Semantic Web 0 (2021) 1
IOS Press

Survey on English Entity Linking on Wikidata

Approaches and Datasets

Cedric Möller [a,*], Jens Lehmann [a,b] and Ricardo Usbeck [a]
[a] *NetMedia Department, Fraunhofer IAIS, Zwickauer Straße 46, 01069 Dresden, Germany*
E-mail: firstname.lastname@iais.fraunhofer.de
[b] *University of Bonn, Endenicher Allee 19a, 53115, Bonn, Germany*
E-mail: jens.lehmann@cs.uni-bonn.de

Abstract. Wikidata is an always up-to-date, community-driven, and multilingual knowledge graph. Hence, Wikidata is an attractive basis for Entity Linking, which is evident by the recent increase in published papers. This survey focuses on four subjects: (1) How do current Entity Linking approaches exploit the specific characteristics of Wikidata? (2) Which unexploited Wikidata characteristics are worth to consider for the Entity Linking task? (3) Which Wikidata Entity Linking datasets exist, how widely used are they and how are they constructed? (4) Do the characteristics of Wikidata matter for the design of Entity Linking datasets and if so, how?

Our survey reveals that most Entity Linking approaches use Wikidata in the same way as any other knowledge graph missing the chance to leverage Wikidata-specific characteristics to increase quality. Almost all approaches employ specific properties like labels and sometimes descriptions but ignore characteristics like the hyper-relational structure. Thus, there is still room for improvement, for example, by including hyper-relational graph embeddings or type information. Many approaches also include information from Wikipedia which is easily combinable with Wikidata and provides valuable textual information which is Wikidata lacking.

The current Wikidata-specific Entity Linking datasets do not differ in their annotation scheme from schemes for other knowledge graphs like DBpedia. The potential for multilingual and time-dependent datasets, naturally suited for Wikidata, is not lifted.

Keywords: Entity Linking, Entity Disambiguation, Wikidata

1. Introduction

1.1. Motivation

Entity Linking (EL) is the task of connecting already marked mentions in an utterance to their corresponding entities in a knowledge base, see Figure 1.

There are multiple knowledge bases such as DBpedia [68], Freebase [10], Yago4 [108] or Wikidata [120]. In contrast to DBpedia, Yago4, or Freebase, which mostly extract information from existing sources, Wikidata is a curated, community-based Knowledge Graph (KG). That is, the elements are added and edited by the community. The number of active editors is continuously increasing, see Figure 2. This allows Wikidata to stay up-to-date while automatically, one-time generated KGs such as Yago4 or Freebase become outdated over time [91]. Note, DBpedia stays also up-to-date but has a delay of a month.[1] DBpedia Live [21] exists, which is consistently updated with Wikipedia information. But it is more challenging to work with as no full dump is provided. Furthermore, the DBpedia ontology is not continuously updated, for example, with new emerging classes. The addition of new classes only comes with an update

*Corresponding author. E-mail: firstname.lastname@iais.fraunhofer.de.

[1] https://release-dashboard.dbpedia.org/

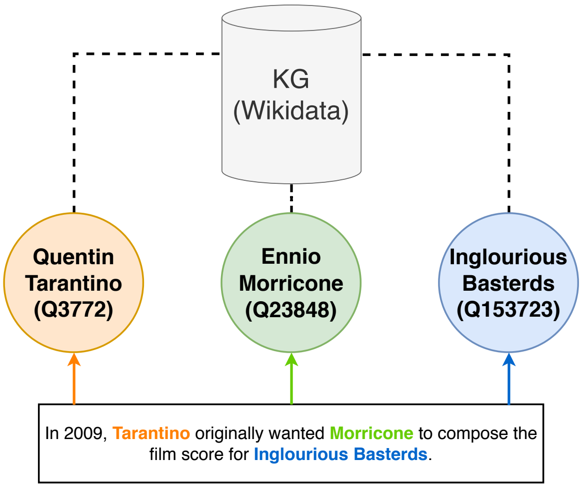


Fig. 1. Entity Linking - Mentions in text are linked to the corresponding entities (color-coded) in a knowledge base (here: Wikidata).

of the mapping-based extraction. On the other hand, new classes in Wikidata can be added continuously by the community. Furthermore, Wikidata is an inherently multilingual knowledge base. Both of these factors attract novel EL research over Wikidata in recent years cf. Figure 3. While Wikidata has its advantages regarding EL, exploiting those, for example in the form of hyper-relational structure (see Figure 4 for an example graph), is also challenging.

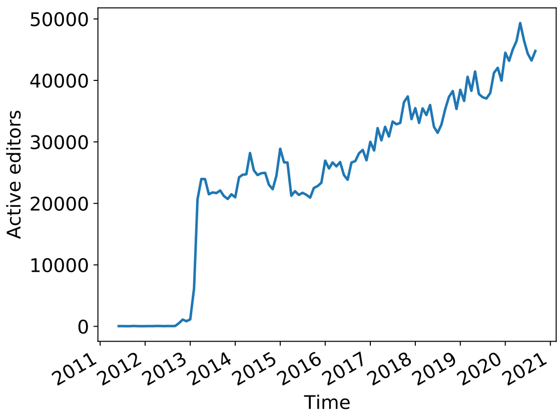


Fig. 2. Active editors in Wikidata [126].

Primarily, this survey strives to expose the benefits and associated challenges stemming from the effective use of Wikidata as the target KG for EL. Additionally, the survey provides a concise overview of existing approaches, which is essential to (1) avoid duplicated research in the future and (2) enable a smoother entry into the field of Wikidata EL. Similarly, dataset landscape is structured, which helps researchers finding the correct dataset for their EL problem.

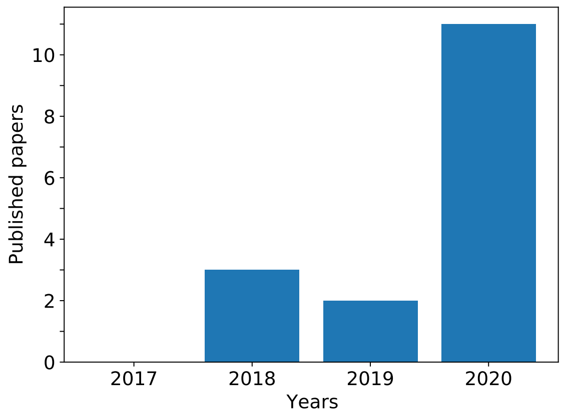


Fig. 3. Publishing years of included Wikidata EL papers.

The focus of this survey lies on EL approaches, which operate on already marked mentions of entities, as the task of Entity Recognition (ER) is much less dependent on the characteristics of a KG. However, due to the only recent uptake of research on EL on Wikidata there is only a low number of EL-only publications. To broaden the survey's scope, we also consider methods that include the task of ER. We do not restrict ourselves to either rule-, statistical- or deep learning-based algorithms on Wikidata. This survey limits itself to the English language as it is the most dominant language in EL, and thus a better comparison of the approaches and datasets is possible. Nevertheless, the topic of multilingualism is still of relevance in the analyses and discussions, as it is an essential characteristic of Wikidata. Since all multilingual Entity Linkers found also target English, none were excluded.

1.2. *Research Questions and Contributions*

EL approaches use many different kinds of information like labels, popularity measures, graph structure, and more. This multitude of possible signals raises the question of how the characteristics of Wikidata are used by the current state of the art of EL over Wikidata. Thus, the first research question is:

RQ 1: How do current Entity Linking approaches exploit the specific characteristics of Wikidata?

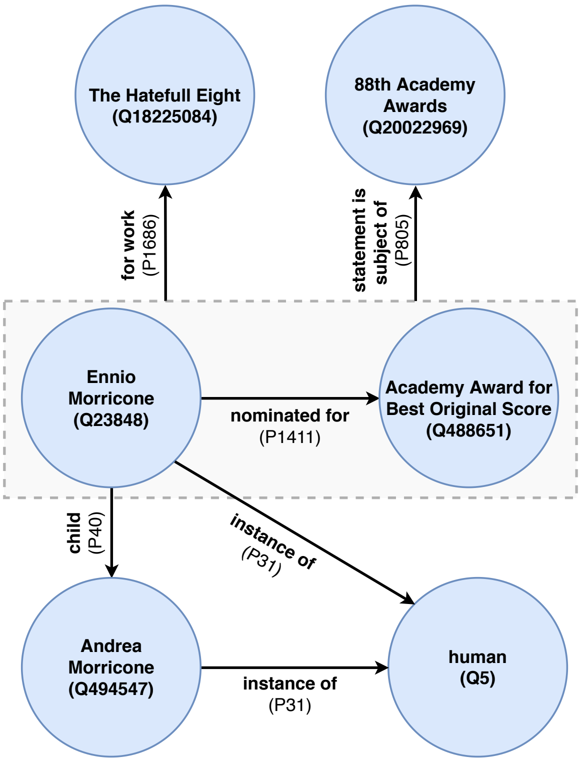


Fig. 4. Wikidata subgraph - Dashed rectangle represents a claim with attached qualifiers.

In particular, which Wikidata-specific characteristics contribute to the solution? We answer this question by gathering all existing approaches working on Wikidata systematically (see Section 2) and analyzing them. The focus lies mainly on the usage of Wikidata's graph characteristics.

Secondly, we identify what kind of characteristics of Wikidata are of importance for EL but are insufficiently considered. This raises the second research question:

> **RQ 2**: Which unexploited Wikidata characteristics are are worth to consider for the Entity Linking task?

We tackle this question by giving an overview of the structure of Wikidata and the amount of information it contains, and then discussing the potential and challenges for EL.

Furthermore, we want to give an overview of which datasets for EL over Wikidata exist. Lastly, it is of interest if it is essential that datasets are designed with Wikidata in mind and if so, in what way? Thus, we post the following two research questions:

> **RQ 3**: Which Wikidata EL datasets exist, how widely used are they and how are they constructed?

> **RQ 4**: Do the characteristics of Wikidata matter for the design of EL datasets and if so, how?

To answer those two last research questions, all current Wikidata-specific EL datasets are gathered and analyzed with the research questions in mind. Furthermore, we discuss how the characteristics of Wikidata might affect the design of datasets.

This survey makes the following contributions:

- A concise list of future research avenues.
- A list and comparison of datasets focusing on Wikidata.
- An analysis of current evaluation results.
- A discussion of the relevance of Wikidata for Entity Linking.

2. Survey Methodology

There are several different types of surveys which desire to accomplish different contributions to the research field [59]:

1. Providing an overview of the current prominent areas of research in a field
2. Identification of open problems
3. Providing a novel approach tackling the extracted open problems (in combination with the identification of open problems)

Our related work section analyses different recent and older surveys on EL and highlights specific areas not covered and our survey's novelties. While some very recent surveys exist, they do not consider the different underlying Knowledge Graphs as a significant factor affecting the performance of EL approaches. Furthermore, barely any approaches included in other surveys are working on Wikidata and take the particular characteristics of Wikidata into account. To fill in the gaps, our survey gives an overview and examines all current EL approaches and datasets, focusing on Wikidata. Additionally, we identify less-utilized but promising characteristics of Wikidata regarding EL. Therefore, this survey provides contributions 2 and 3.

Table 1: Qualifying and disqualifying criteria for approaches.

Must satisfy all	Must not satisfy any
– Approaches that consider the problem of unstructured EL over Knowledge Graphs – Approaches where the target Knowledge Graph is Wikidata	– Approaches conducting Semi-structured EL – Approaches not doing EL in the English language

Until December 18, 2020, we continuously searched for existing and newly released scientific work suitable for the survey. Note, this survey includes only scientific articles that were accessible to the authors.[2]

2.1. Approaches

This survey's qualifying and disqualifying criteria for including papers can be found in Table 1. "Semi-structured" in this table means that the entity mentions do not occur in natural language utterances but more structured formats such as tables. The different approaches were searched for by using multiple different search engines (see Table 3).

To gather a wide choice of approaches the following filters were applied. Any approach where `Wikidata` was not occurring once in the full text was not considered. `Entity Linking` or `Entity Disambiguation` had to occur in the title of the paper. The publishing year was not a criterion due to the small number of valid papers and the relatively recent existence of Wikidata. The systematic search process resulted in 150 papers and theses (including duplicates).

[2]https://www.projekt-deal.de/max-planck-gesellschaft-verzichtet-ab-2019-auf-elsevier/

Following this search, the resulting papers were filtered again using the qualifying and disqualifying criteria. This resulted in 16 papers and one master thesis in the end.

The search resulted in papers in the period from 2018 to 2020. While there exist EL approaches from 2016 [4, 106] working on Wikidata, they did not qualify according to the criteria above.

2.2. Datasets

The dataset search was conducted in two ways. First, a search for potential datasets was performed using multiple search engines, see Table 3. Second, the datasets on which the approaches were evaluated were considered. The criteria for the inclusion of a dataset can be found in Table 2.

We scanned the dataset papers in the following way. First, in the title, `Entity Linking` or `Entity Disambiguation` had to occur once. Due to those keywords, other datasets suitable for EL but constructed for a different purpose like KG population

Table 2: Qualifying and disqualifying criteria for the dataset search.

Must satisfy all	Must not satisfy any
– Datasets that are designed for EL or are used for evaluation of Wikidata EL – Datasets must include Wikidata identifiers from the start	– Datasets without English utterances

Table 3: Search engines.

Search Engines
– Google Scholar – Springer Link – Science Direct – IEEE Xplore Digital Library – ACM Digital Library

were not included. Additionally, `dataset` must occur in the title and `Wikidata` has to appear at least once in the full text. This resulted in 20 papers (including duplicates). Of those, only two included Wikidata identifiers and focused on English.

Eighteen datasets are accompanying the different approaches. Many of those did not include Wikidata identifiers from the start. This makes them less optimal for the examination of the influence of Wikidata on the design of datasets. They are included in the section about the approaches but not in the section about the Wikidata datasets.

After removal of duplicates, 11 Wikidata datasets are included in the end.

3. Problem Definition

EL is the task of linking an entity mention in unstructured or semi-structured data to the correct entity in a KG. The focus of this survey lies in unstructured data, namely natural language utterances.

An utterance is defined as a sequence of n words.

$$s = (w_0, w_1, ...w_{n-1})$$

Since not only approaches that solely do EL were included in the survey, Entity Recognition will also be defined.

There exists no universally agreed on definition of an entity. In general, named entities like a specific person or an organization are desirable to link. But sometimes, also common entities, such as `interview` or `theater`, are included. What exactly needs to be linked, depends on the use case [95].

Entity Recognition. ER is the task of identifying the spans

$$(w_i, ..., w_k)|0 \leqslant i \leqslant k \leqslant n - 1$$

of all entities in an utterance u. Each such a span is called an entity mention m. The word or word sequence referring to an entity is also known as the surface form of the entity. An utterance can contain more than one entity, often also consisting of more than one word. Sometimes, also some broad type of an entity is classified too. Normally, those are `person`, `location` and `organization`. Some of the considered approaches do this classification task and also use it to improve the EL. It is also up to debate what an entity mention is. In general, a literal reference to an entity is considered a mention. But whether to include pronouns or how to handle overlapping mentions depends on the use-case.

Entity Linking. EL is the task of linking the recognized entity mention to the correct entity in a KG. A KG is defined as a directed graph $G = (V, E, \mathcal{R})$ consisting of vertices V, edges E and relations $\mathcal{R}$. Often, vertices correspond to entities $\mathcal{E}$ or literals $\mathcal{L}$, which are concrete values like the height or a name. E is a list $(e_1, \ldots, e_n)$ of edges with $e_j \in V \times \mathcal{R} \times V$ where relations $\mathcal{R}$ specify a certain meaning for the connection between entities. Such edges are also called triples. But there exists no single definition of a KG; vertices and edges can also be defined differently. A concrete definition of the Wikidata KG is provided in the next section.

In general, EL takes the utterance u and all identified entity mentions $M = (m_1, ...m_n)$ in the utterance and links each of them to an element of the set $(\mathcal{E} \cup \{unknown\})$. The *unknown* element is added to the set of vertices to be able to map to an unknown entity that is not available in the KG. Such an entity is also called a NIL or an emerging entity [52].

The goal of EL is to find a mapping function that maps all found mentions to the correct KG entities and also to identify if an entity mention does not exist in the KG.

EL is often split into two subtasks. First, potential candidates for an entity are retrieved from a KG. This is necessary as doing EL over the whole set of entities is often intractable. *Candidate generation* is usually performed via efficient metrics measuring the similarities between entities in the utterance and entities in the KG.

The result is a set of candidates $C = \{c_0, \cdots, c_l\}$ for each entity mention m in the utterance.

