Rule-based Topic Classification Scheme for Zhihu Questions using Wikidata Ontology

Alyssa Shuang Sha[0000-0001-6368-3388] and Armin Haller[0000-0003-3425-0780]

Australian National University, Canberra, Australia

{alyssa.sha,armin.haller}@anu.edu.au

Abstract. We propose a novel rule-based approach[1] to achieve the hierarchical topic classification for content from different platforms by utilizing Wikidata ontology. More specifically, we conduct an experiment on Zhihu (the largest Chinese Q&A platform). By mapping the labels of the questions from Zhihu to Wikidata entities, we can rely on the taxonomic relations from Wikidata to categorize the question topics in a coarse-to-fine manner according to their domain coverage. The proposed topic classification scheme can be applied to classify the items with Wikidata identifiers from different platforms uniformly. This research also sheds light on how the structured knowledge base can help social media or knowledge-sharing platforms to build a more standardized content classification system.

Keywords: Topic classification, Wikidata, Entity Linking, Q&A services.

1 Introduction

Online Question and Answering services (Q&As) are part of a leading group of influential and popular Web 2.0 applications that millions of people use daily to ask and answer questions [7]. These services have also been developed into powerful platforms for creating, sharing, and acquiring a massive amount of human knowledge. Topic classification is a crucial task in Q&As as the structured categorization of questions facilitates information search and knowledge retrieval by the users.

Similar to popular social media (e.g., Twitter, Instagram), popular Q&As (e.g., StackOverflow, StackExchange, Zhihu) mainly rely on tags/hashtags/labels attached by users to achieve topic classification. Among the popular Q&As that provide a "tagging" function, t[1]here are two common methods to produce a topic structure for the questions.

Q&As like StackOverflow and StackExchange only provide a flat topic structure, where all the tags are organized at the same semantic level. Users can find the Tags relevant to the topic they are viewing. They can also browse topics based on popularity, alphabetical order, and update time. However, the tags are unstructured, meaning

---

[1] Our database and programs will be made publicly available once the article has been accepted by this workshop.

no hierarchical relationships exist among all the topics. Other Q&As attempt to build their topic structure by categorizing all the labels attached by users. Zhihu[2] produces a rooted Directed Acyclic Graph (DAG) as a topic tree by assigning parent nodes manually for newly generated labels. As shown in Fig 1, topic tags tend to be more specific from the root-to-leaf nodes. A question can be annotated by either the leaf or the internal nodes at the same time.

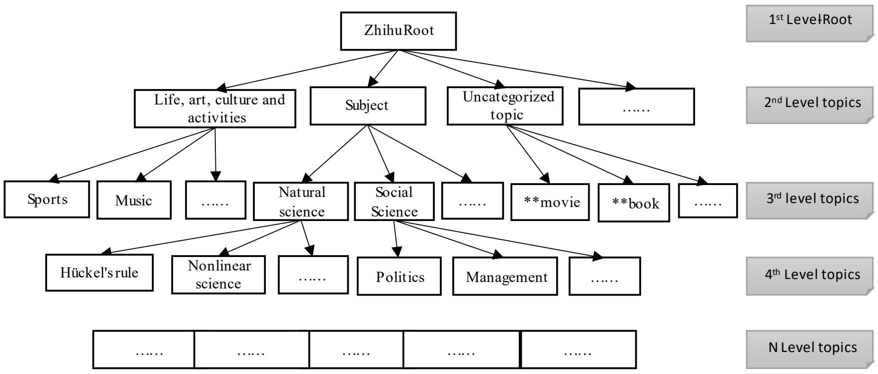

Fig. 1. An example of Zhihu topic structure

However, the topic structure that comes with Q&As like Zhihu does not allow for real-time sorting of newly added labels. Thus, many unstructured labels (represented as Uncategorized topic in Fig. 1) need to be manually categorized afterward. Besides, the topic structure in Zhihu is not rational or consistent enough, as most of the labels are generated by individual users. Thus, in this research, we attempt to utilize the taxonomic relations from the structured knowledge base to improve the standardization and automation of topic classification in Q&As.

