Metadata Enrichment with Named Entity Recognition using GPT-4

Ashwin Nair, Min Hoon Ee and Robin Dresel

National Library Board of Singapore, 100 Victoria Street Singapore 188064

Abstract

To enhance the user experience and resource discoverability of Infopedia, the Singapore encyclopedia, the National Library Board of Singapore (NLB) uses Generative Pre-trained Transformer 4 (GPT-4) for Named Entity Recognition (NER), aiming to automate metadata enrichment of its digital encyclopedia articles. This initiative leverages GPT-4's capabilities in accurately identifying and incorporating relevant Singaporean entities before integrating them into the NLB's Knowledge Graph, improving recommendations of related resources. An evaluation on a subset of 100 articles demonstrates a precision score of 0.975, indicating high entity detection with minimal inaccuracies. The team acknowledges challenges related to GPT-4's black-box nature and the potential for non-reproducibility. This effort illustrates the potential of generative AI to streamline metadata enrichment processes, offering a promising avenue for enhancing metadata of digital libraries.

Keywords

Named Entity Recognition, Large Language Models, Prompt Engineering, Metadata Enrichment, Digital Libraries

1. Introduction

The National Library Board of Singapore (NLB) manages an extensive collection of both physical and digital collections, with resources from The National Archives under its umbrella as well. The Singapore Infopedia [1] is a digital encyclopedia dedicated to Singapore. It contains articles covering the country's history, culture, prominent personalities, and significant events and is accessible via our National Library Online [2] platform. Each article offers a comprehensive overview of its respective topic and includes references for in-depth research. To allow for integrations with our physical and digital resources, one of the services we are developing is the Singapore Infopedia Sidebar. The Sidebar serves as a recommender engine and points the user to resources that are related to the article. The Sidebar uses metadata of the individual article to recommend related Works based on relationships established in our Linked Data Knowledge Graph (KG).

NLB's Knowledge Graph contains data coming from various source systems. These data are presented in diverse formats that require conversion to establish a unified graph centered on entities. The source systems include: The Integrated Library System (ILS), housing all MARC21 records for physical and eBook collections in National and Public libraries; The Content Management System (CMS), containing Dublin Core (DC) records for digital resources and DC records derived from the ISAD-G archival standard detailing The National Archives collections; locally stored authority records utilized by cataloguers and exported to CSV-XML format (Dresel ,2023).

[1] These authors contributed equally.
ashwin_nair_madhavan@nlb.gov.sg (A. Nair); ee_min_hoon@nlb.gov.sg (M.H. Ee); robin_dresel@nlb.gov.sg (R. Dresel)
0000-0002-8386-057X (A. Nair); 0009-0009-7101-9319 (M.H. Ee); 0000-0002-2183-1204 (R. Dresel)

https://doi.org/10.23106/dcmi.952445840

During the development process, we decided to enhance the metadata of Infopedia collection by adding on to the entities that are mentioned in the articles. These entities will be referenced and used to add related resources, thereby enhancing the discovery journey of the users.

2. Problem Statement

It is time consuming for a human to ascertain the named entities for each article in the Infopedia collection. A possible solution we are exploring is to use Generative Artificial Intelligence (GEN AI), specifically through a Large Language Model (LLM), to add entities which are being mentioned, into the metadata of the articles to increase the number of relationships through the KG and thus related resources. Specifically, we used Generative Pre-trained Transformer 4 (GPT 4). It is one of the most robust models, outperforming OpenAI's previous large language models (OPEN AI, 2023).

Through Named Entity Recognition (NER), we can extract named entities in context, effectively harnessing unstructured data to establish connections to additional resources within the extensive NLB collection (Goh, 2018).

The objectives for using GPT-4 are:

1. **Named Entity Recognition**, with the extracted entities used to match against those in our Knowledge Graph
2. **Expand NLB's Knowledge Base** with entities that are extracted but do not currently exist in our Knowledge Graph
3. **Enhance the Infopedia Sidebar**

Our focus is solely on Singaporean entities due to Infopedia's Singapore-centric nature, aiming to utilize NLB's local name authorities and establish connections between these entities. NLB has been actively developing and managing name authorities that are user-centric, Singapore-centric, and Southeast Asian-centric as part of its Knowledge Organization System (KOS) (Puay Eng, Hong, & Jailani, 2018). These authority records feature established terms for entities in Singapore and Southeast Asia, emphasizing local context details like dialect names for ethnic Chinese. Additionally, entries for entities not covered in international authorities, such as the Library of Congress Name Authorities (LCNA), are included to support digital cataloging needs (Ee, 2019).

3. Literature Review

Named Entity Recognition (NER) focuses on identifying entities within unstructured text and classifying them into predefined categories such as people, locations, and organizations (Nasar et al., 2021). This process is useful for extracting structured information, aiding in metadata enrichment and information retrieval within digital libraries.

