Towards Holistic Entity Linking: Survey and Directions

Italo L. Oliveira[a,*], Renato Fileto[a], René Speck[c,e], Luís P. F. Garcia[d], Diego Moussallem[b,e], Jens Lehmann[f,e,g]

[a]*Department of Informatics and Statistics, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil*
[b]*Data Science Group, University of Paderborn, North Rhine-Westphalia, Germany*
[c]*Scalable Data Management, Computing Centre Leipzig University, Saxony, Germany*
[d]*Computer Science Department, University of Brasília, Brasília, Distrito Federal, Brazil*
[e]*Institute for Applied Informatics, Leipzig, Saxony, Germany*
[f]*Smart Data Analytics, University of Bonn, Bonn, NorthRhine-Westphalia, Germany*
[g]*Fraunhofer IAIS, Dresden, Saxony, Germany*

Abstract

Entity Linking (EL) empowers Natural Language Processing applications by linking relevant mentions found in raw textual data to precise information about what they supposedly stand for. However, EL approaches have mostly focused on particular kinds of inputs and frequently fail to properly handle texts from specific sources (e.g., microblogs) that have particularities such as grammatical errors, slangs, lack of contextual information and other problems, besides difficulties to exploit their associated data (e.g., time stamps, geographic indicators, authors' profile data). Some EL approaches have been devised to circumvent such challenges. They exploit several inputs, data features, and EL methods in a synergetic process for more powerful and robust collective EL. This paper reviews recent works that employ such holistic strategies for EL, discusses their limitations, and proposes directions for further advancing holistic EL approaches.

Keywords: Entity Linking (EL), Holistic EL Approaches, Collective EL, Embedding-based EL

---

*Corresponding author

Email addresses: `italo.oliveira@posgrad.ufsc.br` (Italo L. Oliveira), `r.fileto@ufsc.br` (Renato Fileto), `rene.speck@uni-leipzig.de` (René Speck), `luis.garcia@unb.br` (Luís P. F. Garcia), `diego.moussallem@uni-paderborn.de` (Diego Moussallem), `jens.lehmann@cs.uni-bonn.de` (Jens Lehmann)

1. Introduction

An astounding amount of textual data (e.g., news, comments, social media posts) is produced and made available on the Web each day[1]. However, their lack of well-defined machine-processable semantics hinders their use by applications [1]. A way to circumvent this problem is to semantically annotate these data before their use [2, 3].

Manual semantic annotation of vast amounts of data is expensive due to the prohibitive workforce needed and the difficulty of obtaining quality and standardized results [2]. Thus, many automatic tasks for semantic annotation have been proposed in different research fields such as Text Mining (TM), Natural Language Processing (NLP), and the Semantic Web. One of the most important tasks for annotating textual content is Entity Linking (EL) [4].

The EL task links each relevant named entity mention (e.g., *'Jordan'*) found in a text to a descriptor of what that mention refers to (e.g., the famous basketball player or the country called *'Jordan'*) in the context where it appears. The entity descriptors can be taken, for example, from a Knowledge Graph (KG) (e.g., DBpedia[2] [5, 6], Yago[3] [7], Freebase[4] [8], LinkedGeoData[5] [9], WikiData[6] [10]). Commonly, the mentions are recognized one step before by the Named Entity Recognition (NER) task, which is responsible for identifying and tagging these mentions with their respective types (e.g., PERSON, ORGANIZATION, LOCATION). Some works consider that EL is just the disambiguation of the mentions (also called Named Entity Disambiguation (NED) task) [11, 12], while others consider that EL is a combination of the NER and NED tasks [13, 14, 15]. In this work, we consider the first definition for EL.

We use the following formal definition of the EL task, extracted from Wei et al. (2015) [11]. Given a set of entities E (e.g., within a knowledge graph KG) and a set of named entity mentions M within a text document T, the EL task aims to map each mention $m \in M$ to its corresponding entity $e \in E$.

[1]https://www.visualcapitalist.com/what-happens-in-an-internet-minute-in-2019/
[2]https://wiki.dbpedia.org
[3]http://www.yago-knowledge.org/
[4]https://developers.google.com/freebase/
[5]http://aksw.org/Projects/LinkedGeoData.html
[6]https://www.wikidata.org

In case that the corresponding entity e for a mention m does not exist in E (i.e. $e \notin E$), m is labeled as "NIL", whose meaning is unlinkable, i.e., it is not provided in E.

Recent surveys about EL present a good overview of existing approaches, datasets, benchmarks [11, 16], and EL approaches for a specific type of text document, such as microblog posts [17]. However, these surveys do not consider in detail emerging strategies that influence new EL approaches and can be regarded as ways to consider several facets of the EL task concomitantly in more holistic EL processes. Thus, this survey presents a more detailed literature review and analysis of emerging holistic approaches for the EL task. Such a holistic view may provide extra information for EL approaches to tackle ambiguous named entity mentions, for example, in texts with limited context and lots of noise (e.g., typos, grammatical errors, extensive use of slangs, acronyms) like social media posts.

According to Cambridge Dictionary[7], holism is *"the belief that each thing is a whole that is more important than the parts that make it up"*. From this concept, we understand that holism in EL involves several kinds of input (e.g., text documents to be annotated, their associated data and metadata, KGs), a variety of relevant features that can be extracted from these inputs, and a myriad of methods that can be employed in the data processing for EL. Although several works have been employing some degree of holism in the EL task, as detailed in Section 2, to the best of our knowledge, they were never analyzed and summarized from this viewpoint.

This survey aims to provide a comprehensive review of a variety of EL approaches that exhibit some holism. We classify these approaches according to key aspects that allow an overview of holistic techniques applied to EL and a better understanding of their diversity. These key holistic aspects include the exploitation of distinct inputs and data features, the use of diverse NLP tasks for information extraction and, the collective disambiguation of mentions on text and knowledge models, such as embeddings. To the best of our knowledge, these aspects of holistic EL approaches have not been described in the literature yet. They are usually implicit in the EL proposals. They can be useful to understand better, classify, and compare these approaches. This survey gives insights into how a variety of techniques can be combined into more holistic approaches to improve EL results.

[7]https://dictionary.cambridge.org/dictionary/english/holism

Most of the EL approaches from the literature still rely directly on Wikipedia to determine entities, and many of them employ well-known NLP tasks. On the other hand, there is a recent trend to use embeddings. Meanwhile, there are also efforts to boost EL power and reliability by considering semantic coherence of entities in collective EL processes. Based on these findings, we propose some pillars for future holistic EL approaches, that include: (i) handling EL (and semantic annotation) as a general process which can be tailored for different kinds of data (e.g., news, social media posts) and goals, by appropriately selecting, composing and tuning suitable approaches for each one of its constituent tasks; (ii) better-exploiting word embeddings aligned with knowledge embeddings for EL; (iii) using context information extracted from texts and their associated data and metadata (e.g., source, location, time) to disambiguate collections of related mentions holistically (e.g., in the same document or the documents of the same author, geographically or historically related) based on measures of the semantic coherence of the entity candidates. We also outline a holistic EL approach that exploits these pillars to disambiguate more entity mentions more accurately.

1.1. Bibliographical Review Procedure

The steps of the methodology employed in the systematic bibliographical review are presented in Figure 1. The search for papers was done in Google Scholar[8], the ACM Digital Library[9], Springer Link[10] and, Scopus[11]. We chose these platforms because they gave the best results for preliminary searches, including books, conference, proceedings and journals papers. The search string used was *"(Entity Linking"* OR *"Named Entity Disambiguation")* AND *"text document"*.

Although the focus of this survey is holistic approaches for EL, most of the articles do not use the word *"holistic"* or *"holistic"* in their title or contents, justifying its exclusion from the keyword search. Moreover, we did not consider articles tagged as *preview-only* in Springer Link.

In total, we found 23135 articles. To reduce the number of articles (Step 1), we removed duplicates using the Mendeley[12] tool. Then, we considered only

---

[8] https://scholar.google.com.br/
[9] https://dl.acm.org/
[10] https://link.springer.com/
[11] https://www.scopus.com/
[12] https://www.mendeley.com/

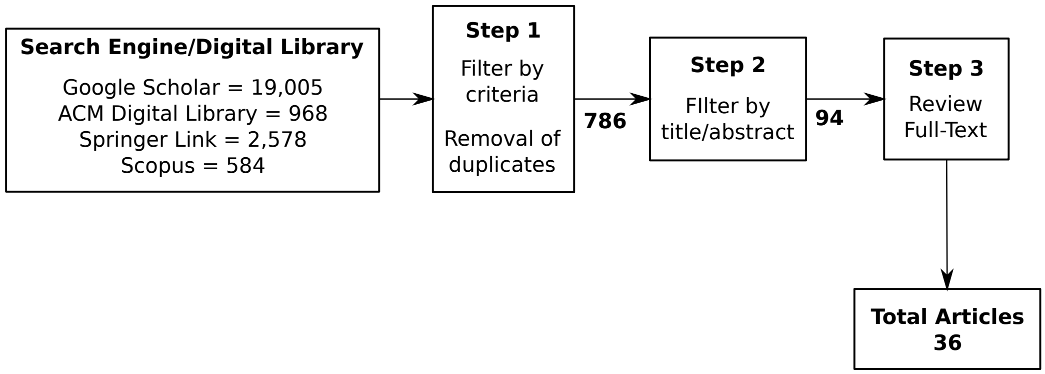

Figure 1: Steps followed to retrieve the articles analyzed in this survey.

articles that satisfied the following criteria: (i) peer-reviewed or published; (ii) published from 2005 onwards; (iii) written in English; (iv) EL approaches that annotate textual data. Regarding criteria (ii), we have found only a few articles about EL before 2005 and, most of them, adopt a manual or semi-automatic method. In this survey, we focus only on fully automatic EL approaches. Step 1 yielded 786 articles.

In Step 2, we manually analyzed the title and abstract of the 786 articles resulting from Step 1 and removed those not related to the EL task, resulting in 84 articles. Lastly, in Step 3, we reviewed the remaining 94 articles and selected the 36 ones having any of the holistic aspects presented in Section 2.

1.2. Outline

The remaining of this paper is structured as follows. Section 2 illustrates how holism can help to generate better EL results and defines key aspects of holistic approaches for EL. Then section 3 reviews the approaches selected from recent literature that present at least one of these holistic aspects. Section 4 provides a comparative analysis of these approaches. Section 5 depicts the potential pillars that can support future holistic approaches. Finally, Section 6 concludes the paper and presents future directions for research.