Wikidata is the collaborative knowledge graph (KG) initiated by the Wikimedia Foundation [22]. KGs are a type of technology that can bring context and depth to various applications, including web search, product recommendations, and intelligent assistants. A KG often spans many domains and is created on top of a conceptual schema, or ontology, that specifies the types of entities (classes) and features that may be used in the graph [13].

Items and Properties are the building blocks of Wikidata[3]. Items refer to physical or abstract entities, as well as classes of entities (e.g., humans, business). Any relationship between entities is expressed using properties (e.g., part of, parent of). Alphanumeric codes as Uniform Resource Identifiers (URIs) are used to identify Items and Properties which combine a letter (q for items, p for properties) and a number. Prior research into the Wikidata ontology has identified Items that perform the role of classes using Properties indicating taxonomic relations—for example, *instance of* (P31) and *subclass of* (P279). Following a similar approach, we define the Wikidata ontology as the set of all its properties and items utilized as classes which can be seen

---

2 https://www.zhihu.com/
3 https://www.wikidata.org

as a directed acyclic graph (DAG). The following diagram (Fig. 2) shows part of the Wikidata top-level ontology supported by the *subclass of* (P279) relationship (*entity* Q35120 as the ontology root is the superclass of all items in Wikidata). Depending on the distance to *entity* Q35120, the top-level items can be classified as first-layer entities, second-layer entities, third-layer entities, etc.

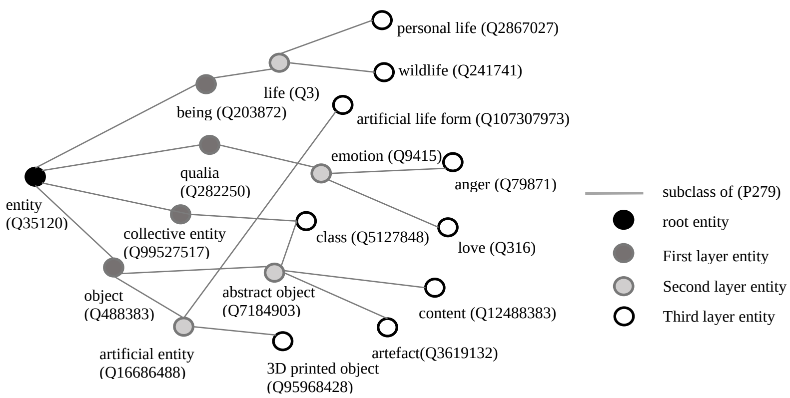

Fig. 2. Example of the top-level ontology in Wikidata

Besides the existing items, properties, and statements, Wikidata provides external identifiers used in external systems (databases, authority control files, online encyclopedias, etc.) and displayed as links in Wikidata items if a *formatter URL* (P1630) is defined. Therefore, information from other platforms can be linked to Wikidata entities and categorized following taxonomic relationships in structured KG using external identifiers. For example, as presented in Fig. 3, the questions with a Starbucks label attached are grouped by a unique Topic ID (19557474), which serves as one external identifier that can link the Zhihu Starbucks topic with Wikidata *entity* Q37158. Likewise, the information about Starbucks from other systems (Twitter, Quora, SEEK, etc.) can be mapped to the same entity and thus be classified based on the structured knowledge base from Wikidata.