After limiting the space of possible entities, one of the available candidates is chosen for each entity. This is done via a *candidate ranking* algorithm, which assigns a rank to each candidate, signalizing how likely it is the correct one.

$$rank_{local} : C \times M \to \mathbb{R}$$
$$\text{given by } (c, m) \mapsto rank_{local}(c, m)$$

where $rank_{local}$ is a ranking function of a candidate. The goal is then to optimize the objective function:

$$A^* = \arg\max_A \sum_{i=1}^{n} rank_{local}(a_i, m_i) | a_i \in C_i$$

where $A = \{a_1, ..., a_n\} \in \mathcal{P}(\mathcal{E})$ is an assignment of one candidate to each entity mention m_i. $\mathcal{P}(*)$ is the power set operator.

The rank calculation of the candidates of one entity is often not independent of the other entities' candidates. In this case, another global ranking function will include the whole assignment:

$$rank_{global} : \mathcal{P}(\mathcal{E}) \to \mathbb{R} \text{ given by } A \mapsto rank_{global}(A)$$

The objective function is then:

$$A^* = \arg\max_A \left[\sum_{i=1}^{n} rank_{local}(a_i, m_i) \right] + rank_{global}(A) \mid a_i \in C_i$$

Those two different categories of reranking methods are called *local* or *global* [90].

There exists also some ambiguity in the object of linking itself. For example, there exists an Wikidata entity `2014 FIFA World Cup` and an entity `FIFA World Cup`. There is no unanimous solution on how to link the entity mention in the utterance `In 2014, Germany won the FIFA World Cup`.

Sometimes EL is also called Entity Disambiguation, which we see more as part of EL, namely where entities are disambiguated via the candidate ranking.

4. Wikidata

Wikidata is a community-driven knowledge graph edited by humans and machines. As of July 2020, it contained around 87 million items of structured data about various domains. Seventy-three million items can be interpreted as entities due to the existence of a `is_instance` property. As a comparison, DBpedia contains around 5 million entities [108]. Note that the `is_instance` property includes a much broader scope of entities than the ones interpreted as entities for DBpedia. However, Wikidata contains around 8.5 million persons while DBpedia only contains around 1.8 million (in October 2020). Thus, a large difference in size is obvious.

4.1. Definition

Wikidata is a collection of *entities* where each such an entity has a page on Wikidata. An entity can be either an item or a property. Note that an entity in the sense of Wikidata is generally not the same as an entity one links to via EL. For example, Wikidata entities are also properties which describe relations between different items. Linking to such relations is closer to Relation Extraction [7, 71, 103]. Furthermore, many of the items are more abstract classes, which are usually also not considered as entities linked-to in EL. Note that if not mentioned otherwise, if we speak about entities, entities in the context of EL are meant.

Item. Topics, classes, or objects are defined as items. An example of an item can be found in Figure 5. An item is enriched with more information using statements about the item itself. In general, items consist of one label, one description, and aliases in different languages. An unique and language-agnostic identifier identifies items in the form `Q[0-9]+`.

For example, the item with the identifier `Q23848` has the label `Ennio Morricone`, two aliases, `Dan Savio` and `Leo Nichols`, and `Italian composer, orchestrator and conductor (1928-2020)` as description at the point of writing. The corresponding Wikidata page can also be seen in Figure 5.

Not all items are entities in the context of EL. In general, items which are unique instances of some class are interpreted as entities. Of course, this also depends on the use case.

Table 4: KG statistics by [108].

KG	#Entities in million	#Labels/Aliases in million	last updated
Wikidata	78	442	always
DBpedia	5	22	monthly
Yago4	67	371	November 2019

Property. A property specifies a relation between items/literals. Each property also has an identifier similar to an item, specified by $P[0-9]*$. For instance, a property `P19` specifies the place of birth `Rome` for `Ennio Morricone`. In NLP, the term `relation` is commonly used to refer to a certain connection between entities. A property in the sense of Wikidata is a type of relation. To not break with the terminology used in the examined papers, when we talk about relations, we always mean Wikidata properties if not mentioned otherwise.

Statement. A statement introduces information by giving structure to the data in the graph. It is specified by a *claim*, and *references*, *qualifiers* and *ranks* related to the claim. Statements are assigned to items in Wikidata. A claim is defined as a pair of a property and some value. A value can be another item or some literal. Multiple values are possible for a property. Even an `unknown value` and a `no value` exists.

References point to sources making the claims inside the statements verifiable. In general, they consist of the source and date of retrieval of the claim. *Qualifiers* define the value of a claim further by contextual information. For example, a qualifier could specify how long one person was the spouse of another person. *Ranks* are used if multiple values are valid in a statement. If the population of a country is specified in a statement, it might be also useful to have the populations of past years available. The most up-to-date population information usually has then the highest rank and is thus usually the most desirable claim to use.

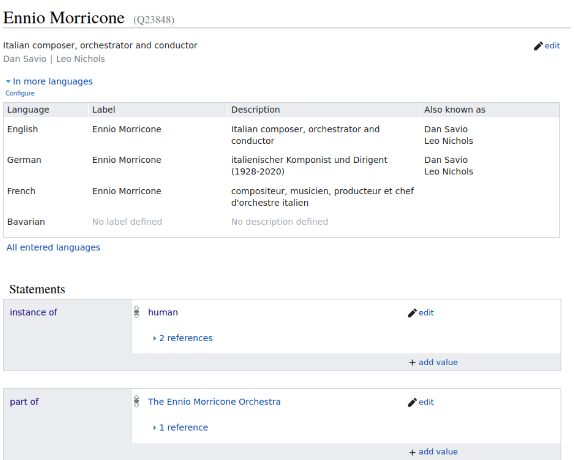
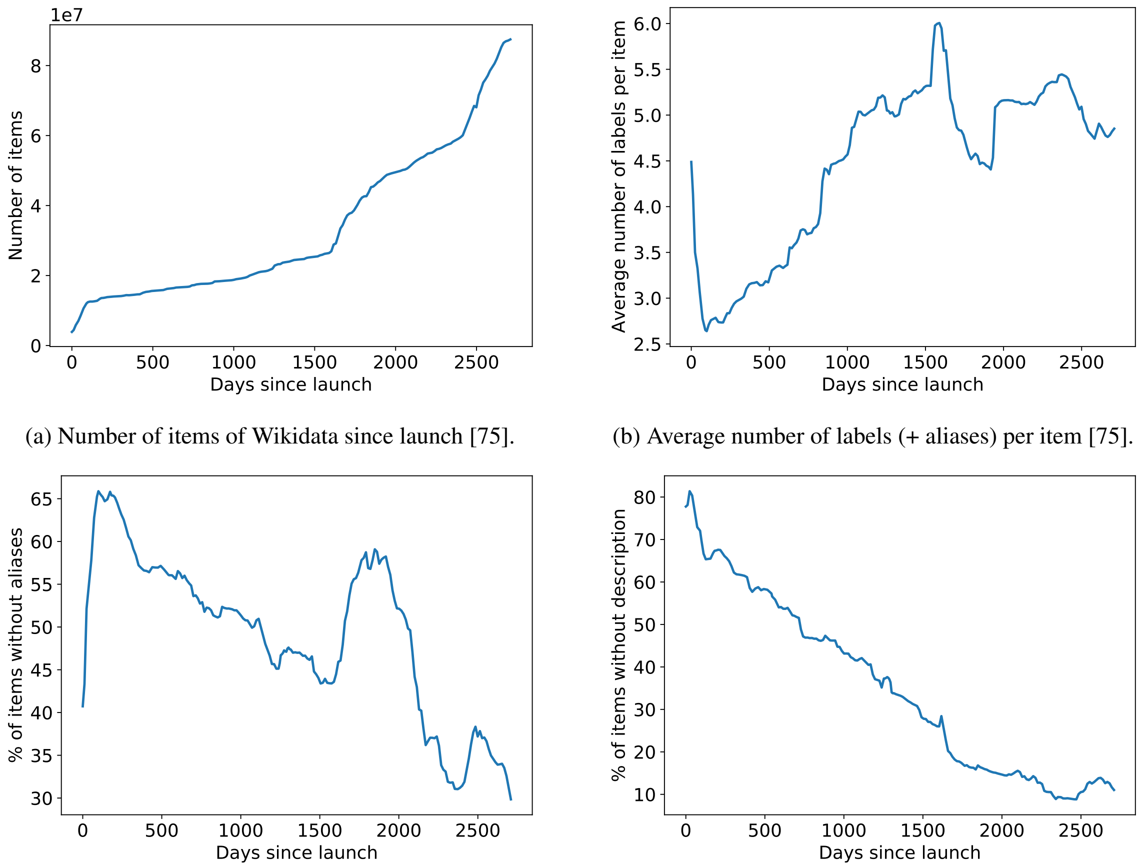


Fig. 5. Example of an item in Wikidata

Statements can be also seen in Figure 5 at the bottom. For example, it is defined that `Ennio Morricone` is an `instance of` the class `human`. This is also an example for the different types of items. While `Ennio Morricone` is an entity in our sense, `human` is a class.

Hyper-Relational Graphs. Wikidata can thus be defined as a hyper-relational knowledge graph as statements can be specified by more information than a single claim. Multiple properties/relations are therefore part of a statement. In case of a hyper-relational graph $\mathcal{G} = (V, E, \mathcal{R})$, E is a list $(e_1, \ldots, e_n)$ of edges with $e_j \in V \times \mathcal{R} \times V \times \mathcal{P}(\mathcal{R} \times V)$ for $1 \leqslant j \leqslant n$, where $\mathcal{P}$ denotes the power set. A hyper-relational fact $e_j \in E$ is usually written as a tuple $(s, r, o, \mathcal{Q})$, where $\mathcal{Q}$ is the set of *qualifier pairs* $\{(qr_i, qv_i)\}$ with *qualifier relations* $qr_i \in \mathcal{R}$ and *qualifier values* $qv_i \in V$. (s, r, o) is referred to as the *main triple* of the fact. We use the notation Q_j to denote the qualifier pairs of e_j [39]. For example, under this representation scheme, the `nominated for` edge in Fig. 4 has two additional claims and would be represented as `(Ennio Morricone, nominated for, Academy Award for Best Original Score, (for work, The Hateful Eight), (statement is subject of, 88th Academy Awards))` Structures similar to qualifiers exist also in some other knowledge graphs, such as the inactive Freebase in the form of Compound Value Types [10].

Other structural elements. The aforementioned elements are essential for Wikidata but more do exist. For example, there are entities (in the sense of Wikidata) corresponding to Lexemes, Forms, Senses or Schemas.



(a) Number of items of Wikidata since launch [75].

(b) Average number of labels (+ aliases) per item [75].

(c) Percentage of items without any aliases [75].

(d) Percentage of items without a description [75].

Fig. 6. Statistics on Wikidata based on [75].

Yet, as those are in general not of relevance for EL, we refrain from introducing them in more detail.

For more information on Wikidata, see the paper by Denny Vrandečić and Markus Krötzsch [120].

4.2. Discussion

Novelties. As already mentioned, a useful characteristic of Wikidata is that the community can openly edit it. Another novelty is that there can be a plurality of facts, as contradictory facts based on different sources are allowed. Similarly, time-sensitive data can also be included easily by qualifiers and ranks. The population of a country, for example, changes from year to year which can be represented easily in Wikidata. Lastly, due to their language-agnostic identifiers, Wikidata is inherently multilingual. Language only starts playing a role in the labels and descriptions of an item.

Strengths. Due to the inclusion of information by the community, recent events will always be included. The knowledge graph is thus much more up to date than other KGs. Freebase is unsupported for years now, and DBpedia updates its dumps only every month. Thus, Wikidata is much more suitable and useful for industry applications such as smart assistants since it is the most complete open accessible data source to date. In Figure 6a, one can see that number of items in Wikidata is increasing steadily. The existence of labels and additional aliases (see Figure 6b) helps EL as a too small amount of possible surface forms often lead to a failure in the candidate generation. DBpedia does for example not include aliases, only a single exact label; to compensate, additional resources like Wikipedia are often used to extract a label dictionary of adequate size [77]. Even each property in Wikidata has a label [120]. Fully language-model based approaches are therefore more naturally usable [80]. Also, nearly all

Table 5: Statistics - Languages Wikidata (Extracted from dump [125])

	Items	**Properties**
Number of languages	457	427
(average, median) of # languages per element (labels + descriptions)	29.04, 6	21.24, 13
(average., median) of # languages per element (labels)	5.59 , 4	21.18, 6
(average, median) of # languages per element (descriptions)	26.10, 4	9.77, 6
% elements without English labels	15.41%	0%
% elements without English descriptions	26.23%	1.08%

items have a description, see Figure 6d. Thus, this short natural language phrase can be used for context similarity measures with the utterance. The inherent multilingual structure is intuitively useful for multilingual Entity Linking. Table 5 shows information about the use of different languages in Wikidata. As can be seen, are item labels/aliases available in up to 457 languages. Of course, not all items have labels in all languages. On average, labels/aliases/descriptions are available in 29.04 different languages. However, the median is only 6 languages. Many entities will therefore certainly not have information in many languages. The most dominant language is English but not all elements have label/alias/description information in English. For less dominant languages, this is of course more severe. German labels exist for example only for 14 %, and Samoan labels for 0.3 %. Context information in the form of descriptions is also given in multiple languages but many languages are again not covered for each entity (as can be seen by a median of only 4). While the multilingual label and description information of items might be useful for language model based variants, the same information for properties enables multilingual language models. Because, on average, 21.18 different languages are available per property for labels, one could train multilingual models on the concatenations of the labels of triples to include context information. But of course, there are again many properties with a lower number of languages, as the median is also only 6 languages. Cross-lingual EL is therefore certainly necessary to use language-model based EL in multiple languages.

By using the qualifiers of hyper-relational statements more detailed information is available, useful not only for Entity Linking but also for other problems like Question Answering. The inclusion of hyper-relational statements is of course also more challenging. Novel graph embeddings have to be developed and utilized which can represent the structure of a claim enriched with qualifiers [39, 96].

Weaknesses. However, this community-driven approach does also introduce challenges. For example, the list of labels of an item will not be exhaustive, as shown in Figures 6b and 6c. The graphs consider labels and aliases of all languages. While the average number of labels/aliases is around 5, not all are useful for Entity Linking in English. `Ennio Morricone` does not have an alias solely consisting of `Ennio` while he will certainly sometimes be referenced by that. Thus, one can not rely on the exact labels alone. But interestingly, Wikidata has properties for the fore- and surname alone, just not as a label or alias. A close examination of what information to use is essential. However, this is also a problem in other KGs. Also, Wikidata often has items with very long, noisy, error-prone labels, which can be a challenge to link to [80]. Nearly 20 percent of labels have a length larger than 100 letters, see Figure 7. Due to the community-driven approach, false statements, due to errors or vandalism [49], also occur.

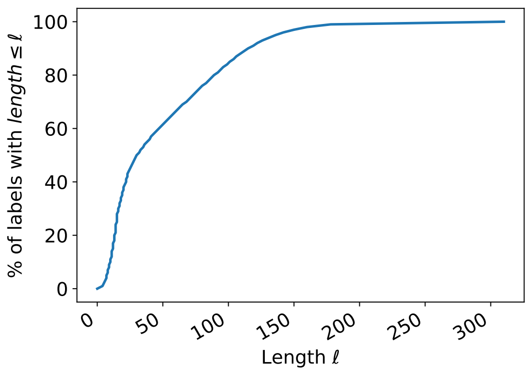


Fig. 7. Percentiles of English label lengths (Extracted from dump [125])

Another problem may be the lack of facts (here defined as statements not being labels, descriptions, or aliases) for some entities. According to Tanon et al. [108], in March 2020, DBpedia had, on average, 26 facts per entity while Wikidata had only 12.5. This is still more than YAGO4 with 5.1. However, those entities with fewer facts are probably also not occurring in DBpedia, which has a much lower amount of entities [108]. To tackle such long-tail entities, different approaches are necessary. The lack of descriptions can also be a problem. Currently, around 10% of all items do not have a description, as shown in Figure 6d. However, the situation is increasingly improving.

A general problem of Entity Linking is that a label or alias can reference multiple entities, see Table 6. While around 70 million mentions point each to an unique item, 2.9 million do not. Not all of those are entities by our definition but, e.g., also classes or topics. Also, longer labels or aliases often correspond to non-entity items. Thus, the percentage of entities with overlapping labels/aliases is certainly larger than for all items. To use Wikidata as a Knowledge Graph, one needs to be cautious of the items one will include as entities. For example, there exist `Wikimedia disambiguation page` items which often have the same label as an entity in the classic sense. Both, `Q76` vs `Q61909968` have `Barack Obama` as the label. Including those will make disambiguation more difficult. Also, the possibility of contradictory facts will make EL over Wikidata harder.

In Wikification, also known as EL on Wikipedia, large text documents for each entity exist in the knowledge base, enabling text-heavy methods [127]. Such large textual contexts (besides the descriptions and the labels of triples itself) do not exist in Wikidata requiring other methods or the inclusion of Wikipedia. However, as Wikidata is closely related to Wikipedia, an inclusion is easily doable.