2. Holistic EL

The goal of this section is to draw how holism can manifest and contribute to better results in EL approaches. First, Section 2.1 exemplifies the potential benefits of holism in EL, in a real-world scenario where entity mentions must be disambiguated in microblog posts. Then, section 2.2, delineates key aspects of holistic EL approaches that we have derived from our studies and that establish essential criteria to identify, classify and analyze holistic EL approaches.

2.1. Motivating Example

Figure 2 illustrates how a holistic view can be helpful for EL on real microblog posts by considering several data features, methods, and semantic coherence to disambiguate entity mentions. Assume that different methods and tools annotated the two tweets presented on the top of the figure, using the information resources represented as labeled boxes. The green dashed lines represent the links from each mention to the respective resource used to describe it correctly, while red dotted lines represent links to resource descriptions that refer to incorrect disambiguations of mentions. The blue and yellow boxes refer to, respectively, concepts and entities from a KG. The gray boxes are word senses described in a lexical base. The resource description corresponding to the mention *'uncountable'* is not shown in the figure just for simplicity.

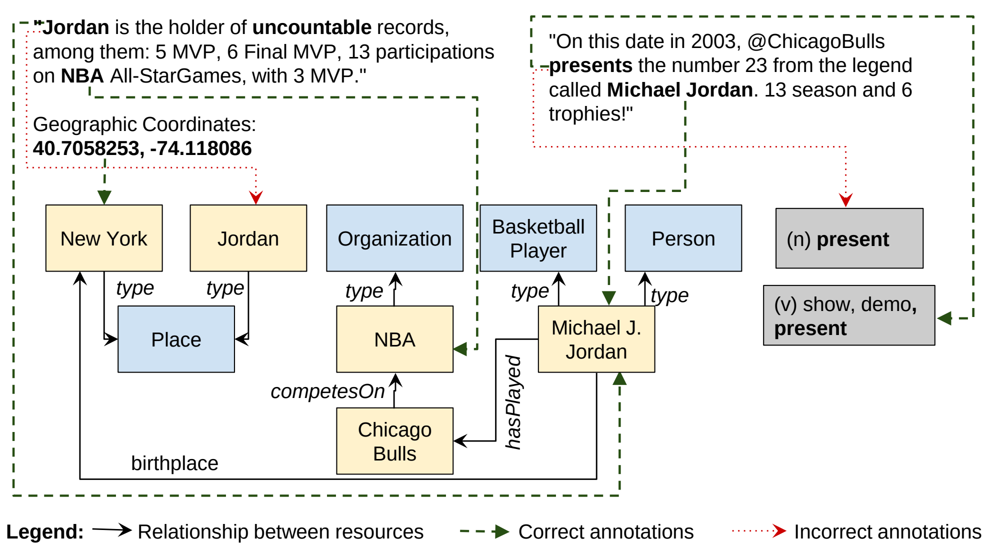

Figure 2: Collaborative disambiguation of entity mentions based on coherence. (Color required)

Different data features are considered by different NLP tasks and annotation approaches for the tweet on the left. Geographic coordinates and the indication of a place in the tweet metadata can also be used to determine the PoI from where the tweet was sent, which is, in this case, the city of *'New York'*. This, allied with the annotation of the mention *'NBA'* (National Basketball Association), provides evidence to disambiguate the ambiguous mention *'Jordan'* to its correct description, i.e., the basketball player *'Michael J. Jordan'*, instead of the country *Jordan*.

Then, if the tweet on the left is related to the one on the right (e.g., both come from the same user around the same time), we can use the annotations generated for the former to help the EL task on the latter (and vice-versa). Therefore, the previous annotations are used as what we can call a semantic context to disambiguate the mention *'Michael Jordan'* successfully to its correct description, based on semantic coherence. However, though the tweet on the right has the mention *'@ChicagoBulls'* (a basketball team where *'Michael J. Jordan'* played), which could also help to assemble the semantic context, most tools that we have tried could not identify this mention. It occurs due to the use of the special character '@' and the lack of spaces to separate the words constituting the mention *'@ChicagoBulls'*. Noise data like this, misspellings, and grammar errors can hinder the performance of current EL approaches. More holistic strategies can help face such challenges by bringing more contextual information that can be useful in the EL process.

2.2. Key Aspects of Holistic EL Approaches

The previous example evidences some key aspects of holistic EL approaches that we have identified in our studies. They include the exploitation of distinct inputs and data features, the use of diverse NLP tasks. However, these aspects of holistic EL approaches have not been adequately described in the literature yet. They are usually implicit in the EL proposals. We believe that they can be useful to understand better, classify, and compare these approaches, besides giving insights about promising research directions. Hence, in this survey, we define these key aspects of holistic EL approaches as follows.

Distinct inputs and data features: Different EL approaches can have distinct inputs. Some approaches are tailored to annotate some specific kinds of text documents (e.g., microblog posts) and may also exploit data and metadata associated with them (e.g., the location of the user posting a tweet, the tweet timestamp). Besides, alternative external inputs can be used to provide entity descriptions and further information/knowledge to help EL. Most current EL approaches employ particular features extracted from a limited number of inputs. Nevertheless, synergetic exploitation of distinct data and data features can boost EL results, especially if context information is limited, as in some microblog posts.

Diverse NLP tasks: The EL task is usually preceded by the NER task, which can rely on other NLP tasks and tools to identify mentions in the text. Moreover, several other NLP tasks (e.g., Word Sense Disambiguation, Entity Saliency) can be used before, after, or concomitantly with the EL task to improve its results.

Dismabiguation methods: The disambiguation step of the EL task may employ a myriad of methods, that can be combined or not. Two promising directions are collective disambiguation and embedding-based methods. Collective disambiguation considers several named entity mentions simultaneously and the coherence of entity candidates to drive the disambiguation process. Embedding-based methods use as inputs word, entity, or KG embeddings. Embeddings-based methods can exploit global relations in a more efficient way than traditional approaches like graph-based ones.

3. Current EL Approaches exhibiting Holism

This section reviews works that present holistic approaches to tackle the EL task. These works were collected in an extensive and systematic review, as explained in Section 1.1. We considered works that already exploit, explicitly or implicitly, any form of holism in their EL processes. Each subsection refers to one of the aspects of holistic EL described in Section 2. The works are organized among these subsections according to their premises, purposes, and proposed approaches. Some works comprise more than one aspect of holism. In such cases, the work is described in the section referring to the holistic aspect that we consider most exploited or relevant in that work.

3.1. Distinct Inputs and Data Features

EL can rely on a variety of features extracted from the text documents to be annotated, as well as their associated data and metadata. For example, the temporal context of microblog posts (e.g., tweets) [18, 19], their ordering, and their associated data. According to Hua et al. (2015) [18], EL approaches that focus just on the context around named entity mentions are unsuitable for queries and tweets. This occurs because both queries and tweets are concise, and, therefore, present little context around the named entity mentions. Therefore, the authors propose the use of other features besides that. More specifically, they propose the use of entity popularity and recency to determine user interests by his/her social interactions. Hua et al. (2015) [18] define entity recency as the recent popularity of a specific entity. Their user interests model captures the most relevant entities for the user. It is produced through the analysis of the user social interactions. The scores of the entity candidates during the EL task are obtained by summing the scores of entity popularity, entity recency, and user interests. The correct entity candidate for its respective mention is the one with the highest score.

Tran et al. (2015) [19] aim to annotate hashtags found in tweets, instead of named entity mentions. To disambiguate hashtags, they propose the use of a ranking learning algorithm using temporal information. According to the authors, the meaning of a hashtag can change depending on the posting time. For example, the hashtag *#sochi* usually refers to the city of Sochi, in Russia. However, around 2014, that hashtag was extensively used to talk about the Winter Olympic Games that happened in that city in 2014. To perform such disambiguation, the authors use Wikipedia pages, their edit history, and their page view statistics. Based on them, the proposed method builds a graph, called *influence graph*, having the entity candidates for the hashtags as nodes and their hyperlinks as edges. The edges and several similarity measures determine the influence of an entity candidate on the graph. This graph is used to train a ranking learning algorithm, which is similar to the PageRank algorithm. Then, the entity candidates with the biggest influence are chosen as the correct disambiguation for their respective hashtags.

3.2. Diverse NLP Tasks

Since the EL task requires the mentions to be previously spotted and demarcated in the text by the NER task, which can be done manually or automatically, most works consider these tasks independently, with NER completely preceding EL. However, some papers disagree with such an independent approach [20, 21, 22, 23], arguing that "mutual dependency between the two tasks is ignored" [20] and that "errors caused by NER will propagate to EL without the possibility of recovery" [21]. Consequently, the EL task does not take advantage of the features extracted for the NER task, except the named entity mentions, which are recognized and sometimes also classified. Based on this, Luo et al. 2015 [20] propose the Joint Entity Recognition and Linking (JERL) model. According to the authors, NER is usually defined as finding a sequence of named entity mentions and labeling them (with their classes), while the EL task is defined as a ranking task. Thus, they consider that the biggest challenge to model the NER and EL tasks jointly is to combine the sequence labeling and ranking tasks. To achieve this, the JERL model extends a semi Conditional Random Field (CRF) model for modeling the entities distribution and "mutual dependency over-segmentation". To infer which entity candidate is the correct one to describe a named entity mention, Luo et al. (2015) [20] extend the Viterbi algorithm [24].

Meanwhile, Wang & Iwaihara (2019) [22] and Martins et al. (2019) [23] propose a joint neural network model to tackle both tasks simultaneously. In Wang & Iwaihara (2018) [22], the proposed model is a deep neural network based on Tree recursive Neural Networks (TNNs) and Convolutional Neural Networks (CNNs). For the NER task, the authors use the system BRNN-CNN [25], a deep neural network model for NER that also uses TNNs. For the EL task, the authors propose a model that computes the semantic similarity between the recognized mentions and their respective entity candidates. This is achieved by comparing the representation of: (i) context (text around the named entity or introductory text of an entity), performed by TNNs with Long Short-Term Memory (LSTM) neural network; (ii) the whole document, performed by CNNs with an attention mechanism and; (iii) type of the named entity and the entity candidate. Lastly, Wang & Iwaihara (2019) [22] combine the loss functions of the NER and the EL models to perform their joint training.