As explained by Tjong Kim Sang and De Meulder (2003), NER relied heavily on rule-based systems, which used hand-crafted rules to identify entities based on lexical and syntactic cues. While effective in specific contexts, these systems were limited by their inflexibility and scalability, particularly in adapting to new domains or languages. Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs), represented a significant advance, leveraging annotated corpora to learn entity recognition patterns. Despite their improved generalization capabilities however, these models often required extensive, domain-specific datasets for training, limiting their applicability in dynamic or niche areas (McCallum & Li, 2003).

https://doi.org/10.23106/dcmi.952445840

The emergence of transformer-based large language models (LLMs), especially GPT-3.5 and GPT-4, has marked a paradigm shift in the field of Natural Language Processing (Yang et al., 2023). These models, characterized by their deep learning architectures, offer a profound understanding of context and semantics, across various domains and languages. One of the most notable features of LLMs is their capacity for few-shot learning, which allows them to adapt to new tasks by prompting with minimal examples. This adaptation occurs through in-context learning, reducing the necessity of parameter tuning (Brown et al., 2020). Chain-of-Thought (CoT) is a particularly effective method within prompt engineering for enhancing model reasoning and entity recognition capabilities. CoT guides large language models through a logical sequence of reasoning steps (Wei et al., 2022).

Integrating GPT-4 and prompt engineering into the metadata creation holds the opportunity of creating efficiencies in the metadata enrichment processes. This integration will add to the discoverability and accessibility of information, fostering more engaging and user-friendly digital archives. Nevertheless, the deployment of GPT-4 for NER within digital libraries is not devoid of challenges. Concerns over model hallucinations and false positives underscore the need for vigilance. Addressing these challenges, the application of GPT-4 and prompt engineering for Named Entity Recognition (NER) is intended to improve the discoverability and accessibility of digital content.

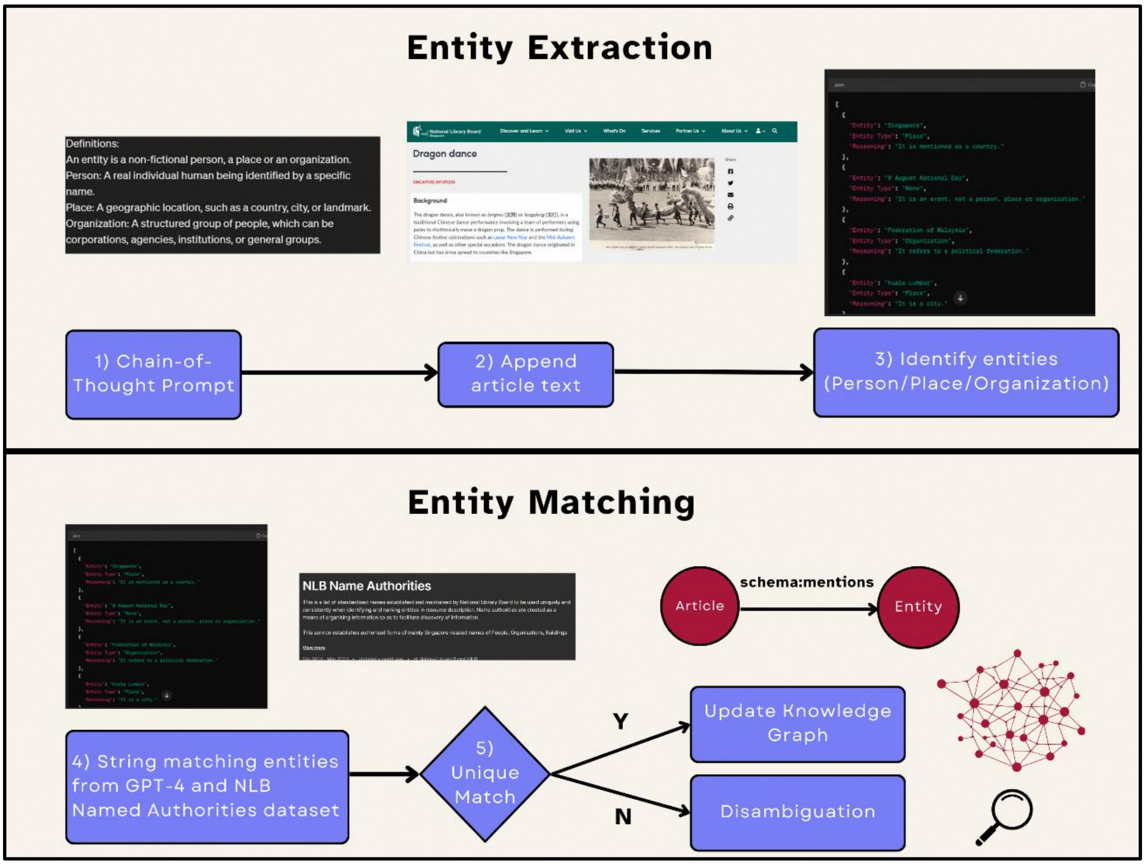
4. Methodology

The methodology draws inspiration from Ashok and Lipton's (2023) study on “PromptNER: Prompting For Named Entity Recognition” and a tutorial on prompt engineering and metadata presented by Qin and Yu (2023).