One can conclude that characteristics of Wikidata, like being up to date, multilingual and hyper-relational, introduce new possibilities while the existence of long-tail entities, noise or contradictory facts is also challenging. Thus, **RQ 2** is answered.

5. Approaches

5.1. Overview

Currently, the number of methods intended to work explicitly on Wikidata is still relatively small, while the amount of the ones utilizing the structure of Wikidata is even smaller.

There exist several KG-agnostic EL approaches [78, 112, 136]. However, they were omitted as their focus is being independent of the KG. Of course, they do use Wikidata information like labels as this information also exists in other KGs, but it is no explicit usage of Wikidata-specific characteristics. While the approach by Zhou et al. [135] does utilize Wikidata aliases in the candidate generation process, the target KB is Wikipedia and was therefore also excluded.

Tools without accompanying publications are not considered due to the lack of information about the approach and its performance. Hence, for instance, the Entity Linker in the DeepPavlov [16] framework is not included, though it targets Wikidata and appears to use label and description information successfully to link entities.

We distinguish three different kind of approaches: (1) Rule-based approaches, (2) approaches employing statistical methods and (3) neural network-based approaches. The vast amount of methods are using neural networks to solve the EL task [6, 13, 14, 18, 54, 61, 67, 80, 81, 85, 88, 89, 104]. Some of those approaches solve the ER and EL jointly as an end-to-end task. Besides those, there exists one purely rule based approach [98] and two based on statistical methods [23, 70].

The approaches mentioned above solve the EL problem as specified in Section 3. That is, other EL methods with a different problem definition also exist. For example, Almeida et al. [4] try to link street names to entities in Wikidata by using additional location information and limiting the entities only to locations. As it uses additional information about the true entity via the location, it is less comparable to the other approaches. Thawani et al. [110] link entities only over columns of tables. It is not comparable since it does not use natural

Table 6: Number of English labels/aliases pointing to a certain number of items in Wikidata (Extracted from dump [125])

# Labels/aliases	70,124,438	2,041,651	828,471	89,210	3329
# Items per label/alias	1	2	$3 - 10$	$11 - 100$	< 100

language utterances. The approach by Klie et al. [63] is concerned with Human-In-The-Loop EL. While its target KB is Wikidata, the focus on the inclusion of a human in EL process makes it incomparable to the other approaches. EL methods working on other languages than English [28, 30, 31, 60, 114] were not considered but also did not use any novel characteristics of Wikidata. In connection to the CLEF HIPE 2020 challenge [28], multiple Entity Linkers working on Wikidata were built. While short descriptions of the approaches are available in the challenge-accompanying paper, only approaches described in an own published paper were included in this survey. The approach by Kristanti and Romary [65] was not included as it used pre-existing tools for EL over Wikidata for which no sufficient documentation was available.

Due to the limited number of methods, we also evaluated methods that are not solely using Wikidata but also additional information from a separate KG or Wikipedia. This is mentioned accordingly. Approaches linking to knowledge graphs different from Wikidata, but for which a mapping between the knowledge graphs and Wikidata exists, are also not included. Such methods would not use the Wikidata characteristics at all and their performance depends only on the quality of the other KG and the mapping.

In the following, the different approaches are described and examined according to the characteristics of Wikidata used. For an overview, see Table 7.

5.1.1. Entity Linking

In the following, we will first focus on methods only doing EL.

In 2018, Cetoli et al. [18] evaluated how different types of basic neural networks perform solely over Wikidata. Notably, they compared the different ways to encode the graph context via neural methods, especially the usefulness of including topological information via GNNs [105, 129] and RNNs [51]. However, there is no candidate generation as it was assumed that the candidates are available. The process consists of combining text and graph embeddings. The text embedding is calculated by applying a Bi-LSTM over the Glove Embeddings of all words in an utterance. The resulting hidden states are then masked by the position of the entity mention in the text and averaged. A graph embedding is calculated in parallel via different methods utilizing GNNs or RNNs. The end score is the output of one feed-forward layer having the concatenation of the graph and text embedding as its input. It represents if the graph embedding is consistent with the text embedding. One crucial problem is that those methods only work for a single entity in the text. Thus, it has to be applied multiple times, and there will be no information exchange between the entities. While the examined algorithms do utilize the underlying graph of Wikidata, the hyper-relational structure is not taken into account. The paper is more concerned with comparing how basic neural networks work on the triples of Wikidata. Due to the pure analytical nature of the paper the usefulness of the designed approaches to a real-world setting is limited. The reliance on graph embeddings make it susceptible to change in the Wikidata KG.

Deeptype [89] is a novel approach using the type information of Wikidata or Wikipedia. Developed in 2018, first, a type system was optimized via stochastic optimization. A type system is a grouping of multiple type axes where a type axis is a set of mutually exclusive types. The idea is to classify entities according to the different type axes. Various methods to generate the type system were compared, such as a genetic algorithm. The objective was a type system which improves the EL performance while also being learnable. The learnability is important to guarantee that a classifier can be trained for the type system. After optimization, it consists of 128 different types. The authors do not mention how the candidates are generated. It is only stated that commonly it is done via a dictionary, therefore, one can only assume that they used a dictionary. Then the words in an utterance are classified via a windowed Bi-LSTM according to the type system. The type probabilities are then used together with a link probability score to get the final score per candidate. This link probability a statistic on how often a mention is linked to an article of an entity in Wikipedia. The approach is multilingual as its learned type system is agnostic to language. Thus it can be easily used with entity mentions in different languages. It is important to note that they used Wikipedia categories to train their type system and Wikipedia articles to train the type classifier. However, the authors claim that the algorithm is easily changeable to Wikidata. Nevertheless, as it is also possible to adapt other algorithms, initially created for different KGs, to Wikidata, this method may not be suitable to be compared to the other algorithms. Assuming it could be used over Wikidata types, it seems to produce quite good results while only using a basic disambiguation algorithm besides the type classifier. The results show that incorporating

Table 7: Comparison between the utilized Wikidata characteristics of each approach.

Approach	**Labels/ Aliases**	**Descrip- tions**	**Knowledge graph structure**	**Hyper- relational structure**	**Types**	**Additional Informa- tion**
OpenTapioca [23]	✓	✗	✓	✓	✓	✗
NED using DL on Graphs [18]	✓	✗	✓	✗	✗	✗
Falcon 2.0 [98]	✓	✗	✓[3]	✗	✗	✗
Arjun [80]	✓	✗	✗	✗	✗	✗
DeepType [89]	✓[1]	✗	✗	✗	✓[1]	Wikipedia[4]
Hedwig [61]	✓	✓	✓	✗	✗	Wikipedia
VCG [104]	✓	✗	✓	✗	✗	✗
KBPearl [70]	✓	✗	✓	✗	✗	✗
PNEL [6]	✓	✓	✓	✗	✗	✗
Mulang et al. [81]	✓	✓[2]	✓	✗	✗	✗
Perkins [85]	✓	✗	✓	✗	✗	✗
Huang et al. [54]	✓	✓	✓	✗	✗	Wikipedia
Boros et al. [13]	✗	✗	✗	✗	✓	Wikipedia, DBpedia
Provatorov et al. [88]	✓	✓	✗	✗	✗	Wikipedia
Labusch and Neudecker [67]	✗	✗	✗	✗	✗	Wikipedia
Botha et al. [14]	✗	✗	✗	✗	✗	Wikipedia
Tweeki [48]	✓	✗	✗	✗	✓	Wikipedia

[1] In paper, just demonstrated for Wikipedia
[2] Appears in the set of triples used for disambiguation
[3] Only querying the existence of triples
[4] Wikidata not used in implementation/evaluation

detailed type information improves EL considerably. As Wikidata contains many more types (≈2,400,000) than other KGs, e.g., DBpedia (≈484,000) [108], it seems to be more suitable for this fine-grained type classificaton. Yet, not only the amount of types plays a role but also how many types are assigned per entity. In this regard, Wikipedia provides much more type information per entity than Wikidata [124]. A shift to Wikidata is, therefore, not that simple. As Wikidata is growing every minute, it may also be challenging to keep the type system up to date.

The approach by Mulang et al. [81] is tackling the EL problem with Transformer [116] models. It is assumed that the candidate entities are given. For each entity, the labels of 1-hop and 2-hop triples are extracted. Those are then concatenated together with the utterance and the entity mention. The concatenation is the input of a pre-trained Transformer model. With a fully connected layer on top, it is then optimized according to a binary cross-entropy loss. This architecture results in a similarity measure between the entity and the entity mention. The examined models are the Transformer models Roberta [73], XLNet [132] and the DCA-SL model [130]. There is no global coherence technique applied. Overall, up to 2-hop triples of any kind are used. For example, labels, aliases, descriptions, or general relations to other entities are all incorporated. It is not mentioned if the hyper-relational structure in the form of qualifiers were used. On the one hand, the purely language-based EL results in less need of retraining if the KG changes. On the other hand, the reliance on the triple information might be problematic for long-tail entities.

The master thesis by Perkins [85] is performing candidate generation by using anchor link probability over Wikipedia and LSH over labels and mention bigrams. Contextual word embeddings of the utterance (ELMo [86]) are used together with KG embeddings (TransE [11]), calculated over Wikipedia and Wikidata, respectively. The context embeddings are sent through a recurrent neural network. The output is concatenated with the KG embedding and then fed into a feed-forward neural network giving a similarity measure between the KG embedding of the entity candidate and the utterance. The KG is used in the form of the calculated TransE embeddings. Hyper-relational structures like qualifiers are not mentioned in the thesis and not considered by the TransE embedding algorithm. Thus, probably not included. The used KG embeddings make it necessary to retrain when the Wikidata KG changes as they are not dynamic.

The approach designed by Botha et al. [14] tackles multilingual EL. It is also crosslingual. That means, it can link entity mentions to entities in a knowledge graph in a language different to the utterance one. The idea is to train one model to link entities in utterances of 100+ different languages to a KG containing not necessarily textual information in the language of the utterance. While the target KG is Wikidata, they mainly use Wikipedia descriptions as input. This is the case as extensive textual information is not available in Wikidata. But as Wikipedia articles are easily linkable to the corresponding Wikidata entities, gathering the desired textual information is easy. Furthermore, as the Wikidata entities have language-agnostic identifiers, Wikidata is suited to be the target KG. The approach resembles the Wikification method by Wu et al. [127] but extends the training process to be multilingual and targets Wikidata. Candidate generation is done via a dual-encoder architecture. Here, two BERT-based Transformer models encode both the context-sensitive mentions and the entities to the same vector space. The mentions are encoded using local context, the mention and surrounding words, and global context, the document title. Entities are encoded by using the Wikipedia article description available in different languages. In both cases, the encoded CLS-token are projected to the desired encoding dimension. The goal is to embed mentions and entities in such a way that the embeddings are similar. The model is trained over Wikipedia by using the anchors in the text as entity mentions. Now, after the model is trained, all entities are embedded. The candidates are generated by embedding the mention and searching for the nearest neighbors. A certain number of neighbors are then the generated candidates. A cross-encoder is employed to rank the entity candidates, fed with the concatenation of the entity description and mention text. Final scores are obtained and the entity mention is linked. Wikidata information is only used to gather all the Wikipedia descriptions in the different languages for all entities. Besides that, one relies mainly on Wikipedia. While that is the case, it is also clear that Wikidata is very suitable as the target KG for multilingual EL as its entities themselves are language-agnostic. The approach was tested on zero- and few-shot settings showing that the model can handle an evolving knowledge base with newly added entities that were never seen before. This is also more easily achievable due to its missing reliance on the graph structure of Wikidata or the structure of Wikipedia. It is the case that some Wikidata entities do not appear in Wikipedia and are therefore invisible to the approach. But this is less problematic here than for other approaches. The model is trained over descriptions of entities in multiple languages. Other approaches only use the English Wikipedia, which misses entities available in other languages. Thus, the amount of available entities is larger.

5.1.2. *Entity Recognition and Entity Linking*

The following methods all include ER in their EL process.

In 2018, Sorokin and Gurevych [104] were doing joint end-to-end ER and EL on short texts. The algorithm tries to incorporate multiple context embeddings into a mention score, signaling if a word is a mention, and a ranking score, signaling the candidate's correctness. First, it generates several different tokenizations of the same utterance. For each token, a search is conducted over all labels in the KG to gather candidate entities. If the token is a substring of a label, the entity is added. Each token sequence gets then a score assigned. The scoring is tackled from two sides. On the utterance side, a token-level context embedding and a character-level context embedding (based on the mention) is computed. The calculation is handled via dilated convolutional networks (DCNN) [134]. On the KG side, one includes the labels of candidate entity, the labels of relations connected to a candidate entity, the embedding of the candidate entity itself, and embeddings of the entities and relations related to the candidate entity. This is again done by DCNNs and, additionally, by fully connected layers. The best solution is then found by calculating a ranking and mention score for each token for each possible tokenization of the utterance. All those scores are then summed up into a global score. The global assignment with the highest score is then used to select the entity mentions and entity candidates. The approach uses the underlying graph, label and alias information of Wikidata. Graph information is used via connected entities and relations. They also use TransE embeddings, and therefore no hyper-relational structure. Due to the usage of static graph embeddings, retraining will be necessary if Wikidata changes.

OpenTapioca [23] is a mainly statistical EL approach published in 2019. While the paper never mentions ER, the approach was evaluated with it. In the code one can see that the ER is done by a SolrTextTagger analyzer of the Solr search platform[3]. The candidates are generated by looking up if the mention corresponds to an entity label or alias in Wikidata stored in a Solr collection. Entities are filtered out which do not correspond to the type person, location or organization. OpenTapioca is based on two main features, which are local compatibility and semantic similarity. First, local compatibility is calculated via a popularity measure and a unigram similarity measure between entity label and mention. The popularity measure is based on the number of sitelinks, PageRank scores, and the number of statements. Second, the semantic similarity strives to include context information in the decision process. All entity candidates are included in a graph and are connected via weighted edges. Those weights are calculated via a statistical similarity measure. This measure includes how likely it is to jump from one entity candidate to another while discounting it by the distance between the corresponding mentions in the utterance. The resulting adjacency matrix is then normalized to a stochastic matrix that defines a Markov Chain. One now propagates the local compatibility using this Markov Chain. Several iterations are then taken, and a final score is inferred via a Support Vector Machine. It supports multiple entities per utterance. OpenTapioca is only trained on and evaluated for three types of entities: locations, persons, and organizations. It facilitates Wikidata-specific labels, aliases, and sitelinks information. More importantly, it also uses qualifiers of statements in the calculation of the PageRank scores. But the qualifiers are only seen as additional edges to the entity. The usage in special domains is limited due to its restriction to only three types of entities but this is just an artificial restriction. It is easily updatable if the Wikidata graph changes as no immediate retraining is necessary.

Falcon 2.0 [98] is a fully linguistic approach and a transformation of Falcon 1.0 [97] to Wikidata. Falcon 2.0 was published in 2019 and its focus lies on short texts, especially questions. It links entities and relations jointly. Falcon 2.0 uses entity and relation labels as well as the triples itself. The relations and entities are recognized by applying linguistic principles. The candidates are then generated by comparing mentions to the labels using the Levenshtein distance. The ranking of the entities and relations is done by creating triples between the relations and entities and checking if the query is successful. The more successful the queries, the higher the candidate will be ranked. If no query is successful, the algorithm returns to the ER phase and splits some of the recognized entities again. As Falcon 2.0 is an extension of Falcon 1.0 from DBpedia to Wikidata, the usage of specific Wikidata characteristics is limited. Falcon 2.0 is tuned for EL on questions and short texts, as well as the English language. It is thus not very generalizable on longer, more noisy, non-question texts. As it only based on rules it is clearly independent of changes in the KG.

Arjun [80] tries to tackle specific challenges of Wikidata like long entity labels and implicit entities. Published in 2020, Arjun is an end-to-end approach utilizing the same model for ER and EL. It is based on an Encoder-Decoder-Attention model. First, the entities are detected via feeding Glove [84] embedded tokens of the utterance into the model and classifying each token as being an entity or not. Afterward, candidates are generated in the same way as in Falcon 2.0 [98]. The candidates are then ranked by feeding the mention, the entity label, and its aliases into the model and calculating the score. Thus, the model is a similarity measure between the mention and the entity labels. It does not use any global ranking. Wikidata information is used in the form of labels and aliases in the candidate generation and candidate ranking. As it relies purely on labels, it is not that susceptible to changes in the KG.