In Martins et al. (2019) [23], the authors extend the Stack-LSTM model [26], which has been used for the NER task [27], to tackle both tasks. They augment the Stack-LSTM with two bi-LSTM, whose inputs are word embeddings, to improve the NER task. The mentions, together with the entity embedding of the entity candidates of their respective mentions, are inputs for an affine transformation layer. The output of such a layer is the score for each entity candidate. The candidate with the highest score is selected as the correct one. As the approach of Kolitsas et al. (2018) [21] presents a jointly model for word and entity embeddings, it will be described in Section 3.3.2.

Word Sense Disambiguation (WSD) is an NLP task similar to EL. Both tasks aim to annotate parts of a text with semantically well-described resources. What distinguishes these tasks are their slightly distinct purposes (namely, disambiguate named entity mentions for EL and disambiguate word senses for WSD) and the different types of resources used for the semantic annotation (KGs for EL and lexicons like Wordnet for WSD). Most of the existing works in the literature handle these tasks separately. According to Moro et al. (2014) [28], it leads to duplicated efforts, such as applying NLP tasks to extract features that are useful for the EL and WSD tasks. Besides, better results may be obtained by combining the contextual information manipulated by these tasks. Thus, they propose an approach that combines the EL and WSD tasks, aiming to improve the results generated by both tasks. This approach has three steps: (i) building the semantic signatures for vertices in a KG called BabelNet [29] that combines a variety of linked data with a multilingual version of the Wordnet; (ii) extracting all the relevant fragments from a given text (e.g., entity mentions, simple words); and (iii) disambiguating the entity and word sense candidates through the use of the semantic signatures. The semantic signature of a given vertex of Babelnet includes all the KG vertices that are densely connected to it and also among themselves [28]. According to the authors, in a KG, many concepts related to an entity or word sense are not directly connected to it. To avoid this issue, the edges of a KG are weighted using the concept of directed triangles.

The Random Walk with Restart algorithm [30] uses the weights from Moro et al. (2014) [28] to build the semantic signatures. The entities and word sense candidates are identified through the use of superstring matching. Moreover, text fragments can overlap. For example, consider the tweet *"After Game of Thrones and Star Wars episode 8, Dubrovnik as a host to Jame Foxx, Jame Dornan and Leo Di Capri in Robin Hood"*. Their approach recognizes the text fragment *Game of Thrones* of this tweet as an entity. On the other hand, the words *Game* and *Thrones* are separetelly recognized as lexical resources. For the disambiguation of the entities and word sense candidates, the candidates are connected when the semantic signature of one has the other. After the connection of semantic signatures, a novel densest subgraph heuristic is applied to the resulting network of semantic signatures. The scores for the semantic signatures are compared, and the entities and lexical resources disambiguated.

Although less popular, there are several other (combinations of) semantic annotation tasks in the literature, for example, Entity Discovery (ED) [31] and Entity Saliency (ES) [32, 12, 33]. The ED task tries to identify and group together entity mentions that refer to a same entity that does not exist in a KG [31]. Wick et al. (2013) [31] propose a joint model to tackle both EL and ED. For this, the authors propose the use of hierarchical trees built from entity mentions and their entity candidates. The leaves of the tree are the entity candidates. The internal nodes are summaries of the entity candidate features. The root is the named entity mention. A temperature-regulated Markov Chain Monte Carlo (MCMC) [31] was used for disambiguation and discovery of new entities.

The ES task determines the relevance of the entity mentions according to with their importance to interpret the contents of a given text document. Usually, ES is a task performed after the EL task. However, some papers argue that the combination of EL and ES can improve both tasks [32, 12, 33]. Since Chen et al. (2018) [33] present a bilinear model to joint learn word and entity embeddings, their approach is described in Section 3.3.2. On the other hand, in Trani et al. (2016) [32] and Trani et al. (2018) [12], the authors propose a Salient Entity Linking algorithm to perform the EL and ES tasks. Their proposal has two steps: Candidate Pruning and Saliency Linking. The Candidate Pruning step aims to reduce the number of entity candidates for each mention in a text document. For this, the authors use a supervised technique that classifies the entity candidates as relevant or irrelevant. The Saliency Linking step also uses a supervised technique, that predicts the entity candidates as *top relevant*, *highly relevant*, *partially relevant* and *not relevant*. One entity candidate will be considered as incorrect if its classification is *not relevant*. Therefore, the proposal [32, 12] addresses both the EL and the ES tasks simultaneously.

3.3. Disambiguation Methods

This section reviews EL proposals presenting approaches for collective disambiguation and some embedding-based approaches. We highlight that these two types of approaches can be combined to improve the EL task results further, as presented by works discussed in the following sections.

3.3.1. Collective Disambiguation Methods

Several works propose methods and algorithms for EL that disambiguate the entity mentions in a document separately, i.e., considering only the textual context around each mention. However, some proposals consider that the named entity mentions in the same document or related documents are semantically related [34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48].

One way to consider entity mentions collectively in a document is through the exploitation of links between entities available in a KG (e.g., DBpedia, a

KG specially built for EL on specific data). In Han et al. (2011) [34] and Guo & Barbosa (2014) [36], the entity mentions, entity candidates, and hyperlinks are modeled as a graph (called Referent Graph) in the former, while the same elements are modeled as a graphical model in the latter. In Han et al. (2011) [34] and Ganea et al. (2016) [39], evidence collected from the document to be annotated and from textual contents associated with the entity candidates are propagated to their respective representations. In Guo & Barbosa (2014) [36], otherwise, the authors exploit the links between entities in a KB to build graphs, called semantic signatures, of the entity candidates and the documents to be annotated. To disambiguate the candidates, after the propagation of the evidence, the candidates that achieve the best score [34] or have the highest probability [39] are chosen as the correct entity representation for their respective entity mentions. Meanwhile, in Guo & Barbosa (2014) [36], the candidates whose semantic signature presents the highest similarity with the semantic signature of the document to be annotated are chosen.

Similarly to [34, 36, 39], Rama-Maneiro et al. (2020) [48] build a graph representing the links between entity candidates to disambiguate the named entity mentions collectively. However, Rama-Maneiro et al. (2020) [48] exploit facts present in DBpedia to model the graph, exploiting some of its existing relations. The authors use the graph to calculate the degree of centrality to identify the most important node in the graph, i.e., the most relevant entity candidate. The authors avoid the combinatorial explosion when building the entity candidates graph by employing several strategies, such as using only the most relevant DBpedia relations, indexing paths between DBpedia nodes and, *"only considering entity candidates that are related to the topic of the document"*. The last strategy is achieved by an inference system that compares the context of an entity candidate (previously built from Wikipedia) with the text document. This way, the authors employ both topic coherence and node centrality in the disambiguation step. Lastly, the authors argue that their approach is capable of building entity candidate graphs up to 8 relations of the distance between entity candidates, while other approaches only achieve 2 or 3 relations of distance.

Although Wikipedia hyperlinks are useful to build graphs used in the disambiguation step of collective approaches, Vaigh et al. (2019) [45] argue that the lack of semantics of these hyperlinks hinders existing approaches. The authors propose the use of semantic relations present in KGs to improve collective EL approaches. Their approach employs a binary logistic regression classifier whose inputs are similarities between the word embedding representations of each mention with each one of its entity candidates, as local score, and, the semantic relatedness between entity candidates, as a global score. The semantic relatedness is calculated by considering the number of relations between entity candidates e_i and e_j divided by the total number of facts in which e_i takes part. To avoid a combinatorial explosion, the authors aggregate the semantic relatedness between entity candidates with a number k of global features.

In the papers [34, 36, 39], the pairwise scores between entity candidates, including incorrect entities, are considered. According to Fang et al. (2019) [44] and Yang et al. (2019) [47], this strategy increases the complexity of the approaches and introduces noise in the results. Thus, both papers propose a sequential collective EL approach, in which they disambiguate the named entity mentions considered less ambiguous and use them to help the disambiguation of the following and more ambiguous named entity mentions. The approach proposed in Fang et al. (2019) [44] is based on an LSTM deep neural network with reinforcement learning with word and entity embedding as inputs. Meanwhile, the approach proposed in Yang et al. (2019) [47] is based on an FFNN with reinforcement learning and attention mechanisms. The inputs are word embeddings representing words and named entity mentions and, entity embeddings representing the already disambiguated named entity mentions and entity candidates.

Similarly to the proposals of [44, 47], Phan et al. (2019) [46] also do not consider the pairwise scores between all pairs of entity candidates. However, instead of considering a sequential collective EL approach, Phan et al. (2019) [46] propose a collective approach by disambiguating the named entity mentions in pairs, considering the pair with the highest confidence in each step. This way, their approach produces a minimum spanning tree of entity candidates that correctly disambiguate the named entity mentions. Their approach is based in a graph whose vertices are the entity candidates, and edges connect entity candidates from distinctly named entity mentions. Edge weights are used as a semantic distance score. The semantic distance score is based on local confidence and pairwise coherence score. The local confidence is calculated by Gradient Boost Tree, which uses the popularity of an entity candidate and the semantic similarity between the word embedding that surrounds the named entity mention and the entity embedding that represents the entity candidate. The pairwise coherence score can be represented by Wikipedia Linked-Based measures [49], the logarithm of the Normalized Jaccard Similarity, and the cosine similarity between entity embeddings. To disambiguate the named entity mentions, the authors employ a heuristic to find the minimum spanning tree. This heuristic is similar to the Kruskal algorithm [50], with the difference that, every time a pair of vertices is selected, the remaining vertices represent entity candidates for the same named entity mentions are removed.

Both Huang et al. (2014) [35] and Chong et al. (2017) [40] also use the idea of handling the entity mentions collectively to annotate tweets semantically. However, due to a lack of context in tweets, such an approach does not work well considering single tweets. Therefore, both works consider a set of semantically related tweets to employ the collective approach. According to Huang et al. (2014) [35], it is challenging to create high quality labeled data to train supervised learning approaches for the EL task due to several factors, including missing and ambiguous resources in a KG and the difficulty to determine the prominence of the mentions in the text. Thus, the authors propose a graph-based semi-supervised learning approach to disambiguate mentions in tweets collectively. Their approach is based on three principles: (i) local compatibility between a mention and its candidate resources; (ii) coreference and (iii) semantic relatedness between different mentions. In Chong et al. (2017) [40], the authors consider that tweets posted geographically and timely close to each other can be semantically related. The proposed method builds a graph with tweets close to each other both in space and time. Moreover, the method considers that semantic relatedness among the entity candidates occurs both inside each tweet (Intra-tweet coherence) and among different tweets (Inter-tweet coherence).

