An LLM Based Method for Domain Specific Mapping of Metadata Terms to a Thesaurus

Mahiro Irie[1,*] and Mitsuharu Nagamori[2,†]

[1] *Master's Programs in Informatics, University of Tsukuba, 1-2, Kasuga, Tsukuba-shi, Ibaraki, Japan*
[2] *Faculty of Library, Information and Media Science, University of Tsukuba,1-2, Kasuga,Tsukuba-shi,Ibaraki,Japan*

Abstract

Metadata terms with high interoperability may assume different roles from their labels depending on the context in which they are used. However, due to insufficient definitions of vocabularies and relationships between terms, metadata schema designers with less knowledge in metadata find it challenging to identify interoperable terms from the domain-specific vocabularies used. This study proposes a method to map interoperable metadata terms to the concepts of a thesaurus that represent the roles of terms in specific contexts employing the Large Language Models (LLMs). By combining interoperable metadata terms with the domains in which they are used, it is possible to associate them with words that represent their roles. Without extensive metadata knowledge, users are expected to discover more interoperable terms using this approach through a thesaurus-based search.

Keywords

Metadata Schema, Metadata Term, Linked Open Data, LLM, Interoperability

1. Introduction

When designing the schema, it is important to select and reuse metadata terms from existing vocabularies due to interoperation of metadata and crosscutting use of LOD dataset. However, it is hard to search suitable terms for metadata schema designers because they need expertise and experience about metadata. One of the causes is that the words the designers recall and the words used in the vocabulary definition of the terms aren't necessarily the same. Interoperable metadata terms have different roles from their labels depending on the context, but the relations between the terms themselves and the roles are not defined, so the designers cannot find interoperable terms from their words directly.

In contrast, we thought that we can associate the interoperable terms with concepts which represent roles in the specific domains the terms are used by combining the terms and the domains. We propose a method that maps the terms and a thesaurus to suit the role of the terms in particular contexts. This allows the designers to explore via the thesaurus and find interoperable terms by following the mapping. We constructed a mapping system applying the method and evaluated the mapping accuracy by manual judging.

* Corresponding author. † These authors contributed equally.
irie.mahiro.qg@alumni.tsukuba.ac.jp(M. Irie); nagamori@slis.tsukuba.ac.jp (M.Nagamori)
0009-0008-5676-0638 (M.Irie); 0000-0002-9545-7825 (M.Nagamori)

https://doi.org/10.23106/dcmi.952406367

2. Metadata Terms Search

2.1. Current Search Method

LOD dataset has different schema depending on the purpose of its creators. The definition of this structure is a metadata schema. Due to improve the interoperability of the dataset, the metadata terms used should be selected from existing metadata vocabularies[1], and it is required to search for the terms that suits the designers' purposes from the existing vocabularies. In this paper, "metadata term" is shortened to "term".

Many existing metadata terms are published on the website of each organization scattered, and they are not stored and published together somewhere. Therefore, without knowledge or experience about metadata, it is difficult to search vocabulary definitions that define requirements of the terms on the network. The vocabulary definitions don't have as much information as web documents, so it is hard to find them through web search engine such as Google, and there is no established method to search the terms currently[1]. As general terms search means, there are the systems such as Linked Open Vocabularies (LOV)[2] and Zazuko Prefix Server[3]. These are search systems of the terms by entering some keywords like a web search engine. Commonly, these systems highly depend on the keyword because users don't find the search result if the keyword doesn't hit the vocabulary definitions of the terms in a full-text search, even with similar terms. There are also means of finding the terms from an existing LOD dataset that has the specific instance the designer wants to describe, or after discovering a vocabulary suitable for what the designer wants to describe, but they can only be done by those who familiar with metadata because these require knowledge and experience with metadata vocabularies.

2.2. Metadata Terms Search Issues

Although there are means of terms search, it is hard to identify all potential terms that meet the schema designer's purposes. The following are some possible reasons.

First, vocabulary definitions are insufficient. There is no general guideline for describing vocabulary definitions, and the quality and quantity of them varies depending on the vocabulary regardless of its interoperability. For example, schema:Book, a class representing a book defined by Schema.org[4], a vocabulary that can be used in any field, has the label "Book" and the definition "A book." and there is no new information beyond the label in the sentence. Such terms are difficult to hit in a search, and even if they do, it is hard to judge whether they are appropriate.