4.1. Entity Extraction

It involves 2 key stages: Entity Extraction and Entity Matching. In the Entity Extraction stage, prompt engineering is exploited to guide GPT-4 in accurately identifying entities such as People, Places, and Organizations within the articles. The Entity Matching process aligns the entities identified by GPT-4 with the NLB Named Authorities Dataset (National Library Board, 2023) through string matching. Finally, the extracted entities are updated into the Knowledge Graph, enriching the metadata associated with the articles. The metadata is used to improve the precision and contextual relevance of sidebar resource recommendations through the insertion of “schema:mentions” in NLB’s Knowledge Graph, thereby enhancing the overall user experience of the sidebar.

https://doi.org/10.23106/dcmi.952445840



Entity Extraction
1) Chain-of-Thought Prompt
2) Append article text
3) Identify entities (Person/Place/Organization)
Entity Matching
4) String matching entities from GPT-4 and NLB Named Authorities dataset
5) Unique Match
Y
N
Update Knowledge Graph
Disambiguation
Article
schema:mentions
Entity

Definitions:
An entity is a non-fictional person, a place or an organization.
Person: A real individual human being identified by a specific name.
Place: A geographic location, such as a country, city, or landmark.
Organization: A structured group of people, which can be corporations, agencies, institutions, or general groups.

```
[
  {
    "Entity": "Singapore",
    "Entity Type": "Place",
    "Reasoning": "It is mentioned as a country."
  },
  {
    "Entity": "9 August National Day",
    "Entity Type": "None",
    "Reasoning": "It is an event, not a person, place or organization."
  },
  {
    "Entity": "Federation of Malaysia",
    "Entity Type": "Organization",
    "Reasoning": "It refers to a political federation."
  },
  {
    "Entity": "Kuala Lumpur",
    "Entity Type": "Place",
    "Reasoning": "It is a city."
```

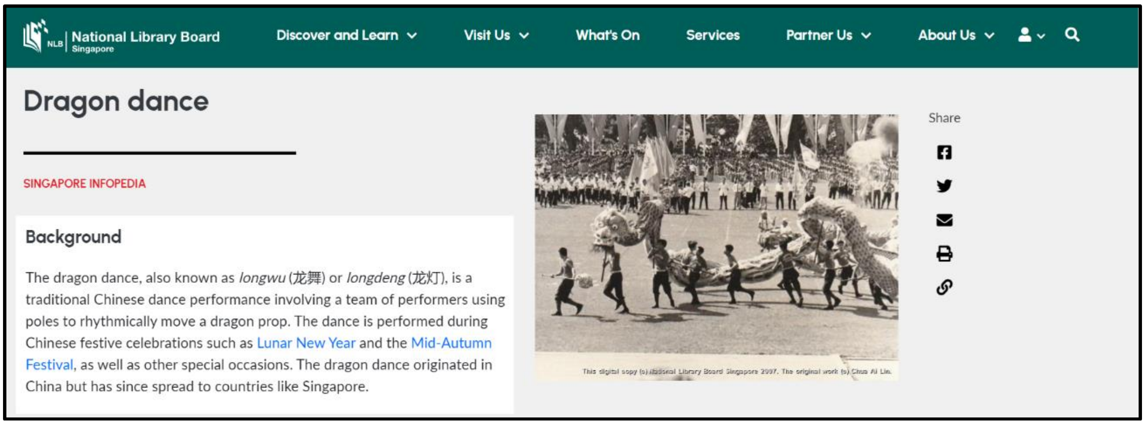
NLB Name Authorities

This is a list of standardised names established and maintained by National Library Board to be used uniquely and consistently when identifying and naming entities in resource description. Name authorities are created as a means of organising information so as to facilitate discovery of information.

This service establishes authorised forms of mainly Singapore-related names of People, Organisations, Buildings

View more

Figure 1: An overview of the automated process for enhancing metadata enrichment in the NLB's Knowledge Graph. The process begins with extracting entities from articles and matching them with the NLB Named Authorities Dataset. Unique matches are added to the knowledge graph, multiple matches undergo disambiguation, and unmatched entities are evaluated for dataset inclusion.



NLB | National Library Board Singapore
Discover and Learn
Visit Us
What's On
Services
Partner Us
About Us
Share
This digital copy (c) National Library Board Singapore 2007. The original work (c) Chua Ai Lin.

Dragon dance

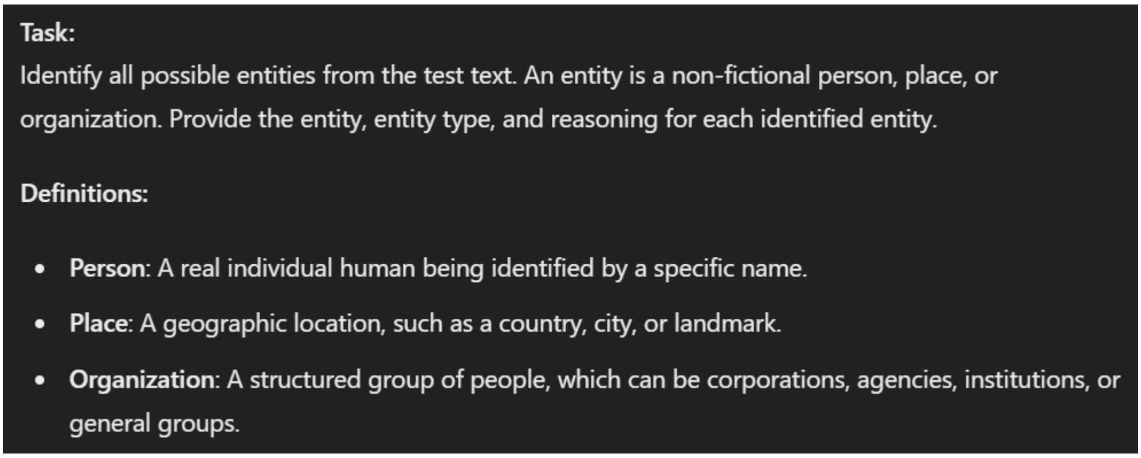
SINGAPORE INFOPEDIA

Background

The dragon dance, also known as *longwu* (龙舞) or *longdeng* (龙灯), is a traditional Chinese dance performance involving a team of performers using poles to rhythmically move a dragon prop. The dance is performed during Chinese festive celebrations such as Lunar New Year and the Mid-Autumn Festival, as well as other special occasions. The dragon dance originated in China but has since spread to countries like Singapore.