Hedwig [61] is a multilingual entity linker specialized on the TAC 2017 task but published in 2020. Another entity linker [62], developed by the same authors, is not included in this survey as Hedwig is partly an evolution of it. The entities to be linked are limited to only a subset of all possible entity classes. Hedwig employs Wikidata and Wikipedia at the same time. The Entity Recgontion uses word embeddings, character embeddings, and dictionary features where the character embeddings are calculated via a Bi-LSTM. The dictionary features are class-dependent, but this is not defined in more detail. Those embeddings and features are computed and concatenated for each token. Afterward, the whole sequence of token features is fed into a Bi-LSTM with a linear chain Conditional Random Field (CRF) layer at the end to recognize the entities. The

[3]https://lucene.apache.org/solr/

candidates for each detected entity mention are then generated by using a mention dictionary. The dictionary is created from Wikidata and Wikipedia information, utilizing labels, aliases, titles or anchor texts. The candidates are disambiguated by constructing a graph consisting of all candidate entities, mentions, and occurring words in the utterance. The edges between entities and other entities, words, or mentions have the normalized pointwise mutual information (NPMI) assigned as their weights. The NPMI specifies how frequent two entities, an entity and a mention or an entity and a word, occur together. Those scores are calculated over a Wikipedia dump. Finally, the PageRank of each node in the graph is calculated via power iteration, and the highest-scoring candidates are chosen. In contrast to DeepType, the type classification is used to determine the types of entities, not mentions. As this is only relevant for the TAC2017 task, the classifier can be ignored. Labels and aliases of multiple languages are used. It also uses sitelinks to connect the Wikidata identifiers and Wikipedia articles. The paper also claims to use descriptions but does not describe anywhere in what way. No hyper-relational or graph features are used. As it employs class-dependent features, it is limited to the entities of classes specified in the TAC 2017 task. The NPMI weights have to be updated with the addition of new elements in Wikidata and Wikipedia.

KBPearl [70], published in 2020, utilizes EL to populate incomplete KGs using documents. First, a document is preprocessed via Tokenization, POS tagging, NER, noun-phrase chunking, and time tagging. Also, an existing Information Extraction tool is used to extract open triples from the document. Open triples are non-linked triples in unstructured text. The triples are processed further by filtering invalid tokens and doing canonicalization. Then, a graph of entities, predicates, noun phrases, and relation phrases is constructed. The candidates are generated by comparing the noun/relation phrases to the labels and aliases of the entities/predicates. The edges between the entities/relations and between entities and relations are weighted by the number of intersecting one-hop statements. The next step is the computation of a maximum dense subgraph. Density is defined by the minimum weighted degree of all nodes [53]. As this problem is NP-hard, a greedy algorithm is used for optimization. New entities relevant for the task of Knowledge Graph Population are identified by thresholding the weighted sum of an entity's incident edges. Like used here, global coherence can perform sub-optimally since not all entities/relations in a document are related. Thus, two variants of the algorithm are proposed. First, a pipeline version that separates the full document into sentences. Second, a near neighbor mode, limiting the interaction of the nodes in the graph by the distances of the corresponding noun-phrases and relation-phrases. The approach includes label and alias information of entities and predicates. Additionally, one-hop statement information is used, but hyper-relational features are not mentioned. However, the paper does not claim that its focus is entirely on Wikidata. Thus, the weak specialization is understandable. While it utilizes EL, the focus of the approach is still knowledge base population. No training is necessary which makes the approach suitable for a dynamic graph like Wikidata.

PNEL [6] is an E2E model jointly solving ER and EL focused on short texts. PNEL employs a Pointer network [118] working on a set of different features. An utterance is tokenized into multiple different combinations. Each token is extended into the (1) token itself, (2) the token and the predecessor, (3) the token and the successor, and (4) the token with both predecessor and successor. For each token combination, candidates are searched for by using the BM25 similarity measure. Fifty candidates are used per tokenization combination. Therefore, 200 candidates are found per token. For each candidate, features are extracted. Those range from the simple length of a token to the graph embeddings of the candidate entity. All features are concatenated to a large feature vector. Therefore, per token, a sequence of 200 such features vectors exist. Finally, the concatenation of those sequences of each token in the sentence is then fed into a Pointer network. At each iteration of the Pointer network, it points to one candidate in the network or an `END` token marking no choice. The entity descriptions, labels and aliases are all used. Additionally, the graph structure is included by TransE graph embeddings, but no hyper-relational information was incorporated. E2E models often can improve the performance of the ER. Most EL algorithms employed in industry often use older ER methods decoupled from the EL process. Thus, such an E2E EL approach can be of use. Nevertheless, due to its reliance on static graph embeddings, complete retraining will be necessary if Wikidata changes.

The approach designed by Huang et al. [54] is utilizing deep and shallow models together. It specialized in short texts. The ER is performed via a pretrained BERT model [25] with a single classification layer on top, determining if a token belongs to an entity mention. The candidate search is done via an ElasticSearch[4] index, comparing the entity mention to labels and aliases by exact match and Levenshtein distance. The candidate ranking uses three similarity measures to calculate the final rank. A CNN is used to compute a character-based similarity between entity mention and candidate label. This results in a similarity matrix whose entries are calculated by the cosine similarity between each character embedding of both strings. The context is included in two ways. First, between the utterance and the entity description, by embedding the tokens of each sequence through a BERT model. Again, a similarity matrix is built by calculating the cosine similarity between each token embedding of both utterance and description. The KG is also considered by including the triples containing the candidate as a subject. For each such a triple a similarity matrix is calculated between the label concatenation of the triple and the utterance. All measures are then combined and fed into a two-layer perceptron. Wikidata labels, aliases and descriptions are utilized. Additionally, the KG structure is incorporated through the labels of candidate-related triples. This is similar to the approach by Mulang et al. [81], but only 1-hop triples are used. There are also no hyper-relational information considered. Due to its reliance on text alone, it is less susceptible to the changes of Wikidata.

In connection to the *CLEF 2020 HIPE challenge* [28], multiple approaches for ER and EL of historical newspapers on Wikidata were developed. Documents were available in English, French and German. Three approaches with a focus on the English language are described in the following. The documents are noisy as the OCR method for transcribing the newspapers produced errors. The authors often constructed different methods for different languages. From now on, only the English models are described. Differences in the usage of Wikidata between the languages did not exist. Yet, the approaches were not multilingual as different models were used and/or a retraining was necessary for different languages.

Boros et al. [13] tackled ER by using a BERT model with a CRF layer on top, which recognizes the entity mentions and classifies the type. During the training, the regular sentences are enriched with misspelled words to make the model robust against noise. For the EL, a knowledge base is built from Wikipedia, containing Wikipedia titles, ids, disambiguation pages, redirects and calculating link probability between mentions and Wikipedia pages. The link probability between anchors and Wikipedia pages is used to gather entity candidates for a mention. The disambiguation approach follows an already existing method [64]. Here, the utterance tokens are embedded via a Bi-LSTM. The token embeddings of a single mention are combined. Then similarity scores between the resulting mention embedding and the entity embeddings of the candidates are calculated. The entity embeddings are computed according to Ganea and Hofmann [40]. These similarity scores are combined with the link probability and long-range context attention, calculated by taking the inner product between an additional context-sensitive mention embedding and an entity candidate embedding. The resulting score is a local ranking measure and is again combined with a global ranking measure considering all other entity mentions in the text. In the end, additional filtering is applied by comparing the DBpedia types of the entities to the ones classified during the ER. If the type does not match or other inconsistencies apply, the entity candidate gets a lower rank. Here, they also experimented with Wikidata types, but this resulted in a performance decrease. As can be seen, technically, no Wikidata information besides the unsuccessful type inclusion is used. Thus, the approach resembles more of a Wikification algorithm. Yet, they do link to Wikidata as the HIPE task dictates it and therefore, the approach was included in the survey. New Wikipedia entity embeddings can be easily added [40] which is an advantage when Wikipedia changes. Also, its robustness against erroneous texts makes it ideal for real-world use.

Labusch and Neudecker [67] also applied a BERT model for ER. For EL, they used mostly Wikipedia, similar to Boros et al. [13]. They built a knowledge base containing all person, location and organization entities from the German Wikipedia. Then it was converted to an English knowledge base by mapping from the German Wikipedia Pages via Wikidata to the English ones. This mapping process resulted in the loss of numerous entities. The candidate generation is done by embedding all Wikipedia page titles in an Approximative Nearest Neighbour index. Using this index, the neighboring entities to the mention embedding are found and used as candidates. For ranking, anchor-contexts of Wikipedia pages are embedded and fed into a classifier together with the embedded mention-

[4]https://www.elastic.co/elasticsearch/

context, which outputs whether both belong to the same entity. This is done for each candidate for around 50 different anchor-contexts. Then, multiple statistics on those similarity scores and candidates are calculated, which are used in a Random Forest model to compute the final ranks. Similar to the previous approach, Wikidata was only used as the target knowledge base, while information from Wikipedia was used for all the EL work. Thus, no special characteristics of Wikidata were used. The approach is less affected by a change of Wikidata due to similar reasons as the previous approach. Also, this approach lacks performance compared to the state of the art in the HIPE task. The knowledge base creation process produces a disadvantageous loss of entities, but this might be easily changed.

Provatorov et al. [88] used an ensemble of fine-tuned BERT models for ER. The ensemble is used to compensate for the noise of the OCR procedure. The candidates were generated by using an ElasticSearch index filled with Wikidata labels. The candidate's final rank is calculated by taking the search score, increasing it if a perfect match applies and finally taking the candidate with the lowest Wikidata identifier number. They also created three other methods of the EL approach: (1) The ranking was done by calculating cosine similarity between the embedding of the utterance and the embedding of the same utterance with the mention replaced by the Wikidata description. Furthermore, the score is increased by the Levenshtein distance between the entity label and the mention. (2) A variant was used where the candidate generation is enriched with historical spellings of Wikidata entities. (3) The last variant used an existing tool, which included contextual similarity and co-occurrence probabilities of mentions and Wikipedia articles. Also, a global ranking was applied. The approach uses Wikidata labels and descriptions in one variant of candidate ranking. Beyond that, no other characteristics specific to Wikidata were considered. Overall, the approach is very basic and uses mostly pre-existing tools to solve the task. The approach is not susceptible to a change of Wikidata as it is mainly based on language and does not need retraining. However, its poor performance in the HIPE challenge makes it a less desirable method to employ.

Tweeki [48] is an approach focusing on unsupervised EL over tweets. The ER is done by a pre-existing Entity Recognizer [41] which also tags the mentions. The candidates are generated by first calculating the link probability between Wikidata aliases over Wikipedia and then searching for the aliases in a dictionary. The ranking is done using the link probabilities while pruning all candidates that do not belong to the type provided by the Entity Recognizer. It is a relatively simple approach that does not need to be trained, making it very suitable for linking entities in tweets. In that document type, often novel entities with minimal context exist. Regarding features of Wikidata, it uses label, alias and type information. Due to it being unsupervised, changes to the KG do not affect it.

5.2. Evaluation

Table 8 and Table 9 give an overview of all available results for the approaches described in the previous section. The first gives information for EL only approaches and the second for approaches evaluating EL together with ER. The micro F_1 scores are given:

$$F_1 = 2 \cdot \frac{p \cdot r}{p + r}$$

where p is the precision $p = \frac{tp}{tp+fp}$ and r is the recall $r = \frac{tp}{tp+fn}$. *tp* are here the amount of true positives, *fp* the amount of false positives and *fn* the amount of false negatives. Micro F_1 means that the scores are calculated over all linked entity mentions and not separately for each document and then averaged. True positives are the correctly linked entity mentions, false positives incorrectly linked entities which do not occur in the set of valid entities and false negatives entities which occur in the set of valid entities but are not linked to [19]. The approaches were evaluated on many different datasets, which makes comparison very difficult. Additionally, many approaches are evaluated on datasets designed for knowledge graphs different to Wikidata and then mapped. Often, the approaches are evaluated on the same dataset but over different subsets, which complicates a comparison even more. The method by Perkins [85] was also evaluated on the Kensho Derived Wikimedia Dataset [58], but it was only used to compare different variants of the designed approach and focussed on different amounts of training data. Thus, inclusion in the evaluation table is not reasonable.

Inferring the utility of a Wikidata characteristic from the different approaches' F_1-measures is inconclusive due to the sparsity of results. For EL-only, AIDA-CoNLL results are available for three of five approaches, but the results for two are the accuracies in-

Table 8: Results: EL only.

	DeepType [89] [1]	Mulang et al. [81]	LSH-ELMo model [85]	NED using DL on Graphs [18] [2]	Botha et al. [14]
AIDA-CoNLL [53]	0.949 [89] [3]	0.9494 [81] [3,4]	0.73 [85]	-	-
ISTEX-1000 [23]	-	0.9261 [81] [5]	-	-	-
Wikidata-Disamb [18]	0.924 [89] [3]	0.9235 [81] [6]	-	0.916 [18]	-
Mewsli-9 [14]	-	-	-	-	0.91 [14] [7]

[1] Only evaluated on Wikipedia
[2] Model with best result
[3] Accuracy instead of F_1
[4] DCA-SL used
[5] XLNet used
[6] Roberta used
[7] Recall instead of F_1

stead of the F_1-measures. However, considering the results of Deeptype [89] for Wikidata-Disamb, it becomes apparent that the inclusion of type information might help a lot. Still, it was only used with Wikipedia categories. The available labels for each item and property make language-model-based approaches possible that perform quite well [81]. No approaches are available to compare to the one by Botha et al. [14], but the result demonstrates the promising performance of multilingual EL with Wikidata as the target KG. For ER + EL approaches, most results were available for LC-QuAD 2.0. Yet, no conclusion can be drawn as many approaches were evaluated on different subsets of the dataset. Falcon 2.0 performs well, but it does not substantially rely on Wikidata characteristics. The performance is good as it is designed for simple questions that follow its rules very closely. Arjun performs well on T-REx by mainly using label information, but the amount of methods tested on the T-REx dataset is too low to be conclusive. Besides that, PNEL and the approach by Huang et al. also achieve good results; both include a broader scope of Wikidata information in the form of labels, descriptions and graph structure. As HIPE challenge approaches are using Wikidata only marginally and the difference in performance depends more on the robustness against the OCR-introduced noise, comparing them is not providing information on the relevance of Wikidata characteristics.

While some algorithms [80] do try to examine the challenges of Wikidata, like more noisy long entity labels, many fail to use most of the advantages of Wikidata's structure. If the approaches are using even more information than just the labels of entities and relations, they mostly only include simple n-hop triple information. Hyper-relational information like qualifiers is only used by OpenTapioca but still in a simple manner. This is surprising, as they can provide valuable additional information. As one can see in Figure 8, around half of the statements on entities occurring in the LC-QuAD 2.0 dataset have one or more qualifiers. These percentages differ from the ones in all of Wikidata, but when entities are considered, appearing in realistic use cases like QA, qualifiers are much more abundant. Thus, dismissing the qualifier information might be critical. The inclusion of hyper-relational graph embeddings could improve the performance of many approaches already using non-hyper-relational ones. Rank information of statements might be useful to consider, but choosing the best one will probably often suffice.

Of all approaches, only two algorithms [6, 54] use descriptions explicitly. Others incorporate them through triples too, but more on the side [81]. Descriptions can provide valuable context information and many items do have them; see Figure 6d. Hedwig [61] claims to use descriptions but fails to describe how. Three approaches [14, 61, 89] demonstrated the usefulness of the inherent multilingualism of Wikidata, notably in combination with Wikipedia.

As Wikidata is always changing, approaches robust against change are preferred. A reliance on transductive graph embeddings [6, 18, 85, 104], which need to have all entities available during training, makes repeated retraining necessary. Alternatively, the used

Table 9: Results: ER + EL.