Relationships between the terms are also insufficient. Among the 3012 terms registered as classes in the LOV (with rdfs:Class as the object of the property rdf:type), the number that define a superordinate/subordinate relation with terms in different vocabularies (connected to classes in different vocabularies by the property rdfs:subClassOf) is only 519 (as of August 17th, 2023), and the number of connections between different vocabularies is small. Therefore, it is also hard to discover by tracing the relationship between terms.

These factors make terms search difficult when the words recalled by the schema designers don't match the words used in the vocabulary definitions. In particular, it is difficult to find the terms of interoperable vocabulary, such as DCMES[5] defined by DCMI[6], and Schema.org, that can be used commonly in various domains. In this paper, the domain (i.e. rdfs:domain) of

https://doi.org/10.23106/dcmi.952406367

the property is used as contextual information of the property. These interoperable vocabularies are defined to allow cross-cutting descriptions to express concepts from various domains. For example, terms representing “creator” and “contributor” such as dc:creator and dc:contributor in DCMES have the role of “director” or “filmmaker” when the domain is movie, and “author” or “writer” when the domain is book. In this case, terms representing "creator" and "contributor" can be regarded as having a cross-cutting nature. As Figure 1, there is no connection and a gap between these cross-cutting terms and the concepts represented by the words used in a specific domain.

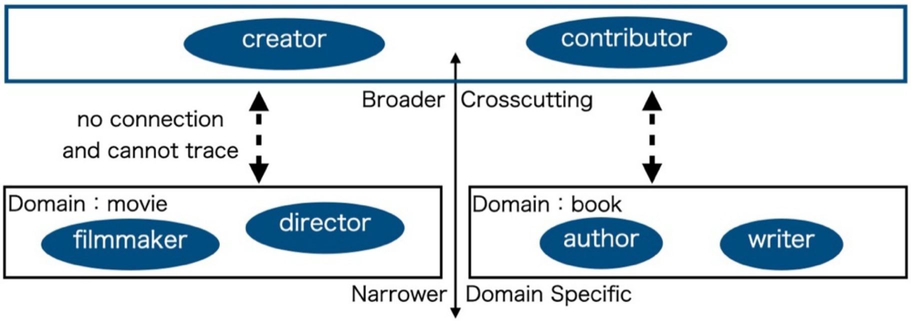

Figure 1: The relation between concepts used cross-cuttingly and in specific domains.

If the designers have some knowledge or experience of metadata, they can list terms represent “creator” and “contributor” as options for terms which can represent “author” and “director”, but if they don’t, they will have to search for the terms from the word “author” and “director”. However, it is challenging to find interoperable terms used in a wide range of domains from words used in a specific context because there are no hits when searching for keywords not in the vocabulary definitions using the existing systems, and the superordinate/subordinate relations between terms are not defined and there are the gaps.

From the above, it is important to use highly interoperable terms for describing schema from the perspective of metadata interoperability, but when the words recalled by the designers don’t match the words used in the metadata vocabularies, due to “insufficient vocabulary definitions” or “insufficient relationship among terms”, it is hard to find and list all the potential terms that suit the designer’s purposes by existing systems.

3. Related Research

There are studies of Muroi[7] and Akiba[8] as researches about metadata terms search. These studies used an approach to recommend terms and refer to related to a term found by keyword search under the opinion that terms search have limit by keyword search only.

The number of connections between the similar terms in different vocabularies have increased and it has become possible to search terms by tracing the connections owing to these studies. However, these methods don’t solve the problem of being unable to find the terms due to a lack of match between the words recalled by schema designers and the words used in the terms. Our study aims to enable schema designers to find interoperable terms from the words they recall by combining the terms with the domain they are used and relating them to concepts in a thesaurus that describe the role of the terms in the domain.

Moreover, There is a study of Mariana[9] about metadata terms search. This study proposes a method about development of Dublin Core Application Profile(DCAP) for metadata schema

https://doi.org/10.23106/dcmi.952406367

creation. The development of DCAP is complex and its guideline don't include a specific design method or standard. She mentioned that terms search and RDF description must be done by an integrator who has knowledge of metadata design.

A research field related to our study is ontology mapping. Lack of connection between terms in different vocabularies is the cause of the terms search problem. In our study of mapping terms to a thesaurus, we referred to the idea of ontology mapping. There is a study of Parundekar[10] about ontology mapping related to our study. This study proposed a method to map ontologies between different Linked Data to improve an interoperability and a usefulness of Linked Data. While this study maps between ontologies, our study maps between metadata terms and thesaurus. The major difference is the use of a thesaurus.