Figure 2: Dragon dance performance. Image source: Singapore Infopedia, National Library Board. This is an article by Infopedia on the dragon dance. (https://www.nlb.gov.sg/main/article-detail?cmsuuid=848b0702-f59e-4ab6-8dc7-3f7f218131b7)

https://doi.org/10.23106/dcmi.952445840



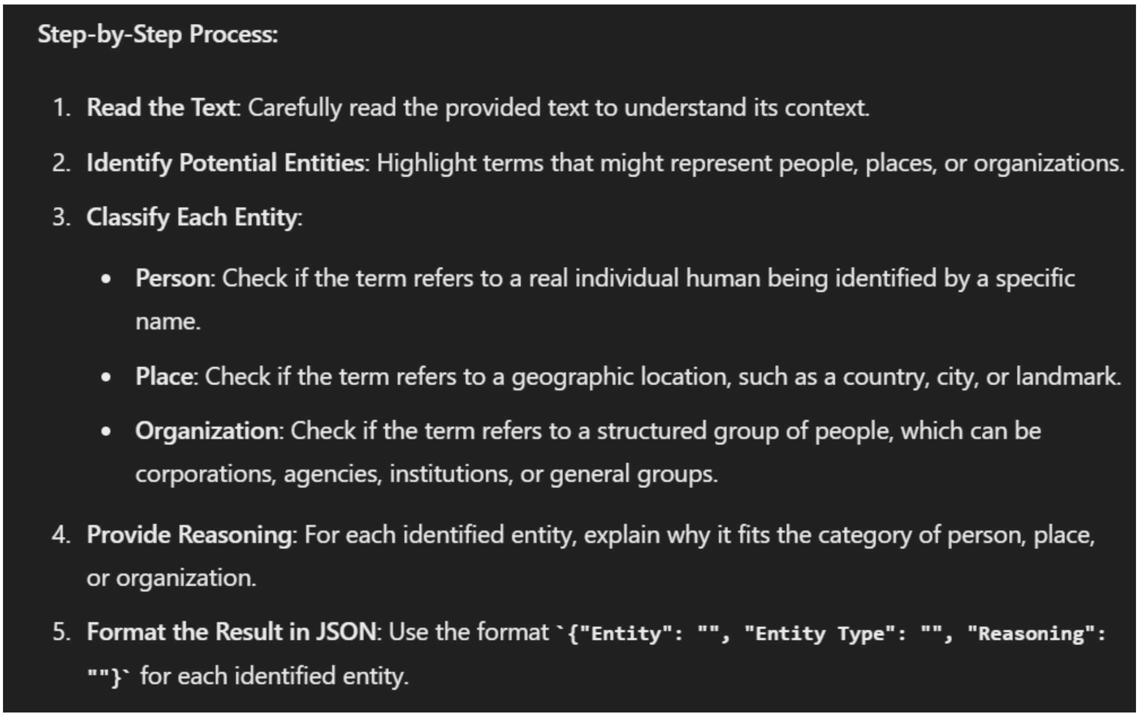
Task:

Identify all possible entities from the test text. An entity is a non-fictional person, place, or organization. Provide the entity, entity type, and reasoning for each identified entity.

Definitions:

- **Person**: A real individual human being identified by a specific name.
- **Place**: A geographic location, such as a country, city, or landmark.
- **Organization**: A structured group of people, which can be corporations, agencies, institutions, or general groups.

Figure 3: The initial stage of our prompt, where we outline the primary task and provide definitions of the types of entities to be identified.