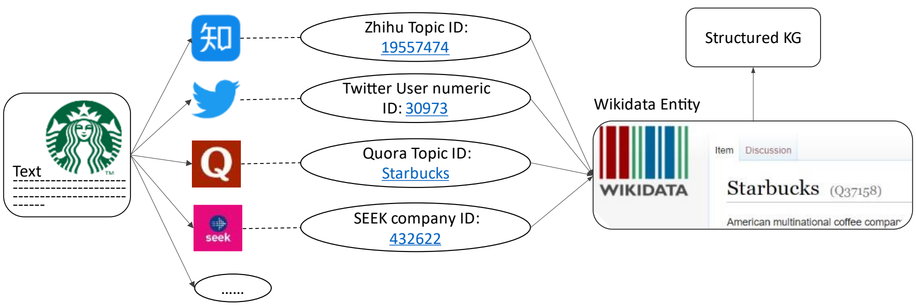

Fig. 3. An example of using Wikidata external identifiers for *entity* Q37158

The purpose of this research is to propose a rule-based topic classification method by using the external identifiers from structured KG (Wikidata). We aim to optimize the existing KG in Q&As by improving the structure and automaticity of topic classification. Our research involves a longitudinal experiment on Zhihu and provides a demonstration to show how the proposed method can assist in the operation and management of the Q&A platform by leveraging the knowledge base of Wikidata.

2 Related Work

For Q&A services, the most common way to capture questions' topics is to focus on tags/hashtags which are called labels in our study. There has been a diverse array of academic work that studies tags, most of them aiming to find relevant information by using tags to predict the popularity or information flow, or to develop clustering methods of online content [1, 6, 12, 17, 20]. Other research looks at tags themselves, trying to analyze their dynamics, popularity, semantics and engagement [1, 2, 8, 10, 19, 24]. However, the tags from many platforms are unstructured, expressed from a single language and semantically duplicated with each other, making the topics on those platforms confusingly categorized without a logical structure. Thus, bridging the gap between unstructured information and structured data is crucial.

Recent approaches that automatically link unstructured data to structured knowledge bases (e.g., Wikidata) revolve around Entity linking (EL) [3, 5, 9, 21]. Entity linking has been widely applied in natural language processing [3, 4, 11, 15, 16] on domains including news, biographical text, and movie/show plots. However, the domain of Q&A services has never been explored by previous research. Besides, most of the EL studies applied Machine learning-based approaches which mainly train neural network-based algorithms and classical machine learning models (e.g., Hidden Markov Model and Conditional Random Fields) on an annotated corpus where mentions of a particular type of entities are annotated [21, 23, 25].

Instead of using Machine learning-based approaches that sometimes appear costly and time-consuming, this research proposes a rule-based approach for topic classification to take advantage of using the built-in external identifiers from Wikidata. Furthermore, our proposed rule can assist in classifying the external topics in a coarse-to-fine manner depending on their semantic coverage by utilizing Wikidata taxonomic relations.

3 Approach

This research maps Zhihu question labels to Wikidata entities to classify topics from Q&As by using a structured knowledge base. The reasons for choosing Zhihu are because 1) it has a large user base, 2) the content features on Zhihu are transparent, and 3) the identifiers of Zhihu labels exist in Wikidata. More importantly, as discussed in the introduction, 4) the current topic structure on Zhihu can be enhanced by utilizing structured data, which provides practical value to this study. This section involves our data collection methods for Wikidata and Zhihu and our proposed rule-based mapping method that can be applied for question topic classification.

3.1 Data Collection

We first extract all the entities from Wikidata that have Zhihu Topic ID identifiers using Wikidata Query Service[4]. The query enabling the creation of a list with Wikidata entities and their corresponding Zhihu Topic IDs is shown in Fig. 4.

```
SELECT ?item ?itemLabel ?zhihuID
WHERE {
  ?item wdt:P3553 ?zhihuID.
  SERVICE wikibase:label
          { bd:serviceParam
            wikibase:language "[AUTO_LANGUAGE],en". }
}
```

Fig. 4. SPARQL query for extracting the entities with Zhihu Topic ID (P3553) identifiers from Wikidata

Note that the original names of the topics are written in Chinese on Zhihu, which are automatically translated to English during this process. There are 14,837 entities from Wikidata with Zhihu Topic ID as one identifier, and 14,836 of them are linked to the Zhihu topic page correctly. One of them (Q9715) is mistakenly labeled and thus removed from our dataset.