4. Domain Based Term Mapping Method

One of the challenges of terms search is finding interoperable terms from the words recalled by the metadata schema designers. We suspect that the reason of it is that the labels for cross-cutting terms don't match the words that describe the term's role in a specific domain, and the relationship between them is not defined, so there is a gap between both concepts.

In contrast, we considered that by combining interoperable terms with their domains that represent the context in which the terms are used, it may be possible to associate the terms with specific concepts that represent the role of the terms in that context and to find the terms across the gap. Therefore, in this study, we propose a method to combine metadata terms and domains and map them to a thesaurus. In this paper, the term “thesaurus” is used as structured vocabulary dictionary for language research or natural language processing, such as WordNet[11] and Wiktionary[12]. This method maps interoperable terms to a thesaurus by inferring words that represent the role of the terms in specific contexts from the combination of interoperable terms and the domains in which the terms are used, and replacing them with the concepts in the thesaurus that they represent. This allows the terms to be associated with concepts in the thesaurus that represent the role of them. By using the thesaurus as a window to explore terms, it is considered that schema designers will be able to find more interoperable terms that express some concepts by tracing connections from the designer's words, as shown in Figure 2, even if they don't know the words used in the metadata vocabularies. This method used WordNet as a mapping destination of the terms.

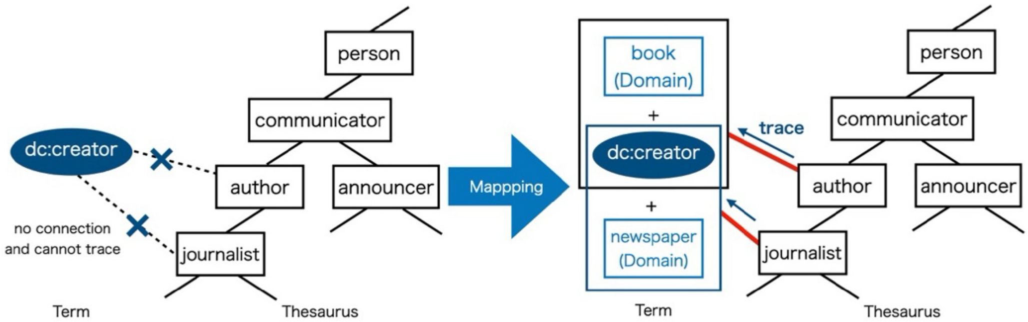

Figure 2: Overview of the method of mapping terms and thesaurus

To use the domain of the term as contextual information, only the properties in the terms of the interoperable vocabularies are subject to mapping in this paper. The domain to be combined with the term is obtained using the LOD Cloud Cache [13], a service that allows cross-searching of multiple LOD datasets, to obtain the class that is actually used as the domain of the term to

https://doi.org/10.23106/dcmi.952406367

be mapped to each term. As a standard to obtain them, the domain words list is manually created in advance, and the domain word and the term pair is selected as the mapping subject if the label of the obtained class is in the list. The number of terms mapped is 76 and the number of pairs of terms and domain words is 1139 in this study.

5. Implementation of Mapping System

We constructed a metadata term and thesaurus mapping system to realize the method. The flow of mapping by the system is shown in Figure 3. The labels and definitions of metadata terms and words representing the domain of the terms are given as input, and the LLM outputs words representing the role of the terms in specific domains, which are then replaced by the thesaurus concepts. The input data consists of term and domain information. For the information on terms, a dump file was obtained from the LOV where the term definitions were compiled, and the URIs, labels and definitions of the terms to be mapped were extracted. Mappings indicating the correspondence between the terms and the concepts in the thesaurus obtained by the system are described and output in RDF format.

Mapping by the system consists of four major steps.