	OpenTapioca [23]	Falcon 2.0 [98]	Arjun [80]	VCG [104]	KBPearl [70] [1]	PNEL [6]	Huang et al. [54]	Boros et al. [13]	Provatorov et al. [88]	Labusch & Neudecker [67]	Hedwig [61]	Tweeki [48]
AIDA-CoNLL [53]	0.482 [23]	-	-	-	-	-	-	-	-	-	-	-
Microposts 2016 [121]	0.087 [23], 0.148 [48]	-	-	-	-	-	-	-	-	-	-	0.248 [48]
ISTEX-1000 [23]	0.87 [23]	-	-	-	-	-	-	-	-	-	-	-
RSS-500 [93]	0.335 [23]	-	-	-	-	-	-	-	-	-	-	-
LC-QuAD 2.0 [27]	0.301 [6]	0.445 [6]	-	0.47 [6]	-	0.589 [6] [2]	-	-	-	-	-	-
LC-QuAD 2.0 [27] [3]	0.25 [98]	0.68 [98]	-		-	-	-	-	-	-	-	-
LC-QuAD 2.0 [27] [4]	-	0.320 [6]	-	-	-	0.629 [6] [2]	-	-	-	-	-	-
Simple-Question	0.20 [6]	0.41 [6]	-	-	-	0.68 [6] [5]	-	-	-	-	-	-
Simple-Question [12] [6]	-	0.63 [98]	-	-	-	-	-	-	-	-	-	-
T-REx [34]	0.579 [80]	-	0.713 [80]	-	-	-	-	-	-	-	-	-
T-REx [34] [7]	-	-	-	-	0.421 [70]	-	-	-	-	-	-	-
WebQSP [133]	-	-	0.730 [6, 104]	-	-	0.712 [6] [8]	0.780 [54]	-	-	-	-	-
CLEF HIPE 2020 [29]	-	-	-	-	-	-	-	0.531 [28] [9]	0.300 [28] [9]	0.141 [28] [9]	-	-
TAC2017 [56]	-	-	-	-	-	-	-	-	-	-	0.582 [61]	
Graph-Questions [107]	-	-	0.442 [104]	-	-	-	-	-	-	-	-	-
QALD-7-WIKI [113]	-	-	-	-	0.679 [70]	-	-	-	-	-	-	-
NYT2018 [69, 70]	-	-	-	-	0.575 [70]	-	-	-	-	-	-	-
ReVerb38 [70]	-	-	-	-	0.653 [70]	-	-	-	-	-	-	-
Knowledge Net [22]	-	-	-	-	0.384 [70]	-	-	-	-	-	-	-
TweekiGold [48]	0.291 [48]	-	-	-	-	-	-	-	-	-	-	0.65 [48]
Derczynski [24]	0.14 [48]	-	-	-	-	-	-	-	-	-	-	0.371 [48]

[1] NN model
[2] L model
[3] 1000 sampled questions from LC-QuAD 2.0
[4] LC-QuAD 2.0 test set used in KBPearl paper
[5] S model
[6] Probably evaluated on train and test set
[7] Evaluation on subset of T-REx data different to the subset used in Arjun paper
[8] W model
[9] Strict mention matching

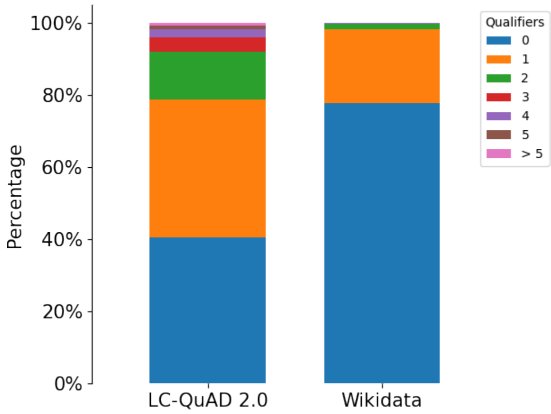


Fig. 8. Percentage of statements having the specified number of qualifiers for all LC-QuAD 2.0 and Wikidata entities.

embeddings would need to be replaced with graph embeddings, which are efficiently updatable or inductive [3, 5, 47, 109, 122, 123, 128]. The rule-based approach Falcon 2.0 [98] is not affected by a developing knowledge base but only usable for correctly-stated questions. Methods only working on text information [54, 80, 81] like labels, descriptions or aliases do not need to be updated if Wikidata changes, only if the text type or the language itself does. For approaches [48, 54, 61] that rely on statistics over Wikipedia, new entities may in Wikidata may sometimes not exist in Wikipedia to a satisfying degree. The approaches by Boros et al. [13], and Labusch and Neudecker [67] are mostly using Wikipedia information. They are, therefore, susceptible to changes in Wikipedia, especially specific statistics calculated over Wikipedia pages. Botha et al. [14] also mainly depends on Wikipedia and thus on the availability of the desired Wikidata entities in Wikipedia itself. But as it uses Wikipedia articles in multiple languages, it encompasses many more entities than the previous approaches that focus on Wikipedia. As it was designed for the zero- and few-shot setting, it is quite robust against changes in the underlying knowledge base. Deeptype [89] relies on a fine-grained type system. As the categories of Wikidata are not evolving as fast as novel entities appear, it is relatively robust against a changing knowledge base. However, it was not yet tested on Wikidata, which's type assignments differs vastly from Wikipedia. Statistical approaches [23, 70] need to update the underlying statistics, but this might be efficiently doable. Overall, the robustness against change is most negatively affected by static/transductive graph embeddings.

This summary and evaluation of the existing Wikidata Entity Linkers answers **RQ 1**.

5.3. *Reproducibility*

Not all algorithms are available as an Web API or even as source code. An overview can be seen in Table 10. The amount of approaches for Wikidata having

Table 10: Availability of approaches.

Approach	Code	Web API
OpenTapioca [23]	✓	✓
NED using DL on Graphs [18]	✓	✗
Falcon 2.0 [98]	✓	✓
Arjun [80]	✓	✗
DeepType [89]	✓	✗
Hedwig [61]	✗	✗
VCG [104]	✓	✗
KBPearl [70]	✗	✗
PNEL [6]	✓	✗
Mulang et al. [81]	✓	✗
Perkins [85]	✗	✗
Huang et al. [54]	✗	✗
Boros et al. [13]	✗	✗
Provatorov et al. [88]	✗	✗
Labusch and Neudecker [67]	✗	✗
Botha et al. [14]	✗	✗
Tweeki [48]	✗	✗

an accessible Web API is meager. While the code for some methods exists, this is still just the case for less than half. The effort to set up the different approaches also varies significantly due to missing instructions or data. Thus, we refrained from evaluating and filling the missing results for all the datasets in Tables 8 and 9.

6. Datasets

6.1. *Overview*

This section is concerned with analyzing the different datasets which are used for Wikidata EL. A comparison can be found in Table 11. Here, information about the purpose, release year, domain and more is given. The majority of datasets on which existing Entity linkers were evaluated, were originally constructed for KGs different from Wikidata. Such a mapping

Table 11: Comparison of used datasets.

Dataset	Domain	Year	Purpose	Annotated mentions	Identifiers
ISTEX-1000 [23]	Research articles	2019	EL	✓	Wikidata
Wikidata-Disamb [18] (based on Wiki-Disamb30[38])	Wikipedia articles	2018 [1]	EL	✗	Wikidata [2]
LC-QuAD 2.0 [27]	General complex questions	2019	Question Answering (QA)	✗	DBpedia, Wikidata
T-REx [34]	Wikipedia abstracts	2015	Knowledge Base Population (KBP), Relation Extraction (RE), Natural Language Geneation (NLG)	✓	Wikidata
Knowledge Net [22]	Wikipedia abstracts, biographical texts	2019	KBP	✓	Wikidata
NYT2018 [69, 70]	News	2018	EL	✓	Wikidata, DBpedia
KORE50DYWC [82]	News	2019	EL	✓	Wikidata, DBpedia, YAGO, Crunchbase
Kensho Derived Wikimedia Dataset [58]	Wikipedia	2020	Natural Language Processing (NLP)	✓	Wikidata, Wikipedia
CLEF HIPE 2020 [29]	Historical newspapers	2020	ER, EL	✓	Wikidata
Mewsli-9 [14]	News in multiple languages	2020	Multilingual EL	✓	Wikidata
TweekiData [48]	Tweets	2020	EL	✓	Wikidata
TweekiGold [48]	Tweets	2020	EL	✓	Wikidata

[1] data from 2010

[2] Original dataset on Wikipedia

can be problematic as some entities labeled for other KGs could be missing in Wikidata. Or some NIL entities that do not exist in other KGs could exist in Wikidata. Eleven datasets were found for which Wikidata [14, 22, 23, 27, 29, 34, 48, 58, 70, 82] identifiers were available from the start.

LC-QuAD 2.0 [27] is a dataset semi-automatically created for Complex Questions Answering providing complex natural language questions. For each question, Wikidata and DBpedia identifiers are provided. The questions are generated from subgraphs of the Wikidata KG. The dataset does not provide annotated mentions.

T-REx [34] was constructed automatically over Wikipedia abstracts. Its main purpose is Knowledge Base Population. According to Mulang et al. [80], this dataset describes the challenges of Wikidata, at least in the form of long, noisy labels, the best.

ISTEX-1000 [23] is a research-focused dataset containing 1000 author affiliation strings. It was manually annotated to evaluate the OpenTapioca entity linker.

KnowledgeNet [22] is a Knowledge Base Population dataset with 9073 manually annotated sentences. The text was extracted from biographical documents from the web or Wikipedia articles.

NYT2018 [69, 70] consists of 30 news documents that were manually annotated on Wikidata and DBpedia. It was constructed for KBPearl, so its main focus is also KBP which is a downstream task of EL.

One dataset, KORE 50 DYWC [82], was found, which was not used by any of the approach papers. It is an annotated EL dataset based on the KORE50 dataset, a manually annotated subset of the AIDA corpus. All sentences are reannotated with DBpedia, Yago, Wikidata and Crunchbase entities.

The Kensho Derived Wikimedia Dataset [58] is an automatically created condensed subset of Wikimedia data. It consists of three levels: Wikipedia text, annotations with Wikipedia pages and links to Wikidata items. Thus, mentions in Wikipedia articles are annotated with Wikidata items. However, as some Wikidata items do not have a corresponding Wikipedia page, the annotation is not exhaustive. It was constructed for NLP in general.

CLEF HIPE 2020 [29] is a dataset based on historical newspapers in English, French and German. Only the English dataset will be analyzed in the following. This dataset is of great difficulty due to many errors in the text, which originates from the OCR method used to parse the scanned newspapers. For the English language, only a development and test set exist. In the other two languages, a training set is also available. It was manually annotated.

Mewsli-9 [14] is a multilingual dataset automatically constructed from WikiNews. It includes nine different languages. A high percentage of entity mentions in the dataset do not have corresponding English Wikipedia pages, and thus, cross-lingual linking is necessary.

TweekiData and TweekiGold [48] are an automatically annotated corpus and a manually annotated dataset for EL over tweets. TweekiData was created by using other existing tweet-based datasets and linking them to Wikidata data via the Tweeki EL. TweekiGold was created by an expert, manually annotating tweets from another dataset with Wikidata identifiers and Wikipedia page-titles.

Table 13 shows the number of documents, the number of mentions, emerging entities and unique entities, and the mentioned ratio. What classifies as a document in a dataset depends on the dataset itself. For example, for T-REx, a document is a whole paragraph of a Wikipedia article, while for LC-QuAD 2.0, a document is just a single question. Due to this, the average amount of entities in a document also varies, e.g., LC-QuAD 2.0 with 1.47 entities per document and T-REx with 11.03. If a dataset was not available, information from the original paper was included. If dataset splits were available, the statistics are also shown separately. The majority of datasets do not contain emerging entities. For the Tweeki datasets, it is not mentioned which Wikidata dump was used to annotate. For a dataset that contains emerging entities, this is problematic. On the other hand, the dump is specified for the CLEF HIPE 2020 dataset, making it possible to work on the Wikidata version with the correct entities missing.

To get an overview how widespread they datasets are in use, see the section 5.2. Thus, **RQ 3** is answered.

6.2. Evaluation

The difficulty of the different datasets was measured by the accuracy of a simple EL method (Table 14) and the ambiguity of mentions (Table 12). The simple EL method searches for entity candidates via an ElasticSearch index, including all English labels and aliases. It then disambiguates by taking the one with the largest

Table 12: Ambiguity of mentions (existence of a match does not correspond to a correct match).

Dataset	**Average number of matches**	**No match**	**Exact match**	**More than one match**
ISTEX-1000 (train)	23.23	8.06%	26.34%	65.61%
ISTEX-1000 (test)	25.85	10.30%	23.88%	65.82%
Wiki-Disamb30 (train)	25.06	0.36%	1.26%	98.38%
Wiki-Disamb30 (dev)	30.39	0.40%	1.18%	98.42%
Wiki-Disamb30 (test)	30.18	0.30%	1.44%	98.26%
Knowledge Net (train)	21.90	10.41%	22.29%	67.3%
T-REx	4.79	31.36%	32.98%	35.65%
KORE50DYWC	28.31	3.93%	7.49%	88.60%
Kensho Derived Wikimedia Dataset	8.16	35.18%	30.94%	33.88%
CLEF HIPE 2020 (en, dev)	24.02	35.71%	11.51%	52.78%
CLEF HIPE 2020 (en, test)	17.78	43.82%	6.74%	49.44%
Mewsli-9 (en)	11.09	16.80%	34.90%	47.30%
TweekiData	19.61	19.98%	12.01%	68.01%
TweekiGold	16.02	7.41%	20.25%	72.34%

tf-idf based BM25 similarity measure score and the lowest Q-identifier number resembling the popularity. Nothing was done to handle inflections.[5] Here, only datasets were included which were accessible. As one can see, is the accuracy positively correlated with the number of exact matches. The more ambiguous the underlying entity mentions are, the more inaccurate a simple similarity measure between label and mention becomes. In this case, more context information is necessary. The simple Entity Linker was only applied to datasets that were feasible to disambiguate in that way. T-REx and the Kensho Derived Wikimedia Dataset were too large. According to the EL performance, ISTEX-1000 is the easiest dataset. Many of the ambiguous mentions reference the most popular one, while also many exact unique matches exist. T-REx, the Kensho Derived Wikimedia Dataset and the Mewsli-9 training dataset have the largest percentage of exact matches for labels. The largest number of ambiguous mentions have the Wiki-Disamb30 datasets, resulting in a low EL but not the lowest accuracy. Deciding on the most prominent entity appears to produce good EL results. This is also the case for the TweekiGold dataset. While the KORE50DYWC dataset is less ambiguous than Wiki-Disamb30, it performs the worst due to references to unpopular entities. The CLEF HIPE 2020 dataset also has a low EL accuracy but not due to ambiguity but many mentions with no exact match. The reason for that is the noise created by OCR. Only the English dataset was examined. The second column of Table 14 specifies the accuracy with all unique exact matches removed. This is based on the intuition that exact matches without any competitors are usually correct. In general, the removal does decrease the accuracy with one exception. The Wiki-Disamb30 datasets constantly achieve better accuracy as a large percentage of the unique exact matches appear to point to wrong entities. Thus, the true entity does not have the label it is referenced by.

Two main characteristics of Wikidata may affect the design of Wikidata EL datasets. First, multilingualism is the main focus of Wikidata, and thus, multilingual datasets should also be a focus. Unfortunately, only two datasets [14, 29] focus on the multilingualism of Wikidata. The CLEF HIPE 2020 dataset is designed for Wikidata and has documents for the languages English, French and German, but each language has a different corpus of documents. The same is the case for the Mewsli-9 dataset, while here, documents in nine languages are available. A dataset similar to VoxEL [94], which is defined for Wikipedia, would be helpful. Here, each utterance is translated into multiple languages, which eases the comparison of the multilingual EL performance. Having the same corpus of documents in different languages would allow a better comparison of a method's performance in various languages. Of course, such translations will never be perfectly comparable.

[5]All source code, plots and results can be found on https://github.com/cedricm-research/ELEnglishWD

Table 13: Comparison of the datasets with focus on the number of documents and Wikidata entities.

Dataset	# documents	# mentions	Emerging entities	Wikidata entities	Unique Wikidata entities	Mentions per document
ISTEX-1000 [23] (train)	750	2073	0%	100%	53.7%	2.76
ISTEX-1000 [23] (test)	250	670	0%	100%	65.8%	2.68
Wikidata-Disamb [18] (train)	100,000	100,000	0%	100%	27.2%	1.0
Wikidata-Disamb [18] (test)	10,000	10,000	0%	100%	57.3%	1.0
Wikidata-Disamb [18] (dev)	10,000	10,000	0%	100%	56.2%	1.0
LC-QuAD 2.0 [27]	6046	44,529	0%	100%	51.2%	1.47
T-REx [34]	4,650,000	51,297,484	0%	100%	9.1%	11.03
Knowledge Net [22] (train)	3977	13,039	0%	100%	30%	3.28
NYT2018 [69, 70]	30	-	-	-	-	-
KORE50DYWC [82]	50	307	0%	100%	72.0%	6.14
Kensho Derived Wikimedia Dataset [58]	14,255,258	121,835,453	0%	100%	3.7%	8.55
CLEF HIPE 2020 (en, dev) [29]	80	470	46.4%	53.6%	31.9%	5.88
CLEF HIPE 2020 (en, test) [29]	46	134	33.6%	66.4%	42.5%	2.91
Mewsli-9 (en) [14]	12,679	80,242	0%	100%	48.2%	6.33
TweekiData [48]	5,000,000	5,038,870	61.2%	38.8%	5.4%	1.01
TweekiGold [48]	500	958	11.1%	88.9%	66.6%	1.92

Table 14: EL accuracy - Kensho Derived Wikimedia Dataset, T-REx and TweekiData are not included due to size, **Acc. filtered** has all exact matches removed.