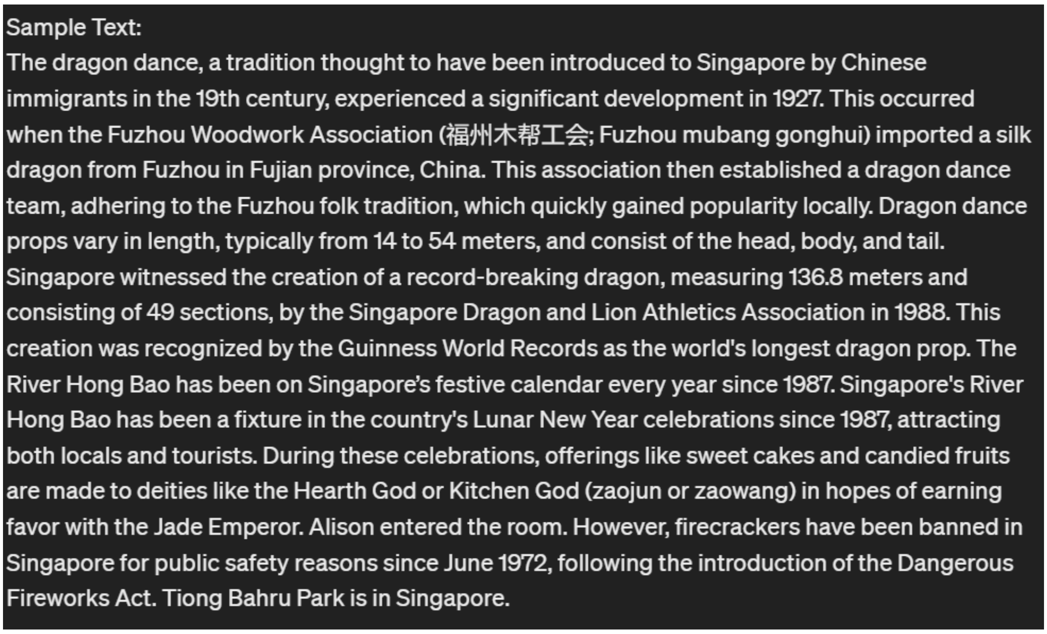


Step-by-Step Process:

1. **Read the Text**: Carefully read the provided text to understand its context.
2. **Identify Potential Entities**: Highlight terms that might represent people, places, or organizations.
3. **Classify Each Entity**:
    - **Person**: Check if the term refers to a real individual human being identified by a specific name.
    - **Place**: Check if the term refers to a geographic location, such as a country, city, or landmark.
    - **Organization**: Check if the term refers to a structured group of people, which can be corporations, agencies, institutions, or general groups.
4. **Provide Reasoning**: For each identified entity, explain why it fits the category of person, place, or organization.
5. **Format the Result in JSON**: Use the format `` `{"Entity": "", "Entity Type": "", "Reasoning": ""}` `` for each identified entity.

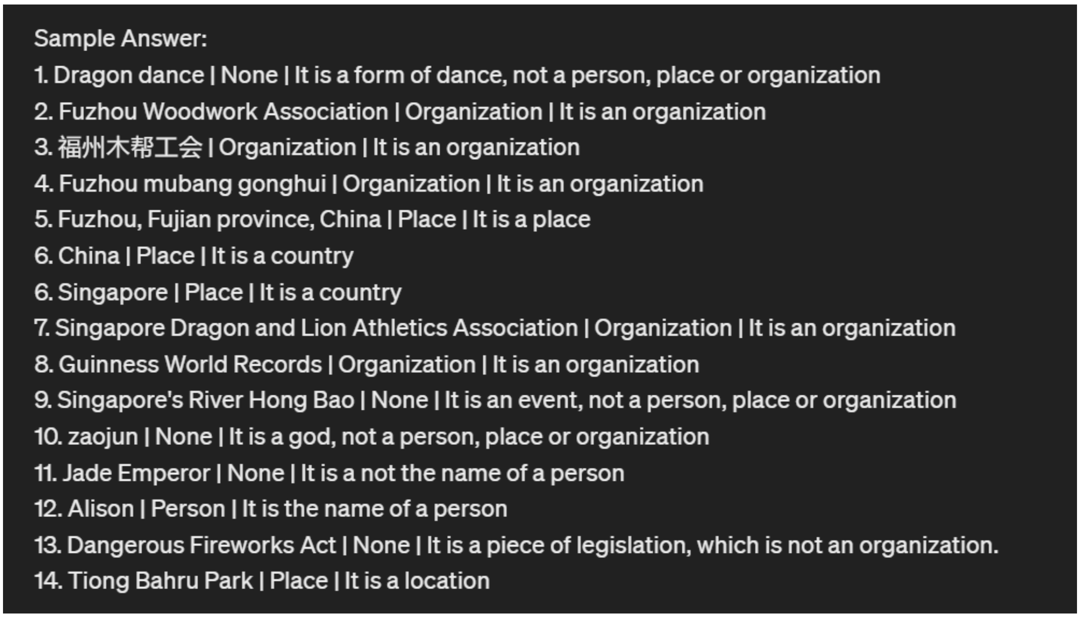
Figure 4: A Chain-Of-Thought prompt that guides GPT-4 through intermediate steps to identify entities, provide reasoning, and generate the information in a structured data format for integration into the knowledge graph.

https://doi.org/10.23106/dcmi.952445840



Sample Text:
The dragon dance, a tradition thought to have been introduced to Singapore by Chinese immigrants in the 19th century, experienced a significant development in 1927. This occurred when the Fuzhou Woodwork Association (福州木帮工会; Fuzhou mubang gonghui) imported a silk dragon from Fuzhou in Fujian province, China. This association then established a dragon dance team, adhering to the Fuzhou folk tradition, which quickly gained popularity locally. Dragon dance props vary in length, typically from 14 to 54 meters, and consist of the head, body, and tail. Singapore witnessed the creation of a record-breaking dragon, measuring 136.8 meters and consisting of 49 sections, by the Singapore Dragon and Lion Athletics Association in 1988. This creation was recognized by the Guinness World Records as the world's longest dragon prop. The River Hong Bao has been on Singapore's festive calendar every year since 1987. Singapore's River Hong Bao has been a fixture in the country's Lunar New Year celebrations since 1987, attracting both locals and tourists. During these celebrations, offerings like sweet cakes and candied fruits are made to deities like the Hearth God or Kitchen God (zaojun or zaowang) in hopes of earning favor with the Jade Emperor. Alison entered the room. However, firecrackers have been banned in Singapore for public safety reasons since June 1972, following the introduction of the Dangerous Fireworks Act. Tiong Bahru Park is in Singapore.