Using an API (api.zhihu.com) and the Python zhihu-oauth package [14], this study randomly collected 1,575 labels (approximately 5% of the total number of existing labels) from Zhihu in April 2022. In order to show a use case of our proposed topic classification method, we attempt to explore the question lifecycle patterns under different topic groups. Questions on Zhihu can be retrieved using a Question ID (QID) as a key. Based on our observation, the question IDs are always eight to nine digits long and the absolute values of the QIDs monotonically increase with time. Hence, we are able to collect popularity panel data (e.g., view numbers, votes, the amount of received answers) for questions that just posed under those labels. The question popularity data were traced 1,440 times for one week period, which will be plotted as question lifecycles for us to explore with different topic categories.

3.2 Rule-based Mapping Method

The proposed rule-based method aims to map Zhihu labels to entities from Wikidata's knowledge graph to produce a more robust and reliable topic classification framework. Specifically, the matching method can be split into two steps: top-down entity extraction and bottom-up label mapping.

---

[4] https://query.wikidata.org

Top-down entity extraction

As we introduced in Fig. 2, the entities on Wikidata can be classified into different layers according to their domain coverage. In this study, we use the existing property *subclass of* (p279) to represent the domain coverage of the entities. Formally, we describe Wikidata as a graph $G = (E, R, S)$ with $S=\{r\ (e_1, e_2) \mid r \in R, e_1, e_2 \in E\}$. In this case, if $e_1 \in G\ (e_2, p279, s)$ which means e_2 is subclass of e_1, then we consider e_1 has a broader domain coverage than e_2. Following the *subclass of* property, we can trace down from the "root" entity (Q35120) and locate the rest entities at different layers to build a hierarchical entity structure.

In order to classify the Zhihu topics into Wikidata entities with similar domain coverage, we need to select entities from one specific layer as the ideal categories for the topic classification task. We decide to extract the third layer entities (shown in Fig. 2) as our classification entries. The reason is to ensure that all categories are at the same level. The entities beyond the third layer are too numerous and specific, by using which we might lose the point of classifying Zhihu topics that are already specific. The first and second layers of labels are again too abstract, thus, it is difficult for users to have a clear understanding of their semantic meanings. Therefore, we chose the third layer labels as a prerequisite for the classification task.

The detailed steps are listed in Algorithm 1. The input is the whole graph from Wikidata with the "root" *entity* (Q35120) and *subclass of* (P279) property specified. The outputs of this algorithm are three sets L_1, L_2, L_3 containing entities for layer1, 2, and 3 respectively.

ALGORITHM 1: COLLECTING 3RD LAYER WIKIDATA ENTITIES

Input: entities E, graph $G = (E, R, S)$, $Q35120 \in E$, $P279 \in R$
Output: L_1, L_2, L_3

$L_1 \leftarrow \emptyset$, $L_2 \leftarrow \emptyset$, $L_3 \leftarrow \emptyset$
for $e \in E$ **do**
    if $Q_{35120} \in G\ (e, P279, s')$ **then**
        $L_1 \leftarrow e$
return L_1
for $e_1 \in L_1$ **do**
    for $e \in E \cup L_1^C$ **do**
        if $e_1 \in G\ (e, P279, s')$ **then**
            $L_2 \leftarrow e$
return L_2
for $e_2 \in L_2$ **do**
    for $e \in E \cup L_1^C \cup L_2^C$ **do**
        if $e_1 \in G\ (e, P279, s')$ **then**
            $L_3 \leftarrow e$
return L_1, L_2, L_3

Bottom-up topic mapping

Secondly, we apply a bottom-up approach to map Zhihu topic labels to the third layer Wikidata entities by using taxonomic relations (*subclass of* and *instance of*) properties forwardly. Zhihu topic ID as one of the Wikidata entity identifiers can be linked to the Wikidata entity URI. After data collection, we generate a set of tuples W with Zhihu identifier i and its corresponding Wikidata URI $f(i)$ using the mapping function provided by Wikidata Query Service (Fig. 4).