Dataset	Acc.	Acc. filtered
ISTEX-1000 (train)	0.744	0.716
ISTEX-1000 (test)	0.716	0.678
Wiki-Disamb30 (train)	0.597	0.600
Wiki-Disamb30 (dev)	0.580	0.584
Wiki-Disamb30 (test)	0.576	0.580
Knowledge Net (train)	0.371	0.285
KORE50DYWC	0.225	0.187
CLEF HIPE 2020 (en, dev)	0.333	0.287
CLEF HIPE 2020 (en, test)	0.258	0.241
TweekiGold	0.565	0.520
Mewsli-9 (en)	0.602	0.490

The second characteristic is the large rate of change of Wikidata. Thus, it would also be advisable that the datasets specify the Wikidata dumps they were created, similar to Petroni et al. [87]. Many of the existing datasets do that, yet not all. In current dumps, entities, which were available while the dataset was created, could have been removed. It is even more probable that emerging entities could now have a corresponding entity in an updated Wikidata dump version. If the EL approach now would detect it as an emerging entity, it is evaluated as correct, but in reality, this is false and vice versa. Concerning emerging entities, another variant of an EL dataset could be useful. Two Wikidata dumps from different time points could be used to label the utterances. Such a dataset would be valuable in the context of Knowledge Graph Population when emerging entities are inserted into the KG. With the true emerging entity available, one could measure the quality of the insertion. Also, constraining that the method needs to perform well on both KG dumps would force EL approaches to be less reliant on a fixed graph structure. This answers **RQ 4**.

7. Related work

While there are multiple recent surveys on EL, none of those are specialized in analyzing the area of EL on Wikidata.

The extensive survey by Sevgili et al. [99] is giving an overview of all neural approaches from 2015 to 2020. It compares 30 different approaches on nine different datasets. Of those, only Deeptype can be seen as focused on Wikidata. The survey also discusses the current state of the art of domain-independent and multilingual neural EL approaches. However, the influence of the underlying KG was not of concern to the authors. It is not described in detail how they found the considered approaches.

In the survey by Al-Moslmi et al. [2], the focus lies on ER and EL approaches over KGs in general. It considers approaches from 2014 to 2019. It gives an overview of the different approaches of ER, Entity Disambiguation, and EL. A distinction between Entity Disambiguation and EL is made, while our survey sees Entity Disambiguation as a part of EL. The roles of different domains, text types, or languages are discussed. The authors considered 89 different approaches and tools. Most approaches were designed for DBpedia or Wikipedia, some for Freebase or YAGO, and some to be KG-agnostic. Again, the only Wikidata contender was Deeptype. F_1 scores were gathered on 17 different datasets. Fifteen algorithms, for which an implementation or a WebAPI was available, were evaluated using GERBIL [92].

Another survey [83] examines recent approaches, which employ holistic strategies. Holism in the context of EL is defined as the usage of domain-specific inputs and metadata, joint ER-EL approaches and collective disambiguation methods. Thirty-six research articles were found which had any holistic aspect - none of the designed approaches linked explicitly to Wikidata.

A comparison of the number of approaches and datasets included in the different surveys can be found in Table 15.

If we go further into the past, the existing surveys [72, 102] are not considering Wikidata at all or only in a small amount as it is still a rather recent KG in comparison to the other established ones like DBpedia, Freebase or YAGO. For an overview on different KGs on the web, we refer the interested reader to the one by Heist et al. [50].

No found survey focused on the differences of EL over different knowledge graphs, respectively, on the particularities of EL over Wikidata.

Table 15: Survey Comparison

Survey	# Approaches	# Wikidata Approaches	# Datasets	# Wikidata Datasets
Sevgili et al. [99]	30	1	9	0
Al-Moslmi et al. [2]	39	1	17	0
Oliveira et al. [83]	36	0	32	0
This survey	17	17	21	11

8. Discussion

8.1. Current Approaches, Datasets and their Drawbacks

Approaches. The number of algorithms using Wikidata is small; the number of algorithms using Wikidata solely is even smaller. Most algorithms employ labels and alias information contained in Wikidata. Some deep learning-based algorithms leverage the underlying graph structure, but the inclusion of that information is often superficial. The same information is also available in other KGs. Additional statement specific information like qualifiers is used by only one algorithm (OpenTapioca), and even then, it only interprets qualifiers as extra edges to the item. Thus, there is no inclusion of the actual structure of a hyper-relation. Information like the descriptions of items which are providing valuable context information is also used seldom. Wikidata includes type information, but almost none of the existing algorithms utilize it to do more than to filter out entities that are not desired to link in general. An exception is Tweeki, which uses it together with ER, and perhaps DeepType, though the evaluated model used Wikipedia categories.

One could claim that the current algorithms are mostly trying to map algorithms also usable on other KGs to Wikidata. Besides utilizing specific characteristics of Wikidata, it is also notable that there is no clear focus on one of the essential characteristics of Wikidata, the continual growth. Many approaches use static graph embeddings, which need to be retrained if the KG changes. EL algorithms working on Wikidata, which are not usable on future versions, seem unintuitive. But there also exist some approaches which can handle change. They often rely on more extensive textual information, which is again challenging due to the limited amount of such data in Wikidata. Wikidata descriptions do exist, but only short paragraphs are provided, in general, insufficient to train a language model. To compensate, Wikipedia is included, which provides this textual information. It seems like Wikidata as the target KG with its language-agnostic identifiers and the easily connectable Wikipedia with its multilingual textual information are the perfect pair.

Most of the approaches tried to use Wikidata due to it being up to date while not utilizing its structure. With small adjustments, many would also work on any other KG. None of the investigated approaches tried to examine the performance between different versions of Wikidata. As continuous evolution is a central characteristic of Wikidata, a temporal analysis would be reasonable.

This survey aimed to identify the extent to which the current state of the art in Wikidata EL is utilizing the characteristics of Wikidata. As only a few are using more information than on other established KGs, there is still much potential for future research.

Datasets. Only a limited amount of datasets were created entirely with Wikidata in mind exist. Many datasets used are still only mapped versions of datasets created for other knowledge bases. Multilingualism is present so far that some datasets contain documents in different languages. However, only different documents for different languages are available. Having the same documents in multiple languages would be more helpful for an evaluation of multilingual Entity Linkers. The fact that the Wikidata is ever-changing is also not genuinely considered in any datasets. Always providing the dump version on which the dataset was created is advisable. Great is that datasets from very different domains like news, forums, research, tweets exist. The utterances can also vary from shorter texts with only a few entities to large documents with many entities. The difficulty of the datasets significantly differs in the ambiguity of the entity mentions. The datasets also differ in quality. Some were automatically created and others annotated manually by experts. There are no unanimously agreed upon datasets used for Wikidata EL. Of course, a single dataset can not exist as different domains and text types make different approaches, and hence datasets necessary.

8.2. Future Research Avenues

In general, Wikidata EL could be improved by including:

- Hyper-relational statements which provide additional information
- Type information for more than limiting the candidate space
- Inductive or efficently trainable knowledge graph embeddings
- Item label and description information in multiple languages for multilingual EL

The qualifier and rank information of Wikidata could be also suitable to do EL on time-sensitive utterances [1]. The problem evolves around utterances which talk about entities from different time points and spans and thus, the referred entity can significantly diverge.

The usefulness of other characteristics of Wikidata, e.g., references, may be limited but could make EL more challenging due to the inclusion of contradictory information. Therefore, research into the consequences and solutions of conflicting information would be advisable.

To reiterate, due to the fast rate of change of Wikidata, approaches are necessary, which are more robust to such a dynamic KG. Continuously retraining transductive embeddings is intractable, so more sophisticated methods like inductive or efficiently retrainable graph embeddings are a necessity.

Multilingual or cross-lingual EL is already tackled with Wikidata but currently mainly by depending on Wikipedia. Using the available multilingual label/description information in a structured form together with the rich textual information in Wikipedia could move the field forward.

It seems like there exist no commonly agreed on Wikidata EL datasets as shown by a large number of different datasets the approaches were tested on. Such datasets should try to represent the challenges of Wikidata like the time-variance, contradictory triple information, noisy labels, and multilingualism.

Acknowledgments

We acknowledge the support of the EU project TAILOR (GA 952215), the Federal Ministry for Economic Affairs and Energy (BMWi) project SPEAKER (FKZ 01MK20011A), the German Federal Ministry of Education and Research (BMBF) projects and excellence clusters ML2R (FKZ 01 15 18038 A/B/C), MLwin (01S18050 D/F), ScaDS.AI (01/S18026A) as well as the Fraunhofer Zukunftsstiftung project JOSEPH.

References

[1] P. Agarwal, J. Strötgen, L.D. Corro, J. Hoffart and G. Weikum, diaNED: Time-Aware Named Entity Disambiguation for Diachronic Corpora, in: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, I. Gurevych and Y. Miyao, eds, Association for Computational Linguistics, 2018, pp. 686–693. doi:10.18653/v1/P18-2109. https://www.aclweb.org/anthology/P18-2109/.

[2] T. Al-Moslmi, M.G. Ocaña, A.L. Opdahl and C. Veres, Named Entity Extraction for Knowledge Graphs: A Literature Overview, *IEEE Access* **8** (2020), 32862–32881. doi:10.1109/ACCESS.2020.2973928.

[3] M. Albooyeh, R. Goel and S.M. Kazemi, Out-of-Sample Representation Learning for Knowledge Graphs, in: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings, EMNLP 2020, Online Event, 16-20 November 2020*, T. Cohn, Y. He and Y. Liu, eds, Association for Computational Linguistics, 2020, pp. 2657–2666. https://www.aclweb.org/anthology/2020.findings-emnlp.241/.

[4] P.D. Almeida, J.G. Rocha, A. Ballatore and A. Zipf, Where the Streets Have Known Names, in: *Computational Science and Its Applications - ICCSA 2016 - 16th International Conference, Beijing, China, July 4-7, 2016, Proceedings, Part IV*, O. Gervasi, B. Murgante, S. Misra, A.M.A.C. Rocha, C.M. Torre, D. Taniar, B.O. Apduhan, E.N. Stankova and S. Wang, eds, Lecture Notes in Computer Science, Vol. 9789, Springer, 2016, pp. 1–12. doi:10.1007/978-3-319-42089-9_1.

[5] J. Baek, D.B. Lee and S.J. Hwang, Learning to Extrapolate Knowledge: Transductive Few-shot Out-of-Graph Link Prediction, in: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan and H. Lin, eds, 2020. https://proceedings.neurips.cc/paper/2020/hash/0663a4ddceacb40b095eda264a85f15c-Abstract.html.

[6] D. Banerjee, D. Chaudhuri, M. Dubey and J. Lehmann, PNEL: Pointer Network Based End-To-End Entity Linking over Knowledge Graphs, in: *The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part I*, J.Z. Pan, V.A.M. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne and L. Kagal, eds, Lecture Notes in Computer Science, Vol. 12506, Springer, 2020, pp. 21–38. doi:10.1007/978-3-030-62419-4_2. https://doi.org/10.1007/978-3-030-62419-4_2.

[7] A. Bastos, A. Nadgeri, K. Singh, I.O. Mulang, S. Shekarpour and J. Hoffart, RECON: Relation Extraction using Knowledge Graph Context in a Graph Neural Network, *CoRR* **abs/2009.08694** (2020). https://arxiv.org/abs/2009.08694.

[8] J. Berant, A. Chou, R. Frostig and P. Liang, Semantic Parsing on Freebase from Question-Answer Pairs, in: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, ACL, 2013, pp. 1533–1544. https://www.aclweb.org/anthology/D13-1160/.

[9] P. Bhargava, N. Spasojevic, S. Ellinger, A. Rao, A. Menon, S. Fuhrmann and G. Hu, Learning to Map Wikidata Entities To Predefined Topics, in: *Companion of The 2019 World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, S. Amer-Yahia, M. Mahdian, A. Goel, G. Houben, K. Lerman, J.J. McAuley, R. Baeza-Yates and L. Zia, eds, ACM, 2019, pp. 1194–1202. doi:10.1145/3308560.3316749.

[10] K.D. Bollacker, C. Evans, P. Paritosh, T. Sturge and J. Taylor, Freebase: a collaboratively created graph database for structuring human knowledge, in: *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008*, J.T. Wang, ed., ACM, 2008, pp. 1247–1250. doi:10.1145/1376616.1376746.

[11] A. Bordes, N. Usunier, A. García-Durán, J. Weston and O. Yakhnenko, Translating Embeddings for Modeling Multi-relational Data, in: *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, C.J.C. Burges, L. Bottou, Z. Ghahramani and K.Q. Weinberger, eds, 2013, pp. 2787–2795. http://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.

[12] A. Bordes, N. Usunier, S. Chopra and J. Weston, Large-scale Simple Question Answering with Memory Networks, *CoRR* **abs/1506.02075** (2015). http://arxiv.org/abs/1506.02075.

[13] E. Boros, E.L. Pontes, L.A. Cabrera-Diego, A. Hamdi, N. Sidère and A. Doucet, Robust Named Entity Recognition and Linking on Historical Multilingual Documents (2020).

[14] J.A. Botha, Z. Shan and D. Gillick, Entity Linking in 100 Languages, *CoRR* **abs/2011.02690** (2020). https://arxiv.org/abs/2011.02690.

[15] D.G. Brizan and A.U. Tansel, A. survey of entity resolution and record linkage methodologies, *Communications of the IIMA* **6**(3) (2006), 5.

[16] M. Burtsev, A. Seliverstov, R. Airapetyan, M. Arkhipov, D. Baymurzina, N. Bushkov, O. Gureenkova, T. Khakhulin, Y. Kuratov, D. Kuznetsov et al., Deeppavlov: Open-source library for dialogue systems, in: *Proceedings of ACL 2018, System Demonstrations*, 2018, pp. 122–127.

[17] J. Callan, M. Hoy, C. Yoo and L. Zhao, Clueweb09 data set, 2009.

[18] A. Cetoli, M. Akbari, S. Bragaglia, A.D. O'Harney and M. Sloan, Named Entity Disambiguation using Deep Learning on Graphs, *CoRR* **abs/1810.09164** (2018). http://arxiv.org/abs/1810.09164.

[19] M. Cornolti, P. Ferragina and M. Ciaramita, A framework for benchmarking entity-annotation systems, in: *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013*, D. Schwabe, V.A.F. Almeida, H. Glaser, R. Baeza-Yates and S.B. Moon, eds, International World Wide Web Conferences Steering Committee / ACM, 2013, pp. 249–260. doi:10.1145/2488388.2488411.

[20] H.-J. Dai, C.-Y. Wu, R.T.-H. Tsai and W.-L. Hsu, From entity recognition to entity linking: a survey of advanced entity linking techniques, in: *26 (2012)*, , 2012, pp. 3M2IOS3b1–3M2IOS3b1.

[21] DBpedia, DBpedia Live. https://wiki.dbpedia.org/online-access/DBpediaLive.

[22] F. de Sá Mesquita, M. Cannaviccio, J. Schmidek, P. Mirza and D. Barbosa, KnowledgeNet: A Benchmark Dataset for Knowledge Base Population, in: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, K. Inui, J. Jiang, V. Ng and X. Wan, eds, Association for Computational Linguistics, 2019, pp. 749–758. doi:10.18653/v1/D19-1069.

[23] A. Delpeuch, OpenTapioca: Lightweight Entity Linking for Wikidata, *CoRR* **abs/1904.09131** (2019). http://arxiv.org/abs/1904.09131.

[24] L. Derczynski, D. Maynard, G. Rizzo, M. van Erp, G. Gorrell, R. Troncy, J. Petrak and K. Bontcheva, Analysis of named entity recognition and linking for tweets, *Inf. Process. Manag.* **51**(2) (2015), 32–49. doi:10.1016/j.ipm.2014.10.006.

[25] J. Devlin, M. Chang, K. Lee and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran and T. Solorio, eds, Association for Computational Linguistics, 2019, pp. 4171–4186. doi:10.18653/v1/n19-1423.

[26] C. Dogan, A. Dutra, A. Gara, A. Gemma, L. Shi, M. Sigamani and E. Walters, Fine-Grained Named Entity Recognition using ELMo and Wikidata, *CoRR* **abs/1904.10503** (2019). http://arxiv.org/abs/1904.10503.

[27] M. Dubey, D. Banerjee, A. Abdelkawi and J. Lehmann, LC-QuAD 2.0: A Large Dataset for Complex Question Answering over Wikidata and DBpedia, in: *The Semantic Web - ISWC 2019 - 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part II*, C. Ghidini, O. Hartig, M. Maleshkova, V. Svátek, I.F. Cruz, A. Hogan, J. Song, M. Lefrançois and F. Gandon, eds, Lecture Notes in Computer Science, Vol. 11779, Springer, 2019, pp. 69–78. doi:10.1007/978-3-030-30796-7_-5. https://doi.org/10.1007/978-3-030-30796-7_5.

[28] M. Ehrmann, M. Romanello, A. Flückiger and S. Clematide, Extended overview of CLEF HIPE 2020: named entity processing on historical newspapers, in: *CLEF*, 2020a.