Although a single tweet provides little context information, Kalloubi et al. (2016) [37] proposes a collective approach for disambiguating the mentions in single tweets. It is justified because the focus of their work is to annotate tweets to retrieve the ones satisfying a query provided by a user. A weighted graph is built, with the entity candidates for all mentions spotted as nodes and the relationships between them as edges. The entity candidates are obtained from DBpedia. Their EL method identifies the most relevant entity in the graph for the respective tweet. This entity, called the central node, is used to calculate the weights of the remaining nodes. An analogous process is applied to the queries of the user. Then, the weighted graph of the query is compared with the weighted graphs of several tweets. The tweet graphs most similar to the user query graph are considered the most relevant to the user and retrieved.

The collective approaches to disambiguate mentions exploit the links between concepts and entities of KGs. However, some KGs for closed domains (e.g., medical, enterprise) have only a few links, if any, between their entities. Consequently, approaches that rely heavily on such links produce unsatisfactory results using these KGs [38]. To circumvent this problem, Li et al. (2016) [38] propose an approach that gathers evidence of named entity mentions in the document to be annotated (e.g., TF-IDF score, relevant words around the entity mentions) and, with such evidence, produces a generative model that simulates the cross-document links among the entities in a KG. An extended version of the Gibbs Sampling is used to disambiguate the entities by inferring the entity that better describes each entity mention.

Different from the other collective approaches, Wei et al. (2019) [41] apply a candidate selection before the disambiguation step. The authors generate a graph with the entity candidates for each mention. Queries executed on Wikipedia and Freebase returns these candidates. The entity candidates are the nodes of the graphs, and they are connected when their respective Wikipedia pages are linked to each other. The authors apply PageRank to select the candidates in the entity candidate graph built for each mention. Only the candidates with higher ranks are considered in the disambiguation step, which is done by an FFNN (Feed-Forward Neural Network). The inputs of the FFNN are the word embeddings of the text; each entity candidate description encoded as a 128-dimension vector and the embedding of left contexts and right contexts (Dual) by Fixed-Size Ordinally Forgetting Encoding (FOFE) [51], i.e., Dual-FOFE. The use of the left and right context is also applied in the approach proposed by Liu et al. (2019) [43], which encode the entity embedding, based on the entity context and its description in Wikipedia pages. They employ these embeddings in their collective disambiguation approach. The entity embedding generation is divided into two neural network models: (i) a Long Short-Term Memory (LSTM) model to encode the entity context, that is composed by the left and right context of the entity mention and; (ii) a Convolutional Neural Network (CNN) model for encoding the entity description. The embeddings are fed to a local model, based on a novel CRF attention network, which produces a local score for the entity candidates. These local scores are used in a Forward-Backward algorithm to calculate the global score and disambiguate all the mentions collectively.

Although collective approaches may provide better results than non-collective approaches, Parravicini et al. (2019) [42] argue that the existing EL approaches are not scalable enough for application/domains that present real-time requirements. They consider as real-time if an EL approach can recognize entity mentions and disambiguate them in a whole text in less than one second. Therefore, the authors propose a collective EL approach based on graph embedding and scalable for real-time requirements. For the disambiguation, the authors use cosine similarity to verify the similarities among the entity candidates of all entity mentions. However, this approach is impracticable when the document presents a high number named entity mentions. As exemplified by them, if a document presents 10 named entity mentions, and each mention presents 10 entity candidates, they must evaluate 10^{10} distinct combinations. For circumventing this, they propose a "heuristic optimization algorithm based on state-space search exploration". This algorithm creates a few numbers of random combinations of entity candidates of distinct entity mentions (for simplicity, we will call these combinations as a tuple of entity candidates) and picks the one with the better score. Next, for the picked tuple, they randomly select a position in the tuple (the position represents an entity mention in the text) and verify if other entity candidates for the same mention have a better score for the tuple. In this case, the entity candidate is replaced by the better one. This optimization algorithm ends after a pre-specified number of runs or when the score of the tuple does not improve after a few runs. This optimization algorithm allows their approach to run in less than one second, and, therefore, is viable for real-time applications.

3.3.2. Embedding-based Methods

A recent strategy for disambiguation is the jointly use of different types of embeddings, like word and entity embeddings [52, 53, 54, 33, 55, 21, 56, 57, 44, 58, 59], document and graph embeddings [60], entity and knowledge embeddings [61], and word and knowledge embedding [62]. In Fang et al. (2016) [52], the authors combine the entity model with the word model by using two alignment techniques proposed by Wang et al. (2014) [63] (based on Wikipedia anchors and entity names, respectively) and one alignment technique from Zhong et al. (2015) [64] (based on entity descriptions). With the models aligned and a few features selected for disambiguation, the authors select the best entity candidate for a given mention in a two-layer disambiguation model. Similarly to Fang et al. (2016) [52], Yamada et al. (2016) [53] propose an embedding model based on skip-gram [65, 66] (skip-gram model, KB graph model, and anchor context model). Besides employing the embedding model, the authors also propose and exploit textual context similarity and coherence to disambiguate entity mentions by using the learning-to-rank algorithm GBRT (Gradient Boosted Regression Trees) [67].

Shi et al. (2020) [61] take a further step and aligns four embedding models to disambiguate named entity mentions in sentence level. This proposal employs Feature-Entity embedding, which represents the context of entity pages in Wikipedia by using Word2Vec. This model is similar to the entity embeddings used in other approaches. Mention-Entity embedding represents the context for a given mention already disambiguated (i.e., $\langle mention, entity \rangle$). This embedding model employs l_2-regularization and Hinge Loss. The knowledge embedding, by its turn, represents the facts in a KG, in this case, Yago. The authors propose knowledge embedding training similar to Word2Vec. Lastly, coherence embedding represents the interactions between disambiguated mentions in the same sentence. In Moreno et al. (2017) [54], the authors proposed an embedding model called Extended Anchor Text (EAT), which is also based on the model proposed by Mikolov et al. (2013) [66]. However, differently from [52, 53], the proposal of Moreno et al. (2017) [54] is based on one model.

Some approaches consider that both words and entities are in the same distributive space [52, 53, 54]. However, Chen et al. (2018) [33] criticize such an assumption. They consider that words and entities are in different distributive spaces because entity surface names can consist of multiple words, and the occurrence scales of words and entities in a text are different. Therefore, the authors propose a Bilinear Joint Learning Model (BJLM), an extension of the skip-gram model [65, 66]. According to the authors, the bilinear model *"simulates the interactions between the word distributive space and the entity distributive space"*.

Kolitsas et al. (2018) [21] propose a neural model that uses both embeddings. Their neural model is composed of a bi-LSTM (bidirectional Long Short-Term Memory) and shallows FFNN. The word embeddings of the mentions and the words around them are fed to the bi-LSTM to generate word embeddings aware of their context. Briefly, the context-aware word embeddings of the named entity mention, the entity embeddings of the entity candidates and other features relevant for EL are fed to an FFNN to get the local score of the entity candidates for a given mention. Lastly, the local score and a partially global mention-entity score are fed to the last FFNN, whose output is used to disambiguate mentions. Mueller & Durrett [57] also do not align the word and entity embeddings. The authors jointly train both embeddings by using the *word2vecf* technique [68] on Wikipedia pages. This allows both embeddings to be in the same distributive space. To perform the EL task, the authors employ the jointly trained word and entity embeddings and lexical features in a Gated Recurrent Unit (GRU) with an attention mechanism. The authors obtained state-of-the-art results in the WikilinksNED dataset [69].

In most KGs, different relations $r \in R$ connect entities. According to Le & Titov (2018) [55], these relations can improve the results of EL approaches. Thus, they propose a CRF model that considers word, entity, and relation embeddings. Furthermore, Le & Titov (2018) [55] proposes three ways to represent relations: general form, relation-norm, and mention-norm. Given a set M of mentions in a text and a set R of latent relations, the pairwise score (m_i, m_j), where $m_i, m_j \in M$, is given by the weighted sum of relation-specific pairwise scores. The relation-norm and mention-norm are the general form with normalization over, respectively, the relation and the mentions.

Besides relations that connect entities in a KG, most entities have at least one type (e.g., person, organization, place), which indicates their classification in a given ontology. However, Chen et al. (2020) [58] stress that existing approaches do not exploit sufficiently entity types in the disambiguation step of EL and propose to imbue entity types information into BERT (Bidirectional Encoder Representations from Transformers) pre-trained entity embedding [70]. They apply the entity similarity score calculated from the entity embeddings into the local context model of the disambiguation step proposed by Ganea & Hofmann (2017) [71]. Chen et al. (2020) [58] consider that the immediate context that surrounds a mention in its entity page (e.g., Wikipedia page) may summarize its types and replace the entity mention in its page by the token [MASK]. Then, they extract the uppermost layer representing the token [MASK] to represent the entity type.

Zhua & Iglesias (2018) [56] argue that current unsupervised EL approaches are not effective for short texts, like queries and social media posts. According to them, this occurs because these approaches depend mainly on features like context similarity and relatedness between the entities. However, short texts have limited context, and few entity mentions, limiting the use of such features. To circumvent these limitations, Zhua & Iglesias (2018) [56] propose an approach based on the contextual similarity between the mention and entity descriptions, which was introduced by them in the so-called Semantic Contextual Similarity-based NED (SCSNED). They also present an approach based on a new embedding model, called Category2Vec, which learns categories from joint embeddings of KG resources and words, based on the entity abstracts and entity categories. Both approaches achieve better results compared with other unsupervised EL approaches and present competitive results against EL approaches in general.

Instead of the joint use of the word and entity embedding, Sevgili et al. (2019) [60] propose the jointly use of document and graph embeddings. The authors present a neural model that exploits both embeddings to disambiguate the named entity mentions. A single FFNN layer composes their neural model, and its inputs are the document vector of the context that surrounds the named entity mention; the document vector of the named entity mention itself; the document vector of a long abstract of the entity candidate and; the graph embedding of the entity candidate. The doc2vec [72] technique generates the document embeddings using English Wikipedia pages, while the DeepWalk [73] technique generates the graph embeddings using DBpedia triples. Lastly, Sevgili et al. (2019) [60] improves the previous approach with graph embeddings. The improvement with graph embeddings results in slightly better precision and recall.