By running the URI in SPARQL query under the *subclass of* and *instance of* relations, we can track the entity's 'parents', 'grandparents', and so on until we reach the previously defined third layer entities. The mapping steps are listed in Algorithm 2. The inputs are the graph from Wikidata with *subclass of* (P279) and *instance of* (P31) properties specified; dataset Z that contains Zhihu topic IDs that we crawled randomly; dataset W that has all Zhihu identifiers and their corresponding entity URIs from Wikidata; L_1, L_2, L_3 that contain the 1st, 2nd and 3rd layer entities.

ALGORITHM 2: MAPPING ZHIHU TOPICS TO THE 3RD LAYER WIKIDATA ENTITIES

Input*:* entities E, graph $G = (E, R, S)$, $P279 \in R$, $P31 \in R$, Z (a set contain zhihu label ids), $W = \{(i, f(i)), i \in N\}$, L_1, L_2, L_3
Output*:* $\Gamma = \{(e_i, z_i), i \in N\}$
for z *in* Z **do**
  if $z \in w_i$ **then**
    $e \leftarrow w_{f(z)}$
    if $e \in L_1$ or $e \in L_2$ **then**
      $\Gamma_z \leftarrow (none, z)$
    else
      while $e \notin L_3$ **do**
        while $G(e, P279, s) = \emptyset$ **do**
          for $\acute{e} \in E$ **do**
            if $\acute{e} \in G(e, P31, s)$ **then**
              $e \leftarrow \acute{e}$
        for $\acute{e} \in E$ **do**
          if $\acute{e} \in G(e, P279, s)$ **then**
            $e \leftarrow \acute{e}$
      $\Gamma_z \leftarrow (e, z)$
  else
    $\Gamma_z \leftarrow \emptyset$
return Γ

During the mapping process, several special occasions were manipulated by our algorithm individually. Zhihu labels that do not have Wikidata identifiers are assigned with an empty value for their output tuple. Zhihu labels that already belong to 1st and 2nd layer entity groups are too general and abstract (e.g., "Metaphysical topics"), thus difficult to be traced down to specific topics. Therefore, we erase the value of their corresponding Wikidata entity. The Zhihu labels cannot be classified by our method in both cases above. In the progress of tracing the “parent” class of entities, if there is no “parent” class for entity e, we find the substitute $\acute{e}$ by using *instance of* property and trace $\acute{e}$’s “parent” instead. Finally, the output for this algorithm is a set of tuples: $\Gamma = \{(e_i, z_i),\ i \in N\}$, where each tuple includes the Zhihu topic id z_i and the successfully-mapped 3rd layer Wikidata entity e_i.

4 Results and Use cases

This section presents a demo of our Zhihu topic classification result using Wikidata knowledge base. A few applications in terms of utilizing the proposed classification scheme to analyze topic features and question lifecycles are also demonstrated.

4.1 Topic classification on Zhihu

After completing the mapping procedures, 581 out of 1,575 Zhihu labels are successfully classified as Wikidata third-layer entities. Table 1 shows some samples from our classified data. Not limited to the third layer entities, our proposed method enables users to choose needed topic classification depths by adjusting the parameters in Algorithm1 & 2.