(1) Replace term with the same concept (resource) in Wikidata [14]
(2) Generate multiple word options with Fill-mask
(3) Select more appropriate word options with Zero-shot Classification
(4) Replace selected words with WordNet concepts

First, in (1), the input term is replaced by the same concept in Wikidata. If the term definitions are used as they are in the prompts, the output of the LLM are likely to be highly influenced by the quality of the vocabulary definitions. On the other hand, the definition of Wikidata resources (hereafter 'Wikidata definition') is basically written in the form 'noun + modifier'. To describe the prompts in a certain format, the term is replaced by a Wikidata resource that represents the same concept as the term. The Wikidata resource data was obtained from a dump file published by Wikidata, and the required data such as labels and definitions were extracted and stored in the full-text search engine Elasticsearch[15]. The

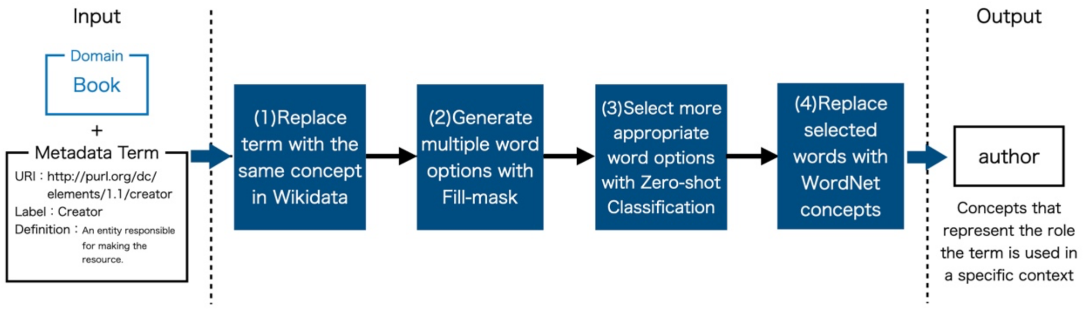

Figure 3: Mapping flow through the system

term is replaced by an Elasticsearch search for a Wikidata resource with a label identical to the label of the term. As an exception, if there is no Wikidata resource corresponding to a term, a term definition is used formatted in a form suitable for the prompt instead of it.

Then, in (2), several potential words to be mapped are generated using the natural language processing task 'Fill-mask'. It should be noted that Fill-mask in this system can't output multiple

https://doi.org/10.23106/dcmi.952406367

words. In (3), words are classified using the natural language processing task 'Zero-shot Classification' to select more suitable words as a mapping destination. The prompts required for Fill-mask and Zero-shot Classification are automatically created using the Wikidata definition obtained in (1), the term labels and the domain words.

Finally, in (4), the sorted words are replaced by concepts in the WordNet. The Python library NLTK, used in natural language processing, is used to replace all WordNet concepts (Synset) with exact matches between words and labels. This four-step process results in a mapping between terms and thesaurus concepts.

6. Experiment

To evaluate the accuracy of the mapping using this method, we conducted an experiment. Due to the lack of quantitative standards for the evaluation of the mapping, it was manually determined whether the mapping was correct or not. However, the total number of mappings output by the system is very large and it would be costly to evaluate them all, so random sampling of the resulting mappings was used for the evaluation.

6.1. An Experimental Method

The experiment was conducted with the following steps.

1. Create mappings with this system
2. Determine the sample size
3. Extract samples from all mappings by random sampling with the sample size determined in step2
4. Judge whether each mapping is correct or not in the samples
5. Calculate the rating index of the sample mapping and indicate the range in which the index fits in the population, using the confidence interval estimation.

For Step 1, not only the mappings judged to be correct by the system, but also those judged to be incorrect are output for evaluation. Mappings judged to be incorrect by the system are pairs of words obtained by Fill-mask that fail to meet the word selection standards in the Zero-shot Classification, as well as the combination of term and domain. The total number of mappings output by the system is 10517. It should be noted that the final mapping destination in this method is WordNet, but in this experiment, the pairs of words output by the LLM and the input terms and domains are treated as mappings before they are replaced by WordNet concepts, and the validity of the connection is evaluated.

For step 4, the correctness of the mapping was manually judged. In order to avoid making a judgement based on the subjective judgement of one specific person, several people, including teachers with knowledge of metadata, made a correct or incorrect judgement of the mapping with reference to the following judgement standards. All of these standards are qualitative and not quantitative.

- The concept represented by the term and the concept represented by the word in the output are in an inclusive relation.
  - ➢ If the output word is equal to the label of the term, it is a correct mapping.

https://doi.org/10.23106/dcmi.952406367

➢ The term is not strictly considered to be an inclusive relation.

- There is a demand to discover terms from output words.
- The output word is appropriate to the definition of the term.
- The combination of the output word and domain is appropriate.
- The output word is appropriate for the purpose of the vocabulary use.