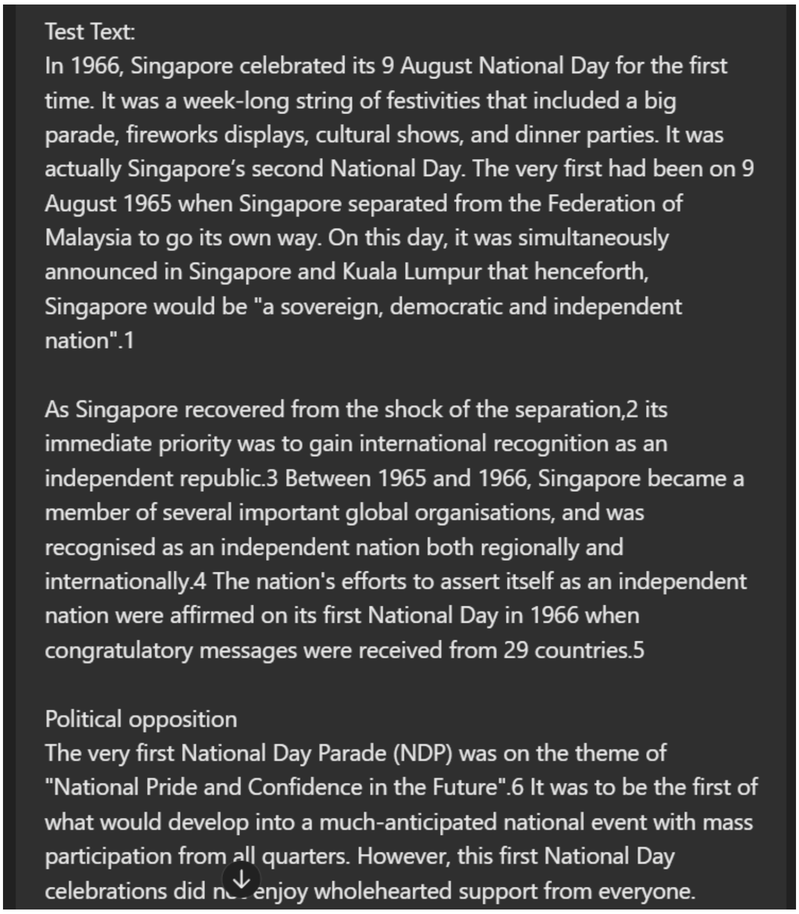
Figure 5: A crafted sample text used for the purpose of identifying entities.



Sample Answer:
1. Dragon dance | None | It is a form of dance, not a person, place or organization
2. Fuzhou Woodwork Association | Organization | It is an organization
3. 福州木帮工会 | Organization | It is an organization
4. Fuzhou mubang gonghui | Organization | It is an organization
5. Fuzhou, Fujian province, China | Place | It is a place
6. China | Place | It is a country
6. Singapore | Place | It is a country
7. Singapore Dragon and Lion Athletics Association | Organization | It is an organization
8. Guinness World Records | Organization | It is an organization
9. Singapore's River Hong Bao | None | It is an event, not a person, place or organization
10. zaojun | None | It is a god, not a person, place or organization
11. Jade Emperor | None | It is a not the name of a person
12. Alison | Person | It is the name of a person
13. Dangerous Fireworks Act | None | It is a piece of legislation, which is not an organization.
14. Tiong Bahru Park | Place | It is a location

Figure 6: Desired answers based on the sample text, along with the intermediate reasoning skills needed to arrive at these conclusions, separated by a delimiter "|".

https://doi.org/10.23106/dcmi.952445840



Test Text:
In 1966, Singapore celebrated its 9 August National Day for the first time. It was a week-long string of festivities that included a big parade, fireworks displays, cultural shows, and dinner parties. It was actually Singapore’s second National Day. The very first had been on 9 August 1965 when Singapore separated from the Federation of Malaysia to go its own way. On this day, it was simultaneously announced in Singapore and Kuala Lumpur that henceforth, Singapore would be "a sovereign, democratic and independent nation".1

As Singapore recovered from the shock of the separation,2 its immediate priority was to gain international recognition as an independent republic.3 Between 1965 and 1966, Singapore became a member of several important global organisations, and was recognised as an independent nation both regionally and internationally.4 The nation's efforts to assert itself as an independent nation were affirmed on its first National Day in 1966 when congratulatory messages were received from 29 countries.5

Political opposition
The very first National Day Parade (NDP) was on the theme of "National Pride and Confidence in the Future".6 It was to be the first of what would develop into a much-anticipated national event with mass participation from all quarters. However, this first National Day celebrations did not enjoy wholehearted support from everyone.

Figure 7: Test data provided for GPT-4 to identify entities.

4.2. Entity Matching

Entities that are extracted using GPT-4 are matched with the NLB Named Authorities Dataset, which consists of standardized names crucial to Singapore's politics, economy, society, culture, or history. This is achieved through:

- **Qualifier Removal:** Removing qualifiers from entity names to focus on core names.
- **Handling Alternate and Preferred Names:** Entities may be associated with multiple variations of names. All alternate and preferred labels will lead to a singular preferred entity through a directed graph.