Table 1. Part of the topic classification results on Zhihu

3rd layer entity	URI of 3rd layer entity	Zhihu Labels
business	Q4830453	hotel, Chicken Soup for the Soul, Patek Philippe & Co., China Unicom, NetEase Games, Starbucks, Alibaba Group, Canon Inc., Amazon, NetEase, Huawei ……
city	Q515	Guangzhou, Shenzhen, Yichang, Chengdu, Yinchuan, Hangzhou, Ogi, Chaozhou, Tianjin, Fuzhou, Shanghai……
science	Q336	analytical chemistry, psychology, medicine, physics, telecommunications engineering, geometry, puberty, economics, computer science, ophthalmology, mathematics……
service provider	Q2169973	Chang'an University, Tianjin University of Technology, Jiangsu University, Shanghai University, Capital Normal University, Zhengzhou University, Civil Aviation University of China, Peking University, China Three Gorges University, East China University of Political Science and Law, Tsinghua University……

4.2 A structural way to explore topic and question features

Previous research has tried to analyze the time sensitivity of Zhihu topics by clustering them according to their recurring frequencies [18]. This way allows users to view the lifecycle of questions under the topics with different time sensitivities. However, it is difficult to relate the topic features (e.g., time-sensitivity) to their semantic meanings without a standard topic classification scheme. By achieving the uniformed classification task, this research offers semantic value with similar domain coverage to label groups aside from the recurring strength and period.

Most questions on Q&As follow a two-stage lifecycle which includes a rapid-growing stage followed by a saturation phase. We refer to the intersection of the two phases as the “knee” point. Fig. 5 depicts the popularity and median knee point axes for part of the topics mapped to Wikidata. The third layer entities are shown in the figure's text. The statistical result shows no significant relationship between the popularity of the questions and the time it takes for the two-stage questions to reach saturation point. The time to reach the second lifecycle stage varies greatly between topics with different levels of popularity.

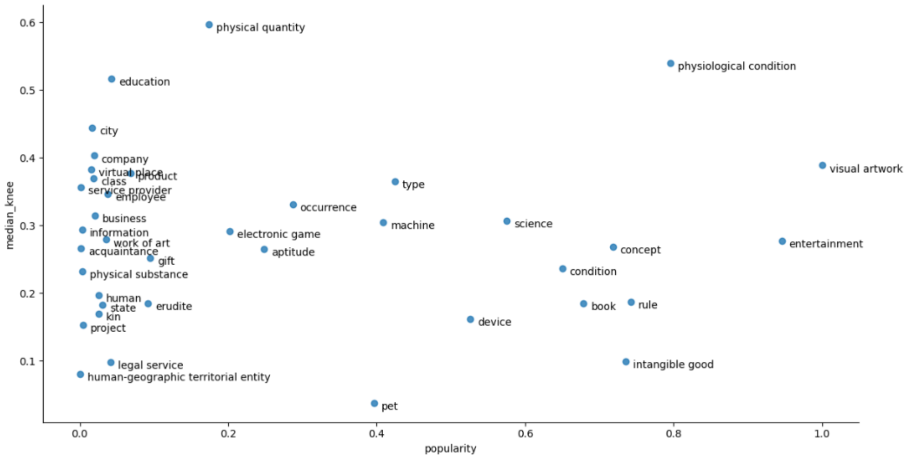

Fig. 5. Knee point and popularity relationships for topics mapped to Wikidata 3rd layer entities

After categorizing Zhihu topics into Wikidata entities, we can closely inspect the topic categories and observe the questions' lifecycle under the categories with similar domain coverage. The figures below (Fig. 6,7) give examples of the lifecycle of questions under *city* (Q515) and *occurrence* (Q1190554) categories. The normalized median knee point of the *city* questions is 0.4437, which is above the median knee (0.3231) of *occurrence* questions according to the category scatter above. When we look at the questions in the city category, we discover that the time sensitivity for these city-related topics is not particularly high. Most questions have been through the relatively flat and stable life stages (e.g., B-Fuzhou and C-Yinchuan), with others maintaining an even growth trend (e.g., A-Guangzhou). The knee point arrival time is fairly late for questions with substantial 2-stage lifecycles. For example, the life pattern of the Shanghai City topic reached a plateau roughly 100 hours after it was posted.

