terms for the following article...
Response:
  Water heaters
  Energy efficiency
  Heat pumps
  Electric power consumption
  Laboratories
  Southeastern States

Find more FAST terms for the following article...
  Water-supply engineering
  Renewable energy sources
  Energy conservation
AI POTENTIALS IN CATALOGING & METADATA

• AI Potentials in Cataloging & Metadata:
  • Generate metadata with more accuracy (summaries, reviews, recommendations; subjects, classification numbers, records);
  • Identifying and correcting metadata errors and inconsistencies;
  • Enhance discoverability and metadata customization via generating metadata more aligned with user interests;
  • Help catalogers to get more information for better decisions;

• For cataloging & metadata professionals: make ourselves relevant to AI; expand our knowledge on machine learning; need to review and verify AI generated content; provide feedback to AI-based cataloging and metadata solutions; be cautious about bias and privacy issues; position ourselves as partners in broader conversations of AI in teaching, learning and research

Image source: https://www.wired.com/2015/11/star-wars-characters-will-now-teach-your-kids-to-code/
Thank you