Differently from approaches that use different techniques to produce different embeddings and then align them, Oliveira et al. (2020) [62] jointly trains both word and knowledge embeddings and exploit them together in a deep neural network to disambiguate entity mentions in microblog posts. They do so by employing the fastText technique, which is capable to jointly train both word and knowledge embeddings in the same vector space. This allows skipping the alignment step performed by other approaches. To exploit both word and knowledge embeddings concomitantly, the authors replace the named entity mentions in the microblog posts by their respective entity candidates, one at a time. Word embeddings represent the words that surround the mentions, while knowledge embeddings represent the entity candidates. These embedded representations of microblog posts are fed into a bi-LSTM, which is followed by an FFNN. If an entity candidate fits the post context, the neural network classifies it as correct. The results surpass most state-of-the-art approaches.

Most EL approaches that use machine learning focus on existing labeled data for training. However, Le & Titov (2019) [59] argue that such approaches fail in domains where few labeled data exist, like scientific domains. To employ EL in such domains, they tackle the EL task as multi-instance learning [74]. In their approach, the multi-instance learning first classifies bags of examples, depending on if they contain or not the correct entity for a given named entity mention. Then, it classifies the instances of a bag that supposedly contains the correct entity. To achieve this, two sets of entity candidates are generated for each named entity mention. The first set, named E^+, is composed by entity candidates found by a surface-match heuristic proposed by Riedel et al. (2010) [75]. The heuristic guarantee that most of the time, the correct entity for a named entity mention is in E^+. The second set, named E^-, is composed of randomly retrieved entity candidates and does not contain the correct entity mention. To disambiguate the entity candidates, the authors propose a neural network model based on a bi-LSTM and an FFNN. The bi-LSTM encodes the context surrounding the named entity mentions, represented by word embeddings. The FFNN has as inputs the bi-LSTM's output and the entity embedding representation of the entity candidates. The binary noise detection classifier is trained jointly with the previous neural network model to improve the results.

4. Comparative Analysis of Holistic EL Approaches

Table 1 provides a comparison summary of proposals found in the literature about the EL task that presents some holism, as reported in Section 3. We list them in chronological ordering, presenting the columns according to relevant aspects of holism that we have proposed in Section 2. The column *External Input* refers to the databases, Knowledge Bases (KBs), or KGs that are used to semantically enrich the data to be annotated, i.e., the sources of resources that can semantically describe what is mentioned in the text. Column *NLP tasks* refers to NLP tasks that are combined or preceded the EL task. Lastly, the column *Method* refers to the methods used to disambiguate the entity candidates. Other methods and tools used to generate features or preprocessing the data are not shown in this table because they are outside the scope of this paper.

The first highlight of Table 1 is the column *Input*. In this column, it is possible to perceive that most of the works use Wikipedia as the source of the entities to semantically enrich the text to be annotated. Some works use KG (e.g., Freebase, DBpedia, Yago) to complement the information available in Wikipedia [28, 20, 41], to help in the generation of entity embeddings [52, 33, 55, 59], graph/knowledge embedding [42, 61, 62], or category embedding [56]. Although DBpedia is the Linked Open Data (LOD) version of Wikipedia [5, 6] (i.e., the KG version of Wikipedia), only the works [37, 56, 60, 42, 62] uses DBpedia directly in the EL task. The annotator proposed by Moro et al. (2014) [28] is the only one that uses the KG BabelNet [29], which combines Wikipedia and WordNet, to jointly perform EL and WSD.

Table 1: Comparison of holistic approaches for EL

Work	Input	NLP tasks	Method
Han et al. 2011 [34]	Wikipedia		Random Graph Walk
Wick et al. 2013 [31]	Wikilinks, Wikipedia	Entity Discovery	Markov Chain Monte Carlo
Moro et al. 2014 [28]	BabelNet	NER, WSD	Random Graph Walk with Restart
Huang et al. 2014 [35]	Wikipedia		Semi-supervised Graph Regularization
Guo & Barbosa 2014 [36]	Wikipedia		Random Graph Walk with Restart
Hua et al. 2015 [18]	Wikipedia		Ranking algorihtm based on user interest, entity popularity and entity recency
Luo et al. 2015 [20]	Wikipedia, Freebase, Sartori	NER	Semi-Conditional Random Fields extended for model entity distribution and mutual dependency over segmentation
Tran et al. 2015 [19]	Wikipedia	NER	Random Graph Walk
Kalloubi et al. 2016 [37]	DBpedia	NER	Graph Centrality Scoring
Ganea et al. 2016 [39]	Wikipedia		Markov Network (Factor Graph) + loopy belief propagation
Li et al. 2016 [38]	Linkless Wikipedia		Gibbs Sampling
Trani et al. 2016 [32, 12]	Wikipedia	Entity Saliency	Gradient Boosting Regression Tree
Fang et al. 2016 [52]	Wikipedia, Freebase		Logistic regression for two-layer model (Word Embedding, Knowledge Embedding)
Yamada et al. 2016 [53]	Wikipedia		Gradient Boosted Regression Trees (Word Embedding, Entity Embedding)
Chong et al. 2017 [40]	Wikipedia	NER (TweetNLP)	Objective function over a graph
Moreno et al. 2017 [54]	Wikipedia		Binary classifiers (Word Embedding, Entity Embedding)
Chen et al. 2018 [33]	Wikipedia, Freebase	Entity Saliency	Pairwise boosting regression tree (Word Embedding, Entity Embedding)
Le & Titov 2018 [55]	Wikipedia, Yago		Conditional random field, loopy belief propagation (Word Embedding, Entity Embedding)
Kolitsas et al. 2018 [21]	Wikipedia	NER	Shallow FFNN and LSTM (Word Embedding, Entity Embedding)
Zhu & Iglesias 2018 [56]	DBpedia		Semantic contextual similarity algorithm (Word Embedding, Category Embedding)
Mueller & Durrett 2018 [57]	Wikipedia		GRU (Word Embedding, Entity Embedding)
Martins et al. 2019 [23]	Wikipedia	NER	Stack-LSTM (Word Embedding, Entity Embedding)
Sevgili et al. 2019 [60]	Wikipedia, DBpedia		FFNN (Word Embedding, Graph Embedding)
Wang & Iwaihara 2019 [22]	Wikipedia	NER	TNN and CNN (Word embedding)
Wei et al. 2019 [41]	Wikipedia, Freebase		FFNN (Word Embedding)
Parravicini et al. 2019 [42]	DBpedia		Semantic similarity (Graph Embedding) and state-space search heuristic
Liu et al. 2019 [43]	Wikipedia		Forward-Backward algorithm (Entity Embedding)
Fang et al. 2019 [44]	Wikipedia		LSTM, Reinforcement Learning (Word Embedding, Entity Embedding)
Yang et al. 2019 [47]	Wikipedia		FFNN, Reinforcement Learning (Word Embedding, Entity Embedding)
Vaigh et al. 2019 [45]	Wikipedia, BaseKB		Binary logistic regression classifier (Word Embedding)
Phan et al. 2019 [46]	Wikipedia		Minimum Spanning Tree (Word Embedding, Entity Embedding)
Le & Titov 2019 [59]	Freebase		bi-LSTM, FFNN (Word Embedding, Entity Embedding)
Chen et al. 2020 [58]	Wikipedia		Conditional Random Field (Word Embedding, Entity Embedding)
Shi et al. 2020 [61]	Wikipedia, Yago		Vector Similarity (Entity Embedding, Knowledge Embedding)
Oliveira et al. 2020 [62]	DBpedia		bi-LSTM (Word Embedding, Knowledge Embedding)
Rama-Maneiro et al. 2020 [48]	Wikipedia, DBpedia		Graph Centrality Scoring, Topic similarity

Although most of the KGs are encoded as RDF (Resource Description Framework) triples, and, therefore, are machine-readable, we reasoning that

Wikipedia is still widely employed by EL approaches for several reasons, including:

- Most KGs have a slow update-cycle[13], and some have been discontinued (e.g., Freebase). Meanwhile, it is possible to get updated dumps from Wikipedia every month[14];

- Although several KGs, like DBpedia and BabelNet, include LOD versions of information extracted from Wikipedia, they present far less textual content than the latter. This makes their use difficult for several approaches that depend on such textual content, like approaches that use word and entity embedding;

- Some metadata about the Wikipedia pages, like page views, are used as features to disambiguate entity candidates. There is no guarantee that such metadata will be available in the KGs.

Regarding the column *NLP tasks*, it is possible to notice that most of the works do not present any other NLP tasks besides EL, frequently disregarding even the NER task. This happens because most of the works consider that all the named entity mentions are already recognized. Therefore, such works focus solely on the named entity mention disambiguation. The works that present the NER task propose an end-to-end EL approach, i.e., they propose a new approach for both NER and EL. Lastly, only a few works present an NLP task besides the NER task. More specifically, the ED and ES tasks.

Since Table 1 is sorted chronologically, we highlight in its last column, *Method*, that the EL approaches are shifting from graph-based methods to approaches that use embeddings of words and entities. Embedding techniques are capable to model local and global interactions between entities or words in low-dimension vectors [76, 77, 78, 79]. Therefore, the use of such embeddings enable approaches to achieve the same or better results as using graph-based methods, with higher scalability. Moreover, some works employ graph embedding [60] and knowledge embedding [52, 61, 62] instead of entity embedding, while Parravicini et al. (2019) [42] uses only graph embedding. These works are essential to show that embeddings generated from DBpedia

[13]Until the submission of this paper, the last public data available from DBpedia is from 2016, while Yago is 2017

[14]https://dumps.wikimedia.org/

are a viable option for EL approaches. Differences between graph embedding and knowledge embedding are presented in Section 5.2.

Given the variety of currently available EL approaches with several distinguishing characteristics, it may be challenging to decide which one is the most appropriate for a specific application or domain. Thus, we propose a Decision Tree (DT) to help make such a choice based on features that correspond to our 3 holistic aspects of EL approaches, as illustrated in Figure 3. If the disambiguated entities need to be semantically coherent, then it is more appropriate to select an approach that makes collective disambiguation instead of non-collective. The choice also depends on the document type: microtext (e.g., social media posts, text snippets) versus any longer or more formal text (e.g., news, books, articles), which is expected to carry more context information and to have fewer typos, grammatical errors, slangs and other kinds of noise. Finally, and optionally, having the NER task integrated with the EL solution may be convenient, for example, to avoid setup of distinct tools. Leaves of the DT in Figure 3 refer to groups of approaches sharing the same selective features. Due to the number of approaches in each group, the citations of the respective works, along with links to repositories containing their open-source code (when available) or links to the GitHub profiles of their author(s) are listed in Table 2.