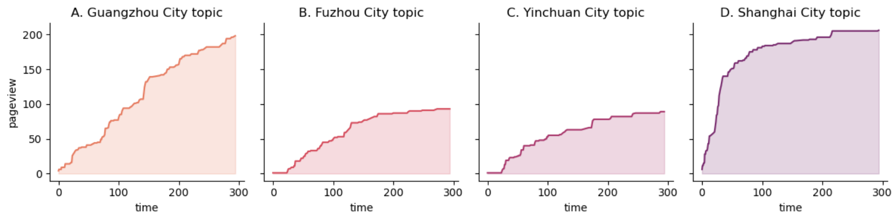

Fig. 6. Lifecycle examples of the questions under city (Q515) category. (Zhihu Question ID: A-447695699, B-447695843, C-447695907, D-447695912)

In relative terms, we find that the occurrence topic is relatively more time-sensitive, as the majority of the questions in our sample group reached their knee points within 40 hours of being posted.

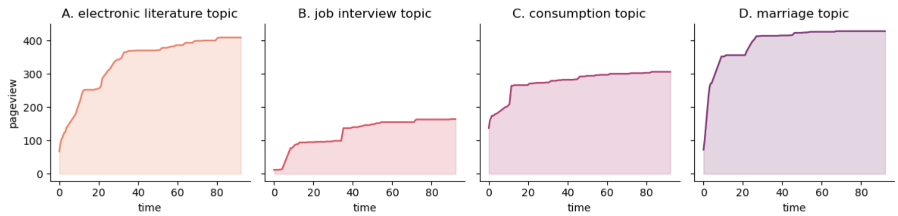

Fig. 7. Lifecycle examples of the questions under occurrence (Q1190554) category. (Zhihu Question ID: A-450461981, B-450462172, C-450462087, D-450461232)

5 Conclusion

This research proposed a novel rule-based approach to classify unstructured topics from Q&As into Wikidata entities with a selected level of domain coverage. Our research builds a bridge between unstructured information and structured data by demonstrating how Wikidata can impact other platforms by optimizing their topic structure. Additionally, the proposed label mapping and classification methods could be utilized on different platforms to achieve real-time classification, especially for Zhihu where the labels are manually classified to the topic structure. Another prospect of our proposed approach is the possibility of classifying and analyzing the Web content with different languages and formats under the same scheme. Replying to the external identifiers from Wikidata, our proposed framework could be easily applied to other social media, news platforms, or KBs. For example, tags/hashtags/labels from Quora, Twitter, or YouTube can also be mapped to the Wikidata entity linked to Zhihu labels. By doing this, we can easily visualize features of the same topic (trend, popularity, time sensitivity, etc.) from multiple platforms.

This research is yet ongoing as certain constraints exist for our proposed method that can be improved. One limitation is that not all Wikidata objects contain a Zhihu topic identifier, which prevents us from relying on external identifiers only to classify a large proportion of Zhihu topics. Besides, the content from many Websites is untagged, which indicates that using labels only to conduct topics classification is inadequate. In the future, we aim to enhance our method by adopting LDA-driven tools with Wikifier to allow direct correspondence from text to Wikidata entities. Lacking evaluation is another constraint of the present exploratory research, considering the performance of our method can serve as a benchmark for similar topic classification tasks. Therefore, our next step is to lean on the large annotated dataset to produce the evaluation matrix with other entity linking models compared.