[29] M. Ehrmann, M. Romanello, A. Flückiger and S. Clematide, Extended overview of CLEF HIPE 2020: named entity processing on historical newspapers, in: *CLEF*, 2020b.

[30] C.B. El Vaigh, G. Le Noé-Bienvenu, G. Gravier and P. Sébillot, IRISA System for Entity Detection and Linking at CLEF HIPE 2020, in: *CEUR Workshop Proceedings*, 2020.

[31] R. Ellgren, Exploring Emerging Entities and Named Entity Disambiguation in News Articles, 2020a.

[32] R. Ellgren, Exploring Emerging Entities and Named Entity Disambiguation in News Articles, 2020b.

[33] J. Ellis, S. Strassel and J. Getman, *TAC KBP Evaluation Source Corpora 2016-2017*, Linguistic Data Consortium, University of Pennsylvania, 2019.

[34] H. ElSahar, P. Vougiouklis, A. Remaci, C. Gravier, J.S. Hare, F. Laforest and E. Simperl, T-REx: A Large Scale Alignment of Natural Language with Knowledge Base Triples, in: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018*, N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis and T. Tokunaga, eds, European Language Resources Association (ELRA), 2018. http://www.lrec-conf.org/proceedings/lrec2018/summaries/632.html.

[35] F. Erxleben, M. Günther, M. Krötzsch, J. Mendez and D. Vrandecic, Introducing Wikidata to the Linked Data Web, in: *The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I*, P. Mika, T. Tudorache, A. Bernstein, C. Welty, C.A. Knoblock, D. Vrandecic, P. Groth, N.F. Noy, K. Janowicz and C.A. Goble, eds, Lecture Notes in Computer Science, Vol. 8796, Springer, 2014, pp. 50–65. doi:10.1007/978-3-319-11964-9_4. https://doi.org/10.1007/978-3-319-11964-9_4.

[36] M. Färber, B. Ell, C. Menne and A. Rettinger, A comparative survey of dbpedia, freebase, opencyc, wikidata, and yago, *Semantic Web Journal* **1**(1) (2015), 1–5.

[37] M. Färber, F. Bartscherer, C. Menne and A. Rettinger, Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO, *Semantic Web* **9**(1) (2018), 77–129. doi:10.3233/SW-170275.

[38] P. Ferragina and U. Scaiella, TAGME: on-the-fly annotation of short text fragments (by wikipedia entities), in: *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010, Toronto, Ontario, Canada, October 26-30, 2010*, J. Huang, N. Koudas, G.J.F. Jones, X. Wu, K. Collins-Thompson and A. An, eds, ACM, 2010, pp. 1625–1628. doi:10.1145/1871437.1871689.

[39] M. Galkin, P. Trivedi, G. Maheshwari, R. Usbeck and J. Lehmann, Message Passing for Hyper-Relational Knowledge Graphs, *CoRR* **abs/2009.10847** (2020). https://arxiv.org/abs/2009.10847.

[40] O. Ganea and T. Hofmann, Deep Joint Entity Disambiguation with Local Neural Attention, in: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, M. Palmer, R. Hwa and S. Riedel, eds, Association for Computational Linguistics, 2017, pp. 2619–2629. doi:10.18653/v1/d17-1277.

[41] M. Gardner, J. Grus, M. Neumann, O. Tafjord, P. Dasigi, N.F. Liu, M.E. Peters, M. Schmitz and L. Zettlemoyer, AllenNLP: A Deep Semantic Natural Language Processing Platform, *CoRR* **abs/1803.07640** (2018). http://arxiv.org/abs/1803.07640.

[42] J. Geiß, A. Spitz and M. Gertz, NECKAr: A Named Entity Classifier for Wikidata, in: *Language Technologies for the Challenges of the Digital Age - 27th International Conference, GSCL 2017, Berlin, Germany, September 13-14, 2017, Proceedings*, G. Rehm and T. Declerck, eds, Lecture Notes in Computer Science, Vol. 10713, Springer, 2017, pp. 115–129. doi:10.1007/978-3-319-73706-5_10. https://doi.org/10.1007/978-3-319-73706-5_10.

[43] G.A. Gesese, M. Alam and H. Sack, Semantic Entity Enrichment by Leveraging Multilingual Descriptions for Link Prediction, *CoRR* **abs/2004.10640** (2020). https://arxiv.org/abs/2004.10640.

[44] M.R. Glass and A. Gliozzo, A Dataset for Web-Scale Knowledge Base Population, in: *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, A. Gangemi, R. Navigli, M. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai and M. Alam, eds, Lecture Notes in Computer Science, Vol. 10843, Springer, 2018, pp. 256–271. doi:10.1007/978-3-319-93417-4_17. https://doi.org/10.1007/978-3-319-93417-4_17.

[45] A. Graves, S. Fernández and J. Schmidhuber, Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition, in: *Artificial Neural Networks: Formal Models and Their Applications - ICANN 2005, 15th International Conference, Warsaw, Poland, September 11-15, 2005, Proceedings, Part II*, W. Duch, J. Kacprzyk, E. Oja and S. Zadrozny, eds, Lecture Notes in Computer Science, Vol. 3697, Springer, 2005, pp. 799–804. doi:10.1007/11550907_126.

[46] S. Guan, X. Jin, Y. Wang and X. Cheng, Link Prediction on N-ary Relational Data, in: *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, L. Liu, R.W. White, A. Mantrach, F. Silvestri, J.J. McAuley, R. Baeza-Yates and L. Zia, eds, ACM, 2019, pp. 583–593. doi:10.1145/3308558.3313414.

[47] T. Hamaguchi, H. Oiwa, M. Shimbo and Y. Matsumoto, Knowledge Transfer for Out-of-Knowledge-Base Entities : A Graph Neural Network Approach, in: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, C. Sierra, ed., ijcai.org, 2017, pp. 1802–1808. doi:10.24963/ijcai.2017/250.

[48] B. Haradizadeh and S. Singh, Tweeki: Linking Named Entities on Twitter to a Knowledge Graph, in: *Workshop on Noisy User-generated Text (W-NUT)*, 2020.

[49] S. Heindorf, M. Potthast, B. Stein and G. Engels, Vandalism Detection in Wikidata, in: *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*, S. Mukhopadhyay, C. Zhai, E. Bertino, F. Crestani, J. Mostafa, J. Tang, L. Si, X. Zhou, Y. Chang, Y. Li and P. Sondhi, eds, ACM, 2016, pp. 327–336. doi:10.1145/2983323.2983740.

[50] N. Heist, S. Hertling, D. Ringler and H. Paulheim, Knowledge Graphs on the Web - An Overview, in: *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*, I. Tiddi, F. Lécué and P. Hitzler, eds, Studies on the Semantic Web, Vol. 47, IOS Press, 2020, pp. 3–22. doi:10.3233/SSW200009.

[51] S. Hochreiter and J. Schmidhuber, Long Short-Term Memory, *Neural Comput.* **9**(8) (1997), 1735–1780. doi:10.1162/neco.1997.9.8.1735.

[52] J. Hoffart, Y. Altun and G. Weikum, Discovering emerging entities with ambiguous names, in: *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014*, C. Chung, A.Z. Broder, K. Shim and T. Suel, eds, ACM, 2014, pp. 385–396. doi:10.1145/2566486.2568003.

[53] J. Hoffart, M.A. Yosef, I. Bordino, H. Fürstenau, M. Pinkal, M. Spaniol, B. Taneva, S. Thater and G. Weikum, Robust Disambiguation of Named Entities in Text, in: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, 27-31 July 2011, John McIntyre Conference Centre, Edinburgh, UK, A meeting of SIGDAT, a Special Interest Group of the ACL*, ACL, 2011, pp. 782–792. https://www.aclweb.org/anthology/D11-1072/.

[54] B. Huang, H. Wang, T. Wang, Y. Liu and Y. Liu, Entity Linking for Short Text Using Structured Knowledge Graph via Multi-grained Text Matching, *Amazon Science* (2020).

[55] A. Ismayilov, D. Kontokostas, S. Auer, J. Lehmann and S. Hellmann, Wikidata through the eyes of DBpedia, *Semantic Web* **9**(4) (2018), 493–503. doi:10.3233/SW-170277.

[56] H. Ji, X. Pan, B. Zhang, J. Nothman, J. Mayfield, P. McNamee and C. Costello, Overview of TAC-KBP2017 13 Languages Entity Discovery and Linking, in: *Proceedings of the 2017 Text Analysis Conference, TAC 2017, Gaithersburg, Maryland, USA, November 13-14, 2017*, NIST, 2017. https://tac.nist.gov/publications/2017/additional.papers/TAC2017.KBP_Entity_Discovery_and_Linking_overview.proceedings.pdf.

[57] S.M. Kazemi, R. Goel, K. Jain, I. Kobyzev, A. Sethi, P. Forsyth and P. Poupart, Representation Learning for Dynamic Graphs: A Survey, *J. Mach. Learn. Res.* **21** (2020), 70:1–70:73. http://jmlr.org/papers/v21/19-447.html.

[58] Kensho R&D group, Kensho Derived Wikimedia Dataset. https://www.kaggle.com/kenshoresearch/kensho-derived-wikimedia-data.

[59] B. Kitchenham, Procedures for performing systematic reviews, *Keele, UK, Keele University* **33**(2004) (2004), 1–26.

[60] M. Klang and P. Nugues, Named entity disambiguation in a question answering system, in: *TheFifth Swedish Language Technology Conference (SLTC2014)*, 2014.

[61] M. Klang and P. Nugues, Hedwig: A Named Entity Linker, in: *Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*, N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk and S. Piperidis, eds, European Language Resources Association, 2020, pp. 4501–4508. https://www.aclweb.org/anthology/2020.lrec-1.554/.

[62] M. Klang, F. Dib and P. Nugues, Overview of the Ugglan Entity Discovery and Linking System, *CoRR* **abs/1903.05498** (2019). http://arxiv.org/abs/1903.05498.

[63] J. Klie, R.E. de Castilho and I. Gurevych, From Zero to Hero: Human-In-The-Loop Entity Linking in Low Resource Domains, in: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, D. Jurafsky, J. Chai, N. Schluter and J.R. Tetreault, eds, Association for Computational Linguistics, 2020, pp. 6982–6993. https://www.aclweb.org/anthology/2020.acl-main.624/.

[64] N. Kolitsas, O. Ganea and T. Hofmann, End-to-End Neural Entity Linking, in: *Proceedings of the 22nd Conference on Computational Natural Language Learning, CoNLL 2018, Brussels, Belgium, October 31 - November 1, 2018*, A. Korhonen and I. Titov, eds, Association for Computational Linguistics, 2018, pp. 519–529. doi:10.18653/v1/k18-1050.

[65] T. Kristanti and L. Romary, DeLFT and entity-fishing: Tools for CLEF HIPE 2020 Shared Task, in: *CLEF 2020-Conference and Labs of the Evaluation Forum*, Vol. 2696, CEUR, 2020a.

[66] T. Kristanti and L. Romary, DeLFT and entity-fishing: Tools for CLEF HIPE 2020 Shared Task, in: *CLEF 2020-Conference and Labs of the Evaluation Forum*, Vol. 2696, CEUR, 2020b.

[67] K. Labusch and C. Neudecker, Named Entity Disambiguation and Linking on Historic Newspaper OCR with BERT (2020).

[68] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P.N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer and C. Bizer, DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia, *Semantic Web* **6**(2) (2015), 167–195. doi:10.3233/SW-140134.

[69] X. Lin and L. Chen, Canonicalization of Open Knowledge Bases with Side Information from the Source Text, in: *35th IEEE International Conference on Data Engineering, ICDE 2019, Macao, China, April 8-11, 2019*, IEEE, 2019, pp. 950–961. doi:10.1109/ICDE.2019.00089.

[70] X. Lin, H. Li, H. Xin, Z. Li and L. Chen, KBPearl: A Knowledge Base Population System Supported by Joint Entity and Relation Linking, *Proc. VLDB Endow.* **13**(7) (2020), 1035–1049. http://www.vldb.org/pvldb/vol13/p1035-lin.pdf.

[71] Y. Lin, Z. Liu and M. Sun, Neural Relation Extraction with Multi-lingual Attention, in: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, R. Barzilay and M. Kan, eds, Association for Computational Linguistics, 2017, pp. 34–43. doi:10.18653/v1/P17-1004.

[72] X. Ling, S. Singh and D.S. Weld, Design Challenges for Entity Linking, *Trans. Assoc. Comput. Linguistics* **3** (2015), 315–328. https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/528.

[73] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer and V. Stoyanov, RoBERTa: A Robustly Optimized BERT Pretraining Approach, *CoRR* **abs/1907.11692** (2019). http://arxiv.org/abs/1907.11692.

[74] L. Logeswaran, M. Chang, K. Lee, K. Toutanova, J. Devlin and H. Lee, Zero-Shot Entity Linking by Reading Entity Descriptions, in: *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, A. Korhonen, D.R. Traum and L. Màrquez, eds, Association for Computational Linguistics, 2019, pp. 3449–3460. doi:10.18653/v1/p19-1335.

[75] M. Manske, Wikidata Stats, 2020-07-20. https://wikidata-todo.toolforge.org/stats.php.

[76] D.N. Milne and I.H. Witten, Learning to link with wikipedia, in: *Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, USA, October 26-30, 2008*, J.G. Shana-

han, S. Amer-Yahia, I. Manolescu, Y. Zhang, D.A. Evans, A. Kolcz, K. Choi and A. Chowdhury, eds, ACM, 2008, pp. 509–518. doi:10.1145/1458082.1458150.

[77] D. Moussallem, R. Usbeck, M. Röder and A.N. Ngomo, MAG: A Multilingual, Knowledge-base Agnostic and Deterministic Entity Linking Approach, in: *Proceedings of the Knowledge Capture Conference, K-CAP 2017, Austin, TX, USA, December 4-6, 2017*, Ó. Corcho, K. Janowicz, G. Rizzo, I. Tiddi and D. Garijo, eds, ACM, 2017, pp. 9:1–9:8. doi:10.1145/3148011.3148024.

[78] D. Moussallem, R. Usbeck, M. Röder and A.N. Ngomo, Entity Linking in 40 Languages Using MAG, in: *The Semantic Web: ESWC 2018 Satellite Events - ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*, A. Gangemi, A.L. Gentile, A.G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J.Z. Pan and M. Alam, eds, Lecture Notes in Computer Science, Vol. 11155, Springer, 2018, pp. 176–181. doi:10.1007/978-3-319-98192-5_33.

[79] I.O. Mulang, K. Singh, A. Vyas, S. Shekarpour, A. Sakor, M. Vidal, S. Auer and J. Lehmann, Context-aware Entity Linking with Attentive Neural Networks on Wikidata Knowledge Graph, *CoRR* **abs/1912.06214** (2019). http://arxiv.org/abs/1912.06214.

[80] I.O. Mulang, K. Singh, A. Vyas, S. Shekarpour, M. Vidal and S. Auer, Encoding Knowledge Graph Entity Aliases in Attentive Neural Network for Wikidata Entity Linking, in: *Web Information Systems Engineering - WISE 2020 - 21st International Conference, Amsterdam, The Netherlands, October 20-24, 2020, Proceedings, Part I*, Z. Huang, W. Beek, H. Wang, R. Zhou and Y. Zhang, eds, Lecture Notes in Computer Science, Vol. 12342, Springer, 2020a, pp. 328–342. doi:10.1007/978-3-030-62005-9_24. https://doi.org/10.1007/978-3-030-62005-9_24.

[81] I.O. Mulang, K. Singh, C. Prabhu, A. Nadgeri, J. Hoffart and J. Lehmann, Evaluating the Impact of Knowledge Graph Context on Entity Disambiguation Models, in: *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, M. d'Aquin, S. Dietze, C. Hauff, E. Curry and P. Cudré-Mauroux, eds, ACM, 2020b, pp. 2157–2160. doi:10.1145/3340531.3412159.

[82] K. Noullet, R. Mix and M. Färber, KORE 50^{DYWC}: An Evaluation Data Set for Entity Linking Based on DBpedia, YAGO, Wikidata, and Crunchbase, in: *Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*, N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk and S. Piperidis, eds, European Language Resources Association, 2020, pp. 2389–2395. https://www.aclweb.org/anthology/2020.lrec-1.291/.

[83] I.L. Oliveira, R. Fileto, R. Speck, L.P. Garcia, D. Moussallem and J. Lehmann, Towards holistic Entity Linking: Survey and directions, *Information Systems* (2020), 101624.

[84] J. Pennington, R. Socher and C.D. Manning, Glove: Global Vectors for Word Representation, in: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, A. Moschitti, B. Pang and W. Daelemans, eds, ACL, 2014, pp. 1532–1543. doi:10.3115/v1/d14-1162.

[85] D. Perkins, Separating the Signal from the Noise: Predicting the Correct Entities in Named-Entity Linking, Master's thesis, Uppsala University, 2020. http://uu.diva-portal.org/smash/record.jsf?pid=diva2:1437921&dswid=6063.