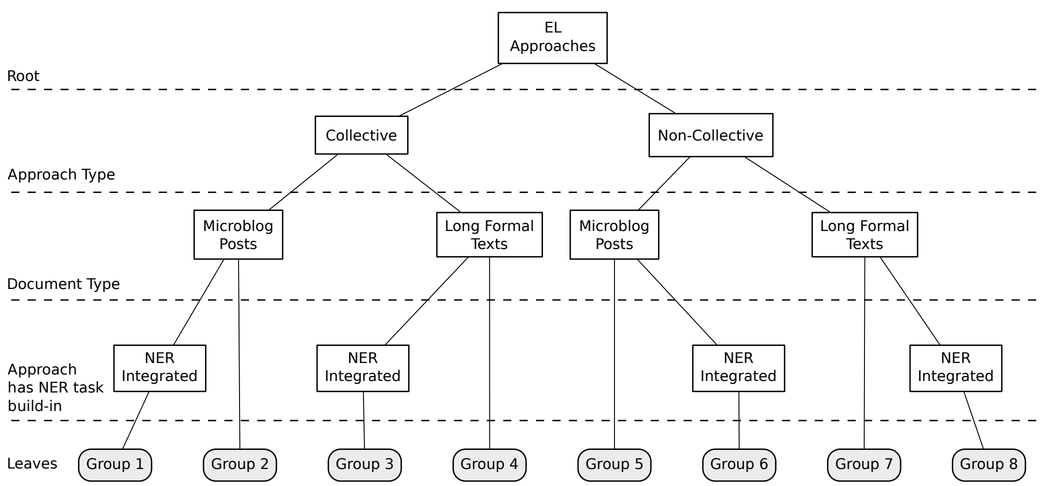

Figure 3: Proposed DT to support the selection of EL approaches. White boxes refer to approach characteristics considered in each DT level. Horizontal dashed lines separate the levels. Gray rounded boxes in the leaves refer to groups of works considered analogous with respect to our decision criteria. Works within each group are listed in Table 2.

Table 2: Groups of works determined by our DT. When available, a link to the source code of the respective approach is provided in the second column. Otherwise, the Github profile of each author is provided in the third column, as a link on the respective full name.

Work	Source code repository	Github of authors
Group 1		
Kalloubi et al. 2016 [37]		https://github.com/fahdkalloubi-ENSA
Chong et al. 2017 [40]		William Cohen
Group 2		
Huang et al. 2014 [35]		Chin-Yew Lin
Group 3		
Moro et al. 2014 [28]		Alessandro Raganato, Roberto Navigli
Group 4		
Han et al. 2011 [34]		
Guo & Barbosa 2014 [36]		Zhaochen Guo, Denilson Barbosa
Li et al. 2016 [38]		Shulong Tan, Huan Sun, Dan Roth
Ganea et al. 2016 [39]	https://github.com/dalab/pboh-entity-linking	
Vaigh et al. 2019 [45]	https://gitlab.inria.fr/celvaigh/ukbscael2019	
Wei et al. 2019 [41]		
Fang et al. 2019 [44]		
Parravicini et al. 2019 [42]		Alberto Parravicini, Davide B Bartolini, Rhicheek Patra, Marco D. Santambrogio
Phan et al. 2019 [46]		Ti Ray, Jialong Han
Yang et al. 2019 [47]	https://github.com/YoungXiyuan/DCA	
Liu et al. 2019 [43]		
Rama-Maneiro et al. 2020 [48]		Efren Rama-Maneiro, Juan C. Vidal
Group 5		
Oliveira et al. 2020 [62]	https://github.com/ItaloLopes/optic	
Group 6		
Tran et al. 2015 [19]		Tuan Tran, Nam K. Tran, Asmelash T. Hadgu, Robert Jäschke
Group 7		
Wick et al. 2013 [31]		Sameer Singh, Harshal Pandya, Andrew McCallum
Hua et al. 2015 [18]		Wen Hua
Trani et al. 2016 [32, 12]		
Fang et al. 2016 [52]		Dilin Wang
Yamada et al. 2016 [53]		
Moreno et al. 2017 [54]		Jose Moreno, Romaric Besançon, Romain Beaumont, Anne-Laure Ligozat, Xavier Tannier
Chen et al. 2018 [33]		
Le & Titov 2018 [55]	https://github.com/lephong/mulrel-nel	
Zhu & Iglesias 2018 [56]		Carlos A. Iglesias
Mueller & Durrett 2018 [57]	https://github.com/davidandym/wikilinks-ned	
Sevgili et al. 2019 [60]	https://github.com/uhh-lt/kb2vec	
Le & Titov 2019 [59]	https://github.com/lephong/dl4el	
Chen et al. 2020 [58]		Chen-Yew Lin, Chen Shuang, Junpeng Wang
Shi et al. 2020 [61]		
Group 8		
Luo et al. 2015 [20]		Chin-Yew Lin
Kolitsas et al. 2018 [21]	https://github.com/dalab/end2end_neural_el	
Martins et al. 2019 [23]		Pedro H. Martins, Zita Marinho, André F. T. Martins
Wang & Iwaihara 2019 [22]		

Notice that Table 2 covers all the works discussed in Sections 3 and 4, grouping them according with the DT of Figure 3. Although not providing links to source code, Moro et al. (2014) [28] and Rama-Maneiro et al. (2020) [48] provide Web applications that demonstrate their proposals, namely Babelfy[15] and ABACO[16], respectively. Some papers provide Github links to the tools and models that support their approaches, such as Trani et al. (2016) [32, 12] with Elianto[17] and Dexter[18] and, Yamada et al. (2016) [53] providing their embedding model Wikipedia2vec[19].

In addition to the criteria considered in our DT (Figure 3) and the availability of the approach as open-source, the quality of the results may also be a crucial decision factor. Thus, in the following (Section 4.1), we provide a performance comparison summary of the approaches analyzed in this survey.

4.1. Evaluation of Holistic EL Approaches

The works analyzed in this survey use several distinct metrics to evaluate their EL approaches. However, we noticed that the most used one is the F1 score. It is is the harmonic mean of precision and recall and tolerates uneven class distributions. Thus, we first compare the approaches using the F1 score, considering the formal definition of the EL task presented in [39, 48].

The majority of the analyzed works that use the F1 score to evaluate their approaches also use the GERBIL benchmark system [80] to produce automatic and reliable evaluations. GERBIL calculates the micro and macro F1 scores to better evaluate the performance of EL approaches. While the micro F1 score evaluates performance over the whole dataset, the macro F1 score evaluates performance for each document and takes the average.

Table 4 presents the F1 score of works analyzed in this survey (listed in the first column of the table, by the chronological order of their publications) that use this metric to evaluate performance on distinct datasets (listed in alphabetic order in the second line of the table header). The last column of Table 4 indicates if the performance of the respective approach was evaluated using a benchmark system or not. The micro F1 score is provided by all works (though not for every dataset listed in the table), while only a few works also provide the macro F1 score (which may appear below the respective micro F1 score). When the respective article provides variations of the F1 score

[15]http://babelfy.org/
[16]https://tec.citius.usc.es/abaco/
[17]https://github.com/dexter/elianto
[18]https://github.com/dexter/dexter
[19]https://github.com/wikipedia2vec/wikipedia2vec

Table 3: EL approaches evaluated with other metrics instead of F1 score. The symbol ≈ indicates that the values are approximate because the works provide them only in graphs.

Metric	Dataset	Wick et al. 2013	Luo et al. 2015	Hua et al. 2015	Tran et al. 2015	Fang et al. 2016	Li et al. 2016	Yamada et al. 2016	Chong et al. 2017	Chen et al. 2018	Mueller & Durrett 2018	Wei et al. 2019	Shi et al. 2020
Accuracy	Custom Tweets (Hua)			≈0.62									
	Custom Tweets (Li)						≈ 0.79						
	TAC-KBP 2009						≈ 0.81						
	CoNLL (pershina)							0.93					
	CoNLL 2013							0.91					0.83
	TAC 2010							0.85					
	CoNLL									0.93			
	TAC 2010									0.88			
	KBP 2010					0.88							
	Wikilinks dev										0.74		
	Wikilinks test										0.75		
	Wikipedia + Wikilinks (Custom)	0.98											
P@1	CoNLL 2013		0.84										
P@5	Custom Tweets (Tran)				0.64								
P@15					0.49								
MAP@15					0.43								
CEAFmC F1	TAC-EDL 2016											0.67	
	TAC-EDL 2017											0.70	
NERLC F1	TAC-EDL 2016											0.65	
	TAC-EDL 2017											0.68	
Change Ratio	Custom Tweets (Chong)								12				

(e.g., due to distinct parameter settings in their method), we only present the highest value that was obtained.

Table 3 summarizes the performance of the works analyzed in this survey that, instead of the standard F1 score, use other metrics (listed in the first column of the table) to evaluate their approaches. Notice that the next most used evaluation metric is by far the accuracy. Differently from the F1 score, the accuracy does not tolerate uneven class distributions. Therefore, depending on the dataset used to evaluate the EL approach, the accuracy may provide a less reliable measure of the performance.

In Tran et al. (2015) [19], the authors consider the EL task as a ranking problem. Therefore, they evaluate if the correct entity for each mention has the highest score among the n best results returned by their approach. In their work, they consider the following metrics: precision at 5 (P@5), precision

Table 4: Evaluation of holistic approaches for EL using F1 micro score (F1@MI) and F1 macro score (F1@MA). The scores are presented as F1@MI/F1@MA. Cells with just one value refer to F1@MI.