There are 3 possible outcomes:

- **Exact Match:** Candidates which are equivalent terms are integrated into the knowledge graph. To indicate that the Singapore Infopedia article mentions these entities, we will link the article and the matched entity in our knowledge graph through “schema:mentions”.
- **Disambiguation:** Manual intervention is required when there are multiple non-exact string matches to resolve potential ambiguities or overlaps between entities.

https://doi.org/10.23106/dcmi.952445840

- **Evaluation:** Entities without matches will be evaluated for dataset inclusion in the NLB Named Authorities Dataset, if deemed significant to Singapore's context.

Entity resolution techniques were considered but ultimately not deployed. Entity resolution involves linking records across data sources that refer to the same entity, aiming to reduce redundancies and ensure data consistency. However, its deployment presents challenges, including technical complexities, resource demands, and the risk of false positives where unrelated records are mistakenly linked. The potential costs and risk of errors from inaccurate matches lead us to a cautious approach, concluding that the resources required and potential for mistakes may outweigh the benefits of automated entity resolution in aligning GPT-4 results with a knowledge graph.

5. Results

We conducted an evaluation of GPT-4's Named Entity Recognition (NER) capabilities on a sample dataset comprising 100 articles from Singapore Infopedia. The evaluation aimed to assess the model's performance in identifying entities relevant to the Singaporean context, including People, Places, and Organizations.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Table 1
Results

Evaluation Metric	Score
Precision	0.975
Hallucinations	1

Our findings include a precision score of 0.975 from 2,066 true positives and 52 false positives and 1 instance of a hallucination, where GPT-4 returned a contextually plausible but not explicitly mentioned entity. The chosen methodology led to the exclusion of false negative assessments in this evaluation. In our context, false negatives refer to entities that were missed by GPT-4. The primary objective of this implementation was to enhance discovery by adding entities, favoring accuracy and the avoidance of incorrect inclusions (hallucinations) over comprehensiveness. This approach was motivated by a desire to reduce manual labor, focusing instead on the accurate identification of entities without extensively verifying false negatives.

6. Analysis

The evaluation results demonstrate GPT-4's effectiveness in identifying named entities within Singapore Infopedia articles. However, the presence of hallucinations and false positives highlighted the need for further refinement. Strategies to mitigate these issues may include refining prompting techniques and incorporating post-processing steps. Moreover, manual verification of the model's predictions is labor-intensive. Therefore, it is crucial to

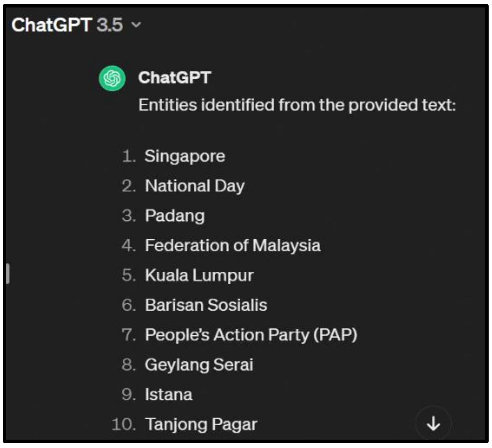
https://doi.org/10.23106/dcmi.952445840

acknowledge the limitations of this study, conducted on a relatively small dataset. Generalizability to larger datasets or different domains may vary. Additionally, further exploration of entity resolution techniques could help reduce the need for disambiguation but may introduce false positives.

Having established the results though, we must admit that the responses using GPT-4 or similar will not necessarily be reproducible, as the language model is a black box to its users. Also, with a constantly evolving algorithm the results are not predictable. In that regard, any work that relies on a language model carries a certain risk and studies that rely on evidence-based results will have limited validity since they cannot be verified by third parties, or even the same team, using the same parameters on a different date.

A relevant observation by the team was that using the later, more "advanced" models yielded better results. When cursory testing GPT-3.5 vs GPT-4 for this task using the same prompts, we saw better categorization and adherence to the instructions given with the latter, while the earlier model failed to output data in the prompted way.

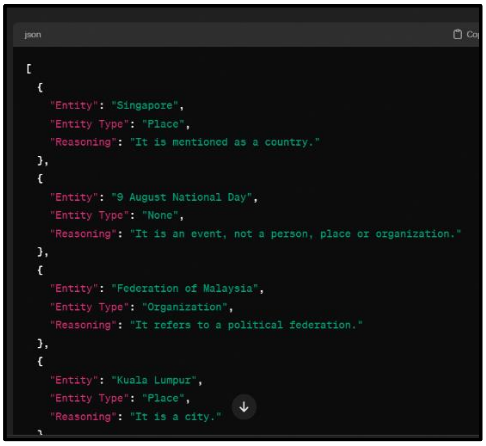
There is also the possibility that Singapore Infopedia is included in GPT-4's training data, particularly through web-crawled text. While this inclusion may enhance the model's accuracy for Named Entity Recognition (NER) of Singapore-related entities, it also introduces the risk of bias or overfitting, potentially limiting the model's adaptability to new datasets.