[86] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee and L. Zettlemoyer, Deep Contextualized Word Representations, in: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, M.A. Walker, H. Ji and A. Stent, eds, Association for Computational Linguistics, 2018, pp. 2227–2237. doi:10.18653/v1/n18-1202.

[87] F. Petroni, A. Piktus, A. Fan, P. Lewis, M. Yazdani, N.D. Cao, J. Thorne, Y. Jernite, V. Plachouras, T. Rocktäschel and S. Riedel, KILT: a Benchmark for Knowledge Intensive Language Tasks, *CoRR* **abs/2009.02252** (2020). https://arxiv.org/abs/2009.02252.

[88] V. Provatorova, S. Vakulenko, E. Kanoulas, K. Dercksen and J.M. van Hulst, Named Entity Recognition and Linking on Historical Newspapers: UvA.ILPS & REL at CLEF HIPE 2020 (2020).

[89] J. Raiman and O. Raiman, DeepType: Multilingual Entity Linking by Neural Type System Evolution, in: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, S.A. McIlraith and K.Q. Weinberger, eds, AAAI Press, 2018, pp. 5406–5413. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17148.

[90] L. Ratinov, D. Roth, D. Downey and M. Anderson, Local and Global Algorithms for Disambiguation to Wikipedia, in: *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA*, D. Lin, Y. Matsumoto and R. Mihalcea, eds, The Association for Computer Linguistics, 2011, pp. 1375–1384. https://www.aclweb.org/anthology/P11-1138/.

[91] D. Ringler and H. Paulheim, One Knowledge Graph to Rule Them All? Analyzing the Differences Between DBpedia, YAGO, Wikidata & co, in: *KI 2017: Advances in Artificial Intelligence - 40th Annual German Conference on AI, Dortmund, Germany, September 25-29, 2017, Proceedings*, G. Kern-Isberner, J. Fürnkranz and M. Thimm, eds, Lecture Notes in Computer Science, Vol. 10505, Springer, 2017, pp. 366–372. doi:10.1007/978-3-319-67190-1_33. https://doi.org/10.1007/978-3-319-67190-1_33.

[92] M. Röder, R. Usbeck and A.N. Ngomo, GERBIL - Benchmarking Named Entity Recognition and Linking consistently, *Semantic Web* **9**(5) (2018), 605–625. doi:10.3233/SW-170286.

[93] M. Röder, R. Usbeck, S. Hellmann, D. Gerber and A. Both, N^3 - A Collection of Datasets for Named Entity Recognition and Disambiguation in the NLP Interchange Format, in: *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014*, N. Calzolari, K. Choukri, T. De-

clerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk and S. Piperidis, eds, European Language Resources Association (ELRA), 2014, pp. 3529–3533. http://www.lrec-conf.org/proceedings/lrec2014/summaries/856.html.

[94] H. Rosales-Méndez, A. Hogan and B. Poblete, VoxEL: A Benchmark Dataset for Multilingual Entity Linking, in: *The Semantic Web - ISWC 2018 - 17th International Semantic Web Conference, Monterey, CA, USA, October 8-12, 2018, Proceedings, Part II*, D. Vrandecic, K. Bontcheva, M.C. Suárez-Figueroa, V. Presutti, I. Celino, M. Sabou, L. Kaffee and E. Simperl, eds, Lecture Notes in Computer Science, Vol. 11137, Springer, 2018, pp. 170–186. doi:10.1007/978-3-030-00668-6_11. https://doi.org/10.1007/978-3-030-00668-6_11.

[95] H. Rosales-Méndez, A. Hogan and B. Poblete, Fine-Grained Entity Linking, *Journal of Web Semantics* (2020), 100600.

[96] P. Rosso, D. Yang and P. Cudré-Mauroux, Beyond Triplets: Hyper-Relational Knowledge Graph Embedding for Link Prediction, in: *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, Y. Huang, I. King, T. Liu and M. van Steen, eds, ACM / IW3C2, 2020, pp. 1885–1896. doi:10.1145/3366423.3380257.

[97] A. Sakor, I.O. Mulang, K. Singh, S. Shekarpour, M. Vidal, J. Lehmann and S. Auer, Old is Gold: Linguistic Driven Approach for Entity and Relation Linking of Short Text, in: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran and T. Solorio, eds, Association for Computational Linguistics, 2019, pp. 2336–2346. doi:10.18653/v1/n19-1243.

[98] A. Sakor, K. Singh, A. Patel and M. Vidal, Falcon 2.0: An Entity and Relation Linking Tool over Wikidata, in: *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, M. d'Aquin, S. Dietze, C. Hauff, E. Curry and P. Cudré-Mauroux, eds, ACM, 2020, pp. 3141–3148. doi:10.1145/3340531.3412777.

[99] Ö. Sevgili, A. Panchenko and C. Biemann, Improving Neural Entity Disambiguation with Graph Embeddings, in: *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28 - August 2, 2019, Volume 2: Student Research Workshop*, F.E. Alva-Manchego, E. Choi and D. Khashabi, eds, Association for Computational Linguistics, 2019, pp. 315–322. doi:10.18653/v1/p19-2044.

[100] Ö. Sevgili, A. Shelmanov, M. Arkhipov, A. Panchenko and C. Biemann, Neural Entity Linking: A Survey of Models based on Deep Learning, *CoRR* **abs/2006.00575** (2020). https://arxiv.org/abs/2006.00575.

[101] F. Shanaz and R.G. Ragel, Wikidata based Location Entity Linking, in: *Proceedings of the 9th International Conference on Software and Computer Applications, ICSCA 2020, Langkawi, Malaysia, February 18-21, 2020*, ACM, 2020, pp. 307–312. doi:10.1145/3384544.3384592.

[102] W. Shen, J. Wang and J. Han, Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions, *IEEE Trans. Knowl. Data Eng.* **27**(2) (2015), 443–460. doi:10.1109/TKDE.2014.2327028.

[103] D. Sorokin and I. Gurevych, Context-Aware Representations for Knowledge Base Relation Extraction, in: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, M. Palmer, R. Hwa and S. Riedel, eds, Association for Computational Linguistics, 2017, pp. 1784–1789. doi:10.18653/v1/d17-1188.

[104] D. Sorokin and I. Gurevych, Mixing Context Granularities for Improved Entity Linking on Question Answering Data across Entity Categories, in: *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, *SEM@NAACL-HLT 2018, New Orleans, Louisiana, USA, June 5-6, 2018*, M. Nissim, J. Berant and A. Lenci, eds, Association for Computational Linguistics, 2018, pp. 65–75. doi:10.18653/v1/s18-2007.

[105] A. Sperduti and A. Starita, Supervised neural networks for the classification of structures, *IEEE Trans. Neural Networks* **8**(3) (1997), 714–735. doi:10.1109/72.572108.

[106] A. Spitz, J. Geiß and M. Gertz, So far away and yet so close: augmenting toponym disambiguation and similarity with text-based networks, in: *Proceedings of the Third International ACM SIGMOD Workshop on Managing and Mining Enriched Geo-Spatial Data, GeoRich@SIGMOD 2016, San Francisco, California, USA, June 26 - July 1, 2016*, A. Züfle, B. Adams and D. Wu, eds, ACM, 2016, pp. 2:1–2:6. doi:10.1145/2948649.2948651.

[107] Y. Su, H. Sun, B.M. Sadler, M. Srivatsa, I. Gur, Z. Yan and X. Yan, On Generating Characteristic-rich Question Sets for QA Evaluation, in: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, J. Su, X. Carreras and K. Duh, eds, The Association for Computational Linguistics, 2016, pp. 562–572. doi:10.18653/v1/d16-1054.

[108] T.P. Tanon, G. Weikum and F.M. Suchanek, YAGO 4: A Reason-able Knowledge Base, in: *The Semantic Web - 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31-June 4, 2020, Proceedings*, A. Harth, S. Kirrane, A.N. Ngomo, H. Paulheim, A. Rula, A.L. Gentile, P. Haase and M. Cochez, eds, Lecture Notes in Computer Science, Vol. 12123, Springer, 2020, pp. 583–596. doi:10.1007/978-3-030-49461-2_34. https://doi.org/10.1007/978-3-030-49461-2_34.

[109] K. Teru, E. Denis and W. Hamilton, Inductive Relation Prediction by Subgraph Reasoning, in: *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, Proceedings of Machine Learning Research, Vol. 119, PMLR, 2020, pp. 9448–9457. http://proceedings.mlr.press/v119/teru20a.html.

[110] A. Thawani, M. Hu, E. Hu, H. Zafar, N.T. Divvala, A. Singh, E. Qasemi, P.A. Szekely and J. Pujara, Entity Linking to Knowledge Graphs to Infer Column Types and Properties, in: *Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching co-located with the 18th International Semantic Web Conference, SemTab@ISWC 2019, Auckland, New Zealand, October 30, 2019*, E. Jiménez-Ruiz, O. Hassanzadeh, K. Srinivas, V. Efthymiou and J. Chen, eds, CEUR Workshop Proceedings, Vol. 2553, CEUR-WS.org, 2019, pp. 25–32. http://ceur-ws.org/Vol-2553/paper4.pdf.

[111] E.F. Tjong Kim Sang and F. De Meulder, Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition, in: *Proceedings of the Seventh Confer-*

ence on Natural Language Learning at HLT-NAACL 2003, 2003, pp. 142–147. https://www.aclweb.org/anthology/W03-0419.

[112] R. Usbeck, A.N. Ngomo, M. Röder, D. Gerber, S.A. Coelho, S. Auer and A. Both, AGDISTIS - Agnostic Disambiguation of Named Entities Using Linked Open Data, in: *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, T. Schaub, G. Friedrich and B. O'Sullivan, eds, Frontiers in Artificial Intelligence and Applications, Vol. 263, IOS Press, 2014, pp. 1113–1114. doi:10.3233/978-1-61499-419-0-1113.

[113] R. Usbeck, A.N. Ngomo, B. Haarmann, A. Krithara, M. Röder and G. Napolitano, 7th Open Challenge on Question Answering over Linked Data (QALD-7), in: *Semantic Web Challenges - 4th SemWebEval Challenge at ESWC 2017, Portoroz, Slovenia, May 28 - June 1, 2017, Revised Selected Papers*, M. Dragoni, M. Solanki and E. Blomqvist, eds, Communications in Computer and Information Science, Vol. 769, Springer, 2017, pp. 59–69. doi:10.1007/978-3-319-69146-6_-6. https://doi.org/10.1007/978-3-319-69146-6_6.

[114] T. van Veen, J. Lonij and W.J. Faber, Linking Named Entities in Dutch Historical Newspapers, in: *Metadata and Semantics Research - 10th International Conference, MTSR 2016, Göttingen, Germany, November 22-25, 2016, Proceedings*, E. Garoufallou, I.S. Coll, A. Stellato and J. Greenberg, eds, Communications in Computer and Information Science, Vol. 672, 2016, pp. 205–210. doi:10.1007/978-3-319-49157-8_18. https://doi.org/10.1007/978-3-319-49157-8_18.

[115] S. Vashishth, P. Jain and P.P. Talukdar, CESI: Canonicalizing Open Knowledge Bases using Embeddings and Side Information, in: *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018*, P. Champin, F.L. Gandon, M. Lalmas and P.G. Ipeirotis, eds, ACM, 2018, pp. 1317–1327. doi:10.1145/3178876.3186030.

[116] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser and I. Polosukhin, Attention is All you Need, in: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, I. Guyon, U. von Luxburg, S. Bengio, H.M. Wallach, R. Fergus, S.V.N. Vishwanathan and R. Garnett, eds, 2017, pp. 5998–6008. http://papers.nips.cc/paper/7181-attention-is-all-you-need.

[117] O. Čerba Ing and M. Otakar Čerba, 7th International Conference on Cartography and GIS, Technical Report, 29, 2018. ISSN 1848-0713. http://id.loc.gov/authorities.

[118] O. Vinyals, M. Fortunato and N. Jaitly, Pointer Networks, in: *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama and R. Garnett, eds, 2015, pp. 2692–2700. http://papers.nips.cc/paper/5866-pointer-networks.

[119] D. Vrandecic, Architecture for a multilingual Wikipedia, *CoRR* **abs/2004.04733** (2020). https://arxiv.org/abs/2004.04733.

[120] D. Vrandečić and M. Krötzsch, Wikidata: a free collaborative knowledgebase, *Commun. ACM* **57**(10) (2014), 78–85. doi:10.1145/2629489.

[121] W3C Microposts Community Group, Making Sense of Microposts (#Microposts2016), 2016. http://microposts2016.seas.upenn.edu/challenge.html.

[122] P. Wang, J. Han, C. Li and R. Pan, Logic Attention Based Neighborhood Aggregation for Inductive Knowledge Graph Embedding, in: *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, AAAI Press, 2019a, pp. 7152–7159. doi:10.1609/aaai.v33i01.33017152.

[123] X. Wang, T. Gao, Z. Zhu, Z. Liu, J. Li and J. Tang, KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation, *CoRR* **abs/1911.06136** (2019b). http://arxiv.org/abs/1911.06136.

[124] G. Weikum, L. Dong, S. Razniewski and F.M. Suchanek, Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases, *CoRR* **abs/2009.11564** (2020). https://arxiv.org/abs/2009.11564.

[125] Wikimedia Foundation, Index of /wikidatawiki/entities/, 2020-08-21. https://dumps.wikimedia.org/wikidatawiki/entities/.

[126] Wikimedia Foundation, Wikistats, 2020-10-09. https://stats.wikimedia.org/#/metrics/wikidata.org.

[127] L. Wu, F. Petroni, M. Josifoski, S. Riedel and L. Zettlemoyer, Zero-shot Entity Linking with Dense Entity Retrieval, *CoRR* **abs/1911.03814** (2019a). http://arxiv.org/abs/1911.03814.

[128] T. Wu, A. Khan, H. Gao and C. Li, Efficiently Embedding Dynamic Knowledge Graphs, *CoRR* **abs/1910.06708** (2019b). http://arxiv.org/abs/1910.06708.

[129] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang and S.Y. Philip, A comprehensive survey on graph neural networks, *IEEE Transactions on Neural Networks and Learning Systems* (2020).

[130] X. Yang, X. Gu, S. Lin, S. Tang, Y. Zhuang, F. Wu, Z. Chen, G. Hu and X. Ren, Learning Dynamic Context Augmentation for Global Entity Linking, in: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, K. Inui, J. Jiang, V. Ng and X. Wan, eds, Association for Computational Linguistics, 2019, pp. 271–281. doi:10.18653/v1/D19-1026.

[131] Y. Yang and M. Chang, S-MART: Novel Tree-based Structured Learning Algorithms Applied to Tweet Entity Linking, *CoRR* **abs/1609.08075** (2016). http://arxiv.org/abs/1609.08075.

[132] Z. Yang, Z. Dai, Y. Yang, J.G. Carbonell, R. Salakhutdinov and Q.V. Le, XLNet: Generalized Autoregressive Pretraining for Language Understanding, in: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, H.M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E.B. Fox and R. Garnett, eds, 2019, pp. 5754–5764. http://papers.nips.cc/paper/8812-xlnet-generalized-autoregressive-pretraining-for-language-understanding.

[133] W. Yih, M. Richardson, C. Meek, M. Chang and J. Suh, The Value of Semantic Parse Labeling for Knowledge Base Question Answering, in: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL*

2016, August 7-12, 2016, Berlin, Germany, Volume 2: Short Papers, The Association for Computer Linguistics, 2016. doi:10.18653/v1/p16-2033.

[134] F. Yu and V. Koltun, Multi-Scale Context Aggregation by Dilated Convolutions, in: *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, Y. Bengio and Y. LeCun, eds, 2016. http://arxiv.org/abs/1511.07122.

[135] S. Zhou, S. Rijhwani, J. Wieting, J.G. Carbonell and G. Neubig, Improving Candidate Generation for Low-resource Cross-lingual Entity Linking, *Trans. Assoc. Comput. Linguistics* **8** (2020), 109–124. https://transacl.org/ojs/index.php/tacl/article/view/1906.

[136] S. Zwicklbauer, C. Seifert and M. Granitzer, DoSeR - A Knowledge-Base-Agnostic Framework for Entity Disambiguation Using Semantic Embeddings, in: *The Semantic Web. Latest Advances and New Domains - 13th International Conference, ESWC 2016, Heraklion, Crete, Greece, May 29 - June 2, 2016, Proceedings*, H. Sack, E. Blomqvist, M. d'Aquin, C. Ghidini, S.P. Ponzetto and C. Lange, eds, Lecture Notes in Computer Science, Vol. 9678, Springer, 2016, pp. 182–198. doi:10.1007/978-3-319-34129-3_12.

[137] entity-fishing, GitHub, 2016–2020.