References

1. Cha, M. et al.: Measuring User Influence in Twitter: The Million Follower Fallacy. Proc. Int. AAAI Conf. Web Soc. Media. 4, 1, (2010).
2. Crane, R., Sornette, D.: Robust dynamic classes revealed by measuring the response function of a social system. Proc. Natl. Acad. Sci. 105, 41, 15649–15653 (2008).
3. Delpeuch, A.: OpenTapioca: Lightweight entity linking for Wikidata. CEUR Workshop Proc. 2773, (2020).
4. Gupta, N. et al.: Entity linking via joint encoding of types, descriptions, and context. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 2681–2690 (2017).
5. Harandizadeh, B., Singh, S.: Tweeki: Linking Named Entities on Twitter to a Knowledge Graph. 2020 EMNLP Workshop W-NUT, 222–231 (2020).
6. Hong, L. et al.: Predicting Popular Messages in Twitter. In: Proceedings of the 20th International Conference Companion on World Wide Web. pp. 57–58 Association for Computing Machinery, New York, NY, USA (2011).
7. Kuang, L. et al.: Spillover Effects of Financial Incentives on Non-Incentivized User Engagement: Evidence from an Online Knowledge Exchange Platform. J. Manag. Inf. Syst. 36, 1, 289–320 (2019).
8. Lehmann, J. et al.: Dynamical Classes of Collective Attention in Twitter. In: Proceedings of the 21st International Conference on World Wide Web. pp. 251–260 Association for Computing Machinery, New York, NY, USA (2012).
9. Li, W. et al.: Topic-Guided Knowledge Graph Construction for Argument Mining. Proc. - 12th IEEE Int. Conf. Big Knowledge, ICBK 2021. 315–322 (2021).
10. Lin, Y.-R. et al.: #Bigbirds Never Die: Understanding Social Dynamics of Emergent Hashtags. Proc. Int. AAAI Conf. Web Soc. Media. 7, 1 SE-Full Papers, 370–379 (2013).
11. Ling, X. et al.: Design Challenges for Entity Linking. Transactions of the Association for Computational Linguistics. 3, 315–328 (2015).
12. Naveed, N. et al.: Bad News Travel Fast: A Content-Based Analysis of Interestingness on Twitter. In: Proceedings of the 3rd International Web Science Conference. Association for Computing Machinery, New York, NY, USA (2011).
13. Paulheim, H.: Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web. 8, 3, 489–508 (2017).
14. Pypi: zhihu-oauth 0.0.42, https://pypi.org/project/zhihu-oauth/.
15. Radhakrishnan, P. et al.: Elden: Improved entity linking using densified knowledge graphs. In: Proceedings of the 2018 Conference of the North American Chapter of the As-

sociation for Computational Linguistics: Human Language Technologies, Volume 1. pp. 1844–1853 (2018).
16. Raiman, J., Raiman, O.: Deeptype: multilingual entity linking by neural type system evolution. In: Proceedings of the AAAI Conference on Artificial Intelligence. (2018).
17. Romero, D.M. et al.: Differences in the Mechanics of Information Diffusion across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. In: Proceedings of the 20th International Conference on World Wide Web. pp. 695–704 Association for Computing Machinery, New York, NY, USA (2011).
18. Sha, A.S. et al.: Effects of Label Usage on Question Lifecycle in Q&A Community. Thirtieth Eur. Conf. Inf. Syst. (ECIS 2022). (2022).
19. Shamma, D.A. et al.: Peaks and Persistence: Modeling the Shape of Microblog Conversations. In: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work. pp. 355–358 Association for Computing Machinery, New York, NY, USA (2011).
20. Suh, B. et al.: Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. In: 2010 IEEE Second International Conference on Social Computing. pp. 177–184 (2010).
21. Turki, H. et al.: Enhancing multilingual and biomedical named entity recognition using Wikidata semantic relations ACM Reference Format. Association for Computing Machinery (2022).
22. Vrandečić, D., Krötzsch, M.: Wikidata: A Free Collaborative Knowledgebase. Commun. ACM. 57, 10, 78–85 (2014).
23. Wang, X. et al.: A comparative study for biomedical named entity recognition. International Journal of Machine Learning and Cybernetics. 9, 3, 373–382 (2018).
24. Yang, J., Leskovec, J.: Patterns of Temporal Variation in Online Media. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. pp. 177–186 Association for Computing Machinery, New York, NY, USA (2011).
25. Yoon, W. et al.: CollaboNet: collaboration of deep neural networks for biomedical named entity recognition. BMC Bioinformatics. 20, 10, 249 (2019).