For step 5, the precision, recall and F-measure(F1-score) are used as evaluation indicators for the mapping. In addition, the indicators follow a binomial distribution, in order to determine only whether the mapping is correct or not in this study. At a 95% confidence interval, with a 5% margin of error, the minimum sample size required is 385. Therefore, the sample size in this experiment was determined in step 2 at 385.

6.2. Result

In the confusion matrix of the results shown in Table 1, positive/negative mappings output by the system are 'prediction' and determined manually are 'Correct'. For each of the "Prediction" and "correct", True mappings are "Positive" and False mappings are "Negative".

The range of indicators across the sampled mappings and the estimated mappings as a whole, calculated from the above results, is shown in Table 2. The estimated population proportion is p.

Table 1

Confusion matrix of sampling results

Correct / Prediction	Positive	Negative
Positive	TP:94	FP:116
Negative	FN:19	TN:156

Table 2

Indicators of sampled mappings and estimated range of indicators for overall mapping

Precision	Recall	F-value	Estimated: Range of Precision	Estimated: Range of Recall	Estimated: Range of F-value
0.448	0.832	0.582	$0.398 \leq p \leq 0.498$	$0.795 \leq p \leq 0.869$	$0.533 \leq p \leq 0.631$

7. Discussion

The evaluation indicators such as precision and recall calculated by the experiments and the output mappings are discussed, and improvement of the method are described.

(1) Discussion about the Precision and Recall

In Table 1, the ratio of FN relative to TP is low and the system is generally correct to judge the actual correct mapping. The precision is low relative to the recall, but this is due to the high FP in Table 1. This may have been influenced by the fact that the word selection standard for Zero-shot Classification, referred to in Chapter 5, intentionally set the standard to output many positive examples. Although this setting was made focusing on increasing the number of

https://doi.org/10.23106/dcmi.952406367

appropriate mappings as much as possible, many inappropriate mappings would make it difficult to search for the terms. There is room for improvement in this method and system, and there is the possibility to improve the precision and recall.

(2) Discussion about the output mapping results

Then, the output mappings are analyzed and the characteristics of the mappings and the causes for which the mappings were outputted are discussed.

First, examples of actual correct mappings that the system was able to predict as correct (corresponding to TP in Table 1) are analyzed. The correct mappings can be mainly classified into three categories. Table 3 shows an example of each mapping.

(A) it paraphrases a subordinate concept in a specific domain
(B) the domain is irrelevant and simply paraphrases the concept expressing the term to another concept
(C) the same words are output

Table 3
Example and Classification of Correct Mapping

Term	Domain word	Output word	Classification	Sampled Number
http://schema.org/image	music album	artwork	(A)	19
http://www.w3.org/2000/01/rdf-schema#seeAlso	animal	reference	(B)	40
http://xmlns.com/foaf/0.1/maker	creative work	maker	(C)	30

(A) is a mapping connecting concepts in a specific domain with interoperable terms representing their superordinate concepts, that the domain may have influenced the output word. This is the most useful mapping that meets the purpose of this study. (B) is a mapping connecting other concepts related to the term, where the domain is not considered to influenced the output word. This mapping is also considered useful for searching terms, as it increases the number of connections to the interoperable terms. (C) is a mapping where the labels of the term and the output word match, and this mapping is not so significant.

Next, examples of incorrect mappings that were actually predicted as correct by the system (corresponding to the FP in Table 1) are analyzed. The incorrect mappings can be classified into four categories. Table 4 shows an example of each mapping.

(D) it influenced by one specific word unrelated to the term in the input
(E) it paraphrased a concept but not appropriate to the domain
(F) it was output of a word that suits the context and is relatively close to the meaning of the term, but was not appropriate as a paraphrase
(G) it departed from both the meaning of the term and the domain

https://doi.org/10.23106/dcmi.952406367

Table 4
Example and Classification of Incorrect Mapping

Term	Domain Word	Wikidata definition	Output word	Classificat ion	Sampled Number
http://purl.org/dc/terms/hasVersion	image	form or variant of a type or original	original	(D)	12
http://schema.org/creator	language	person who does creative work	author	(E)	12
http://schema.org/height	product	distance between the lowest end and highest end of an object	radius	(F)	36
http://schema.org/description	county	description of something, written down or verbal	adjective	(G)	56

(D) is a mapping where a word in the Wikidata definition, which is not directly related to the concept represented by the term, has a big influence. (E) is a mapping where the relationship between the term and the output word is an inclusive relationship, but the combination with the domain is not appropriate. In the example in Table 4, combining the term schema:creator with the domain word "language", appropriate output words could be "linguist" or "developer"(for programming languages). However, the connection between schema:creator and "author" is appropriate for the domain "book", but not for "language". (F) is a mapping where the concept represented by the term and by the output word are used in similar context but shouldn't be connected. (G) is a mapping where the concept represented by the term and by the output word are so different each other.