ChatGPT

Entities identified from the provided text:

1. Singapore
2. National Day
3. Padang
4. Federation of Malaysia
5. Kuala Lumpur
6. Barisan Sosialis
7. People's Action Party (PAP)
8. Geylang Serai
9. Istana
10. Tanjong Pagar



```json
[
  {
    "Entity": "Singapore",
    "Entity Type": "Place",
    "Reasoning": "It is mentioned as a country."
  },
  {
    "Entity": "9 August National Day",
    "Entity Type": "None",
    "Reasoning": "It is an event, not a person, place or organization."
  },
  {
    "Entity": "Federation of Malaysia",
    "Entity Type": "Organization",
    "Reasoning": "It refers to a political federation."
  },
  {
    "Entity": "Kuala Lumpur",
    "Entity Type": "Place",
    "Reasoning": "It is a city."
```

Figure 8: Comparison of output request for JSON between GPT-3.5 (left) and GPT-4 (right). The left image shows the output request for JSON not being followed, while the right image demonstrates GPT-4 following the format.

The focus here was primarily on adding additional entities as keywords - differentiated from subjects through the "schema:mentions" property.

The approach was to not introduce any erroneous data while comprehensiveness was secondary. If a specific article had 5 potential entities, the team decided to rather add only 2 entities into the metadata description in the knowledge graph, that we are certain are accurate, than adding all possible 5 with a drop in certainty. As such, missed entities were

https://doi.org/10.23106/dcmi.952445840

accepted to support the automation of the process without the need to review the results after the initial exploration.

7. Conclusion

As this exploration has shown, the application of GenAI carries some advantages to libraries and specifically the metadata work they do. LLMs can be utilized to identify, categorize, and extract entities within a text with some success. Specifically, metadata enrichment can be automated using GPT-4 with a reasonable amount of precision. In our given scenario, adding metadata that otherwise would be unavailable, became a strong use case.

Due to the complexity of the task, the team did not explore subject indexing, an effort usually undertaken by a subject matter expert in a given area, following specific rules. The National Library Board (NLB) uses Library of Congress Subject Headings (LCSH) for this task. Previous preliminary exploration indicated that generating LCSH unsupervised using LLMs carries too many risks due to inaccuracies, which are likely based on the complexity of rules governing LCSH. A future study on how to assign subjects with the help of GenAI tools may establish efficiencies here, considering the resources required vs the efforts saved in a process flow.

References

[1] National Library Board Singapore. Infopedia. Retrieved from National Library Online: https://www.nlb.gov.sg/main/onesearch/result?type=infopedia&nlonline=true

[2] National Library Board Singapore. National Library Online. Retrieved from https://www.nlb.gov.sg/main/onesearch/result?nlonline=true

[3] Dresel, R. (2024, January 31). Designing a Linked Data Service across borders and timezones: the National Library Board's experience. National Library Board, Singapore. DOI: 10.23106/dcmi.953335608. Retrieved from https://dcpapers.dublincore.org/files/articles/953335608/dcmi-953335608.pdf

[4] OPEN AI. (2023, March 15). Retrieved from arXiv: https://arxiv.org/abs/2303.08774

[5] Goh, R. (2018). Using Named Entity Recognition for Automatic Indexing. IFLA WLIC 2018 – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies in Session 115 - Subject Analysis and Access. Retrieved from https://library.ifla.org/id/eprint/2214/1/115-goh-en.pdf

[6] Puay Eng, T., Hong, G., & Jailani, H. (2018). "Authoritative Content to Build Trust in an Age of Information Overload: The National Library Board of Singapore's Experience." Malaysia: IFLA.

[7] EE, M. H. (2019). Mining Text, Linking Entities – NLB's Journey. IFLA WLIC 2019 - Athens, Greece - Libraries: dialogue for change in Session 114 - Knowledge Management with Information Technology and Big Data. Athens: IFLA. Retrieved from https://library.ifla.org/id/eprint/2448

[8] Nasar, Z., Jaffry, S. W., & Malik, M. K. (2021). Named entity recognition and relation extraction: State-of-the-art. ACM Computing Surveys (CSUR), 54(1), 1-39.

[9] Tjong Kim Sang, E. F., & De Meulder, F. (2003). "Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition." Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4, 142-147. https://doi.org/10.3115/1119176.1119195

[10] McCallum, A., & Li, W. (2003). "Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons." Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, 4, 188-191.

https://doi.org/10.23106/dcmi.952445840

[11] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). "Language Models are Few-Shot Learners." arXiv preprint arXiv:2005.14165.

[12] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." In Advances in Neural Information Processing Systems 35 (NeurIPS 2022) Main Conference Track.

[13] Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., ... & Hu, X. (2023). Harnessing the power of llms in practice: A survey on chatgpt and beyond. ACM Transactions on Knowledge Discovery from Data.

[14] Ashok, D., & Lipton, Z. C.D. Ashok, Z. C. Lipton, PromptNER: Prompting For Named Entity Recognition, Carnegie Mellon University, 2023. doi:10.48550/arXiv.2305.15444

[15] Qin, J., & Yu, B. (2023). Tutorial on Prompt Engineering and Metadata [Presentation slides]. Syracuse University. Retrieved from https://github.com/scienceIQ/DCMI2023-tutorial-prompt

[16] National Library Board. (2023). NLB Name Authorities. Retrieved from https://beta.data.gov.sg/collections/1476/datasets/d_a6d93a061d4148fe55106a28ddd6193c/view

https://doi.org/10.23106/dcmi.952445840