Work	Dataset																														Benchmark
	ACE 2004	AQUAINT	MSNBC	AIDA/CoNLL-Test B	N3-Reuters-128	N3-RSS-500	AIDA/CoNLL	WNED-CWEB	AIDA/CoNLL-Test A	DBpediaSpotlight	KORE50	WNED-WIKI	CoNLL 2003	IITB	Micropost2014-Test	Micropost216-Test	AIDA/CoNLL-Training	Custom Tweets (Meij)	Derczynski	Micropost2014-Train	Micropost2015-Test	OKE 2015	OKE 2015 Task 1 eval set	OKE 2016	OKE 2016 Task 1 eval set	OKE 2018 Task 1 train set	OKE 2018 Task 2 train set	OKE 2018 Task 4 train set	TAC-EDL 2015	Wikinews	
Han et al. 2011														0.73																	
Moro et al. 2014							0.82				0.71																				
Huang et al. 2014																		0.525													
Guo & Barbosa 2014	0.87/0.88	0.88/0.88	0.92/0.92																												
Ganea et al. 2016	0.87/0.90	0.86/0.86	0.89/0.89	0.87/0.86	0.76/0.83	0.71/0.78	0.86/0.86		0.86/0.85	0.79/0.80	0.61/0.55			0.62/0.61	0.74/0.84		0.86/0.87			0.73/0.81											GERBIL
Trani et al. 2016													0.72																	0.72	
Fang et al. 2016	0.80	0.85	0.75																												
Moreno et al. 2017																													0.74		
Le & Titov 2018	0.90	0.88	0.93	0.93				0.77				0.78																			
Kolitsas et al. 2018			0.73/0.72	0.82/0.82	0.54/0.54	0.46/0.42			0.86/0.89		0.40/0.46								0.48/0.42			0.62/0.66		0.57/0.58							GERBIL
Zhu & Iglesias 2018						0.59										0.59															
Martins et al. 2019							0.81/0.81																								GERBIL
Sevgili et al. 2019					0.66/0.61					0.79/0.79																					GERBIL
Wang & Iwaihara 2019	0.76	0.76					0.81																								GERBIL
Parravicini et al 2019	0.84	0.86	0.92		0.82	0.72																									GERBIL
Liu et al. 2019	0.86	0.87		0.87				0.73	0.91																						
Fang et al. 2019	0.91	0.87	0.92					0.78				0.82																			
Yang et al. 2019	0.90	0.88	0.94					0.75				0.78																			
Vaigh et al. 2019				0.87	0.79	0.79			0.90																						GERBIL
Phan et al. 2019	0.88	0.87	0.91		0.85	0.82				0.84	0.78				0.81																GERBIL
Le & Titov 2019							0.37																								
Chen et al. 2020	0.88	0.89	0.93	0.93				0.77				0.80																			
Oliveira et al. 2020															0.29/0.57	0.50/0.95					0.33/0.45										GERBIL
Rama-Maneiro et al. 202	0.78	0.76	0.85	0.70	0.69	0.67				0.83				0.44									0.71		0.73	0.76	0.70	0.84			GERBIL

at 15 (P@15) and, Mean Average Precision (MAP).

The remaining metrics, appearing at the bottom of the first column of Table 3, are variations of the F1 score. In Wick et al. (2013), the authors tackle both the Entity Linking and the Entity Discovery tasks. Due to this, they employ the Pairwise F1 metric for performance evaluation. Unlike the standard F1 score, the pairwise F1 score takes into account pairs of entity mentions in the text document considered to refer to the same entity. The metrics CEAFmC F1 and NERLC F1 are the standard F1 score restricted to specific types of experiments performed on the TAC dataset. CEAFmC denotes *typed_mention_ceaf* and NERLC denotes *strong_typed_all_match*.

Among all the works analyzed in this survey, only Kalloubi et al. (2016) [37] is not present in either Table 4 nor Table 3. It happens because the authors propose an EL approach to semantically enrich tweets for improving their retrieval. Therefore, their experiments evaluate the quality of their retrieval approach and not the EL task itself.

The authors of Chong et al. (2017) [40] aim to measure how their collective EL approach, based on temporal and geospatial features, compare with a non-collective EL approach for tweets. Therefore, they measure the ratio of positive and negative changes in their approach over a non-collective approach. They consider as a positive change when their approach fixes an incorrectly disambiguated entity yield by the baseline. Conversely, a negative change is when their approach transforms a correctly disambiguated entity into an incorrect one.

Finally, Table 5 provides pointers to further details about the datasets mentioned in Tables 4 and 3, as links to the respective home pages, when they are available. When such a link is unavailable, we provide a reference to the paper or challenge in which the dataset appears.

5. Potential Pillars for Future Holistic Approaches

For the best of our knowledge, the trends for EL holistic systems have not been adequately identified and described yet. Therefore, this section describes the pillars that we have identified from our bibliographical review and previous experience.

5.1. The General Semantic Annotation Process

We have observed that a sequence of tasks always occurs in the EL approaches that we have analyzed, independently of the data used and the

Table 5: Datasets employed by EL approaches to evaluate their performance

Dataset Name	Link
ACE 2004	https://cogcomp.seas.upenn.edu/page/resource_view/4
AQUAINT	https://catalog.ldc.upenn.edu/LDC2002T31
MSNBC	https://cogcomp.seas.upenn.edu/page/resource_view/4
AIDA/CoNLL	https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/aida/downloads/
N3-Reuters-128	https://github.com/AKSW/n3-collection
N3-RSS-500	https://github.com/AKSW/n3-collection
IITB	http://www.cse.iitb.ac.in/~soumen/doc/CSAW/Annot/
WNED-CWEB	Guo & Barbosa 2018 [81]
CoNLL 2013	http://www.cnts.ua.ac.be/conll2003/
DbpediaSpotlight	http://www.yovisto.com/labs/ner-benchmarks/
KORE50	http://www.yovisto.com/labs/ner-benchmarks/
WNED-Wiki	Guo & Barbosa 2018 [81]
Microposts2014	http://scc-research.lancaster.ac.uk/workshops/microposts2014/
Custom Tweets (Meij)	Meij et al. 2012 [82]
Derczynski	http://www.derczynski.com/sheffield/resources/ipm_nel.tar.gz
Microposts2015	http://scc-research.lancaster.ac.uk/workshops/microposts2015/
Microposts2016	http://microposts2016.seas.upenn.edu/challenge.html
OKE 2015	Open Knowledge Extraction at ESWC 2015
OKE 2016	Open Knowledge Extraction at ESWC 2016
OKE 2018	https://project-hobbit.eu/open-challenges/oke-open-challenge/
TAC-EDL 2015	https://tac.nist.gov/2015/KBP/data.html
TAC KBP 2010	https://tac.nist.gov/2010/KBP/
Wikipedia: test	Eshel et al. 2017 [69]
Wikinews	Trani et al. (2016) [32]
Custom Tweets (Hua)	Hua et al. 2015 [18]
Custom Tweets (Li)	Li et al. 2013 [83]
TAC KBP 2009	http://pmcnamee.net/kbp.html
CoNLL (Pershina)	Pershina et al. 2015 [84]
Wikilinks	http://www.iesl.cs.umass.edu/data/data-wiki-links
Custom Tweets (Tran)	Tran et al. 2015 [19]
Wikipedia + Wikilinks	Wick et al. 2013 [31]
TAC EDL 2016	http://nlp.cs.rpi.edu/kbp/2016/
TAC KBP 2017	http://nlp.cs.rpi.edu/kbp/2017/

specific methods employed to realize these tasks. It can be regarded as a general process for EL, as shown in Figure 4. Firstly, for several reasons, the *Text documents* given as input for EL can be noisy (e.g., due to spelling errors), and the *External inputs* used to help the EL task can be heterogeneous (e.g., because they may be from different sources). Thus, a *Preprocessing* stage is usually required for cleaning these data and standardizing them in some format. The next stage is to extract the relevant features from the preprocessed data (*Features Extraction*). Looking at resources from external inputs (e.g., Wikipedia, KGs) may contribute to this task to produce good results. The features extracted are used in the *Candidates Selection & Disambiguation* task to disambiguate the possible entity candidates for each named entity mention. This process can be generalized for semantic annotation in general, just by changing the implementation of its tasks and sometimes allowing other kinds of data to be annotated instead of only *Text documents*. The sequence of tasks remains the same, even though each one can be done in very different ways.

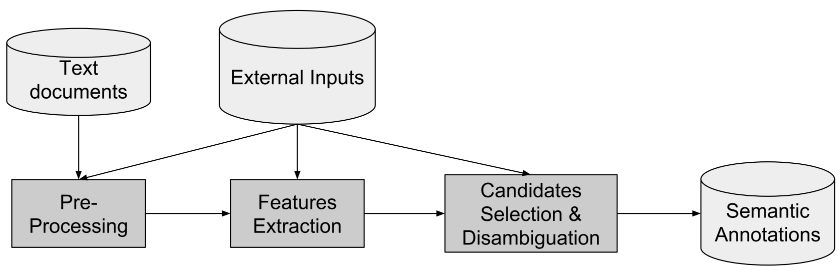

Figure 4: General process for EL.

5.2. Use of KGs and Knowledge Embedding

As presented in Sections 3.2 and 4, only a few works effectively use a KG (e.g., DBpedia, Yago, Freebase, Babelnet) as an external input for EL. Although we present several reasons for existing approaches to choose Wikipedia instead of KGs as their primary external input, we envision that knowledge bases such as KGs are essential for future holistic EL approaches. According to the papers [85, 86], KGs are multi-relational graphs that store facts about entities and their relations. These facts are represented as triples (*head entity, relation, tail entity*), usually encoded in RDF [7, 5, 6]. Therefore, KGs provide knowledge about entities in a standardized and machine-processable way. Future holistic EL approaches can take advantage of KGs encoded as RDF triples to quickly exchange the KG (extract) used for EL in accordance with the domain of the text to be annotated. For example, DBpedia can be used to annotate news and social media posts, as it presents entities of several domains. Meanwhile, for medical reports, which deal with more specialized knowledge, EL approaches can use medical knowledge graphs [87]. Moreover, the adoption of KGs in future holistic EL approaches enables the use of knowledge embedding on them.

Knowledge embedding [86] is a specialized topic from graph embedding [79], whose objective is to embed the entities and relations of a KG in low-dimension vectors. From the works analyzed in Section 2 and compared in Section 4, only Fang et al. (2016) [52] employs knowledge embeddings. However, we believe that the full potential of knowledge embedding in the EL task has not been fully explored yet. Knowledge embedding techniques can capture local, long-range, and global statistics of dependencies present in KGs into embeddings [85]. These dependencies may be useful for EL approaches that disambiguate entities collectively. Moreover, we envision that works that employ graph embedding [60, 42] may improve their results by using knowledge embedding, as shown in the papers [61, 62]. This may happen because graph embedding techniques, like DeepWalk and node2vec [88], are meant for any graph and, therefore, may not exploit effectively KG features (e.g., a high number of distinct relations) as well as knowledge embedding. Lastly, word embedding and knowledge embedding techniques, like fastText [89], are progressing to allow training models not only with facts but also with textual properties of the entities, like entity labels and abstracts. Such combinations of word embeddings with knowledge embeddings may enable improved EL approaches.