There were many examples of (F) and (G), but there are two possible reasons why such mappings were output. First, the input sentence sometimes directly affected the output, as seen when similar words influenced the output, like 'published' leading to 'publisher', although not identical. However, there were by no means many examples of incorrect mappings where the input sentence seemed to have had a direct influence, and the main cause was considered to be the influence of the LLM. Due to Fill-mask, outputs included not only suitable word choices but also alternatives like 'following', 'latter', and 'source', leading to higher scores in Zero-shot Classification despite their limited relevance to the term's role.

(3) Discussion and improvements about the overall method

From the discussion, there are several points where the mapping system could be improved.

Improvement of prompts to enter LLMs: Not only the term labels and definition, but also other information related to the term, such as co-occurring terms and instances actually used in the dataset, could improve the system by giving the input other information related to the term. However, the number of terms used by the dataset is limited and it is difficult to use information from the dataset for all terms.

Improvement of LLMs themselves: The models DistilRoBERTa-base [16] and BART-large-mnli [17] were used in this study, but other models exist that can carry out Fill-mask and Zero-shot classification. It is expected that trends in output words will differ depending on the model, so changing the model could improve the mapping. In addition, although the words were

https://doi.org/10.23106/dcmi.952406367

selected using Zero-shot Classification without learning, the mapping could potentially be improved by training the model based on the data obtained in this study.

Improvement of the overall mapping method: Although WordNet was used in this study as the thesaurus to map to, this method is not limited to WordNet, but can be connected to various thesauruses. There is no need to limit the mapping to a single destination, and mapping in combination with other thesauruses could also be considered. It is thought that by combining WordNet with other thesauruses or databases that have detailed concepts that WordNet doesn't have, more words could be used for term search.

In this method, the mapping was judged and evaluated on two scales, True or False, as to whether the mapping was appropriate or not. However, there are cases where it is difficult to judge whether the mapping is not wrong, even though there is a more appropriate destination, it is possible that it was not appropriate to judge based on two scales of True and False. A better way of evaluating mapping could consider quantifying how appropriate the mapping is, using the scores output by the LLMs.

This system cannot map to all terms because the words that can be output and the terms that can be used as input are limited. For example, the Fill-mask task in the model used in this study can only output one word corresponding to the masked part, and the method cannot map a class used cross-cutting a field to a subordinate class in a specific context.

8. Conclusion

It is important to select and reuse terms from the existing metadata vocabulary that meet the designer's purposes in schema design for cross-cutting use of LOD datasets. One of the challenges of term search is the impossibility of finding interoperable terms due to the lack of match between the words recalled by schema designers and those used in the metadata vocabulary. This paper proposed a method to combine metadata terms and their domain of context and map them to a thesaurus. By mapping terms to a thesaurus containing commonly used words, it is expected to find interoperable terms through the thesaurus. Due to the large number of terms and the very high cost of manual mapping, the mapping was done mechanically using LLM.

Experiments to evaluate the accuracy of the mappings output by the system showed that while the system using LLM was able to output a certain number of mappings between concepts in a specific domain and interoperable terms that represent superordinate concepts, the accuracy of these mappings was not high. The cause of the incorrect mapping output is thought that a combination of factors such as the quality of the prompts, the quality of the LLMs and the structure of the system may have contributed to the incorrect mapping. There is still room for improvement in the evaluation method.

In the future, we will review the system design, including the validation of the optimal LLMs and prompts, to improve and extend the functionality of the mapping system. Besides improving the accuracy of the mapping, the aim is to enable the mapping of classes with high interoperability and mapping with concepts represented by multiple words, which could not be achieved in this study. It is also necessary to evaluate the effectiveness of the mappings on term search through subject experiments. How influenced the mapping obtained by this system is

https://doi.org/10.23106/dcmi.952406367

on term search, including the validity of the information presentation items, including the UI, will also be examined.

Acknowledgements

This study was supported by JSPS KAKENHI Grant Number 24K15651.

References

[1] Tom Heath, and Christian Bizer. Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool. 2011.
[2] Linked Open Vocabularies. URL: https://lov.linkeddata.es/dataset/lov/.
[3] Zazuko Prefix Server. https://prefix.zazuko.com/.
[4] Schema.org. https://schema.org/.
[5] DCMI. "Dublin Core™ Metadata Element Set Version 1.1: Reference Description". URL: https://www.dublincore.org/specifications/dublin-core/dces/.
[6] DCMI. DCMI. URL: https://www.dublincore.org/.
[7] A. Muroi. "A Method of Assisting in Exploration of Metadata Terms Through Presentation and Comparison of Relevant Terms". Graduation Thesis. University of Tsukuba, Ibaraki, Japan. 2022.
[8] R. Akiba. "Metadata Terms Select Assistance Based on Class Structure for Integrated Use of Web API and LOD". Graduation Thesis. University of Tsukuba, Ibaraki, Japan. 2020.
[9] Mariana Curado Malta, Ana Alice Baptista. "A Method for the Development of Dublin Core Application Profiles (Me4DCAP V0.2): Detailed Description". Proceedings of the DCMI International Conference on Dublin Core and Metadata Applications 2013, 2013.
[10] Rahul Parundekar, Craig A. Knoblock, Jose Luis Ambite. "Linking and Building Ontologies of Linked Data". Proceedings of the 9th International Semantic Web Conference, 2010.
[11] Princeton University. "WordNet". https://wordnet.princeton.edu/.
[12] Wiktionary. https://www.wiktionary.org/.
[13] OpenLink Software, "LOD Cloud Cache". https://lod.openlinksw.com/fct/.
[14] Wikidata. https://www.wikidata.org/.
[15] Elastic Co. "Elasticsearch". https://www.elastic.co/jp/elasticsearch.
[16] Hugging Face Inc. "distilroberta-base". https://huggingface.co/distilroberta-base.
[17] Hugging Face Inc. "facebook/bart-large-mnli". https://huggingface.co/facebook/bart-large-mnli.

A. Domain Words List

This is the domain words list used for mapping, referred to in Chapter 4.

person, organization, place, creativework, item, event, dataset, document, class, property, ontology, image, website, product, online account, program, record, book, webpage, article, newspaper, tv series, musicalbum, video, movie, video game, painting, comic, photograph, bibliographic resource, publication, text, track, town, village, business, county, concept, softwaresourcecode, animal, city, company, software, language, tvseries, country, videoobject, imageobject

https://doi.org/10.23106/dcmi.952406367

B. Prompts and Word Select Standards

Table 5 shows the prompts used as input sentences in each task in our system, examples (for the term dc:creator, and domain word "book") and the standards used to select the words to be mapped in each task.

Table 5

Prompts given to LLMs and standards for scores output by LLMs

	Fill-mask	Zero-shot Classification
Prompts	The [Labels of term] in the context of [Domain Word] is the [MASK]. It is the [Wikidata definition].	It is the [Wikidata definition] in the context of [Domain Word].
Prompts Example	The creator in the context of book is the [MASK]. It is the person who does creative work.	It is the person who does creative work in the context of book.
Standard	All words scored 0.01 or above (excluding words that match domain word).	Half of the number of input labels (rounded up to the nearest whole number) are selected in order of increasing score. However, the score of the selected words must be at least 0.30, and words with a score of less than 0.30 are not selected. If the score is 0.80 or higher, the words are selected regardless of the number of input labels.

For [Labels of term] in the prompt, the uppercase letters are converted to lowercase in advance, and the words are split and stop words are removed. In addition, [MASK] in the prompt indicates the masked part in Fill-mask, and the one word that fills this [MASK] is predicted. Both prompts contain 'in the context of [Domain Word]', with the intention to reflect the meaning of the domain.

C. Example of Mapping Judgement

This is an example of output mappings and correct/incorrect judgement.

Term: http://schema.org/creator
Label: creator
Definition of the term: The creator/author of this CreativeWork. This is the same as the Author property for CreativeWork.
Wikidata Definition: person who does creative work
Domain word: painting
Output word: artist
Judgement of the system: True
Judgement of human: True

https://doi.org/10.23106/dcmi.952406367