5.3. Building and Exploiting Historical Contexts

A historical context captures the most critical entities for a specific (group of) agent(s) involved in the production of the text to be annotated (e.g., author of a book, social media user). The knowledge and experience expressed in historical contexts can help to disambiguate highly ambiguous named entity mentions. For example, considering the left tweet in Figure 2, if the historical context of the tweet sender presents entities related to basketball, this may help to disambiguate the mention "Jordan" to the basketball player *Michael J. Jordan*. Historical context can also include entities from old books of an author and help to identify mentions to these old entities in his/her new book. Differently, from the previous potential pillar, some works propose concepts similar to historical contexts, like semantic profiles [90, 91] for social media users. However, to the best of our knowledge, EL approaches have not considered the use of such profiles or historical contexts.

5.4. A Reference Approach for Holistic EL

The combination of potential pillars in a seamless process, where each pillar can efficiently exploit each other, can further enhance the benefits of holism for the EL task. Therefore, we propose as last pillar a *Reference Approach for Holistic EL* based on the pillars previously described.

Figure 5 illustrates the reference approach that we propose for holistic EL. It derives from the generic EL process presented in Figure 4, but supports all the pillars for semantic approaches, and allows cycles for capturing historical contexts from annotations and use them to disambiguate entity candidates and semantic expansion. Each stage of this process can be adapted to fit some text type (e.g., social media posts, news) or EL approach.

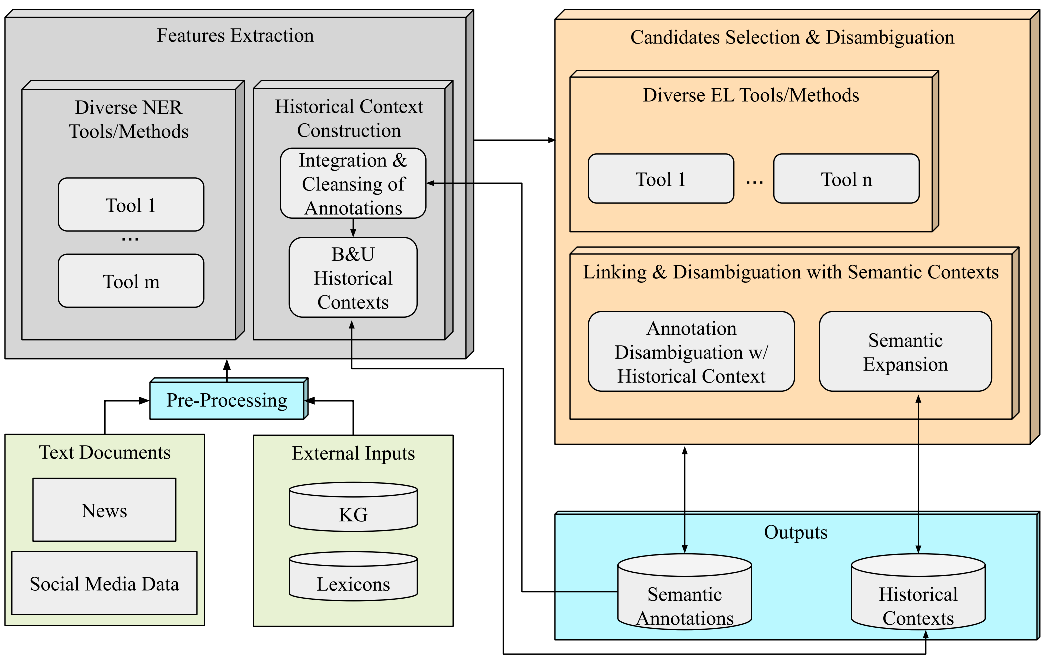

Figure 5: Reference Approach for Holistic EL. (Color required)

In *Feature Extraction* task, several tools and methods can be used to recognize different kinds of named entity mentions (e.g., people, places) in a text. Work-flows can combine them to suit some specific domain better. Besides the recognizing of named entity mentions, we also consider the building and updating historical contexts as feature extraction.

The building of historical contexts takes as input semantic annotations previously created and existing historical contexts in case if it is necessary to update them. Several semantic annotation features can be used to build semantic contexts, like the entity pointed by them, their creation timestamp, which user or application created the annotation and others. However, it is necessary to clean and integrate the semantic annotations before their use in historical contexts. A repository of semantic annotations can have inconsistency among its annotations. For example, for the same mention in a text document, two EL tools can disambiguate to different entities (e.g., the mention “Jordan” in Figure 2 can point to the country *Jordan* or the basketball player *Michael J. Jordan*). The output of this step is a new historical context for an agent or an updated existing historical context.

Similarly to the *Features Extraction* stage, the *Resource Linking & Disambiguation* stage also can employ several EL tools/approaches. However, these tools focus on disambiguating the mentions recognized in the *Features Extraction* stage. Different from the existing works, we propose the use of historical contexts in the EL task. The entity candidates considered can be compared with the preference of an agent expressed in historical contexts already built to better disambiguate the named entity mentions.

Lastly, we propose a step called *Semantic Expansion*. We envision that historical contexts also can express correlations between some entities. Such correlations can be used to generate new semantic annotations not explicitly expressed in the text. For instance, most of the people go to restaurants to eat or go to shopping malls to work or buy things. Therefore, it is possible to infer the activity of an agent from the places visited by him/her and vice versa. This step adds more context to the text, and that can be useful for data with little contexts, like social media posts.

6. Conclusions and Future Directions

Holistic approaches have the potential to boost EL by exploiting several data features and processing methods to make the highest possible number of semantically coherent links. In this paper, we reviewed and compared EL approaches that present some (potential for) holism, aiming to motivate, inspire, and give some directions for research in this field. We classified these approaches according to holism aspects that we have identified in our studies. Besides, we proposed potential pillars for future holistic EL approaches and a reference approach for holistic EL that exploits all these pillars.

Some of the analyzed EL approaches already employ some holism. However, these approaches do not adequately cover all the potential pillars proposed in this paper. For example, regarding the variety of data features that can be exploited in the EL process, few approaches exploit the fact that some social media posts (e.g., tweets) can be associated with geographic coordinates or labels of specific places, which can help to disambiguate some named entity mentions. Moreover, holistic EL approaches could consider historical contexts determined by previous dependable annotations while exploiting cutting-edge technologies like KGs and embeddings for collective disambiguation of entities based on coherence. Future approaches for EL could also tackle challenges such as:

- multiple and dynamic surface names, i.e., terms used for referring to an entity;
- noisy text, i.e. texts with typos, grammatical errors, slangs, etc.;
- lack of contextual information of certain kinds of text (e.g., social media posts), which might be compensated by considering the context determined by the history of annotations;
- efficient and effective use in holistic EL approaches of the knowledge present in a variety of models such as KGs and embeddings;
- combine EL with several annotation tasks and approaches which are based on a myriad of data, data features and annotation methods.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and 88881.189286/2018-01 of program PDSE, and 88881.121467/2016-01 of program Senior Internship, and by the Brazilian National Council for Scientific and Technological Development (CNPq) (grant number 385163/2015-0 of the program CNPq/INCT-INCoD).

References

[1] A. H. Laender, B. A. Ribeiro-Neto, A. S. da Silva, J. S. Teixeira, A brief survey of web data extraction tools, ACM Sigmod Record 31 (2) (2002) 84–93.

[2] K. Bontcheva, D. P. Rout, Making sense of social media streams through semantics: A survey, Semantic Web 5 (5) (2014) 373–403.

[3] P. Oliveira, J. Rocha, Semantic annotation tools survey, in: 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), IEEE, 2013, pp. 301–307.

[4] D. Moussallem, R. Usbeck, M. Röeder, A.-C. N. Ngomo, Mag: A multilingual, knowledge-base agnostic and deterministic entity linking approach, in: Proceedings of the Knowledge Capture Conference, 2017, p. 9.

[5] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, Dbpedia: A nucleus for a web of open data, in: The Semantic Web, Springer Berlin Heidelberg, 2007, pp. 722–735.

[6] J. Lehmann, C. Bizer, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, S. Hellmann, DBpedia - a crystallization point for the web of data, Journal of Web Semantics 7 (3) (2009) 154–165.

[7] M. Fabian, K. Gjergji, W. Gerhard, Yago: A core of semantic knowledge unifying wordnet and wikipedia, in: Proceedings of the 16th International Conference on World Wide Web, Association for Computing Machinery, 2007, pp. 697–706.

[8] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor, Freebase: a collaboratively created graph database for structuring human knowledge, in: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, ACM, 2008, pp. 1247–1250.

[9] C. Stadler, J. Lehmann, K. Höffner, S. Auer, Linkedgeodata: A core for a web of spatial open data, Semantic Web 3 (2012) 333–354.

[10] D. Vrandečić, M. Krötzsch, Wikidata: a free collaborative knowledgebase, Communications of the ACM 57 (10) (2014) 78–85.

[11] W. Shen, J. Wang, J. Han, Entity linking with a knowledge base: Issues, techniques, and solutions, IEEE Transactions on Knowledge and Data Engineering 27 (2) (2015) 443–460.

[12] S. Trani, C. Lucchese, R. Perego, D. E. Losada, D. Ceccarelli, S. Orlando, Sel: A unified algorithm for salient entity linking, Computational Intelligence 34 (1) (2018) 2–29.

[13] M. A. Khalid, V. Jijkoun, M. De Rijke, The impact of named entity normalization on information retrieval for question answering, in: Proceedings of the IR Research, 30th European Conference on Advances in Information Retrieval, Springer-Verlag, 2008, pp. 705–710.

[14] W. Zhang, J. Su, C. L. Tan, W. T. Wang, Entity linking leveraging: automatically generated annotation, in: Proceedings of the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics, 2010, pp. 1290–1298.

[15] C. Choudhay, C. ORiordan, A graph-based collective linking approach with group co-existence strength, in: Proceedings for the 26th AIAI Irish Conference on Artificial Intelligence and Cognitive Science, CEUR-WS, 2018, pp. 267–278.

[16] G. Wu, Y. He, X. Hu, Entity linking: an issue to extract corresponding entity with knowledge base, IEEE Access 6 (2018) 6220–6231.

[17] L. Derczynski, D. Maynard, G. Rizzo, M. Van Erp, G. Gorrell, R. Troncy, J. Petrak, K. Bontcheva, Analysis of named entity recognition and linking for tweets, Information Processing & Management 51 (2) (2015) 32–49.

[18] W. Hua, K. Zheng, X. Zhou, Microblog entity linking with social temporal context, in: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Association for Computing Machinery, 2015, pp. 1761–1775.

[19] T. Tran, N. K. Tran, T. H. Asmelash, R. Jäschke, Semantic annotation for microblog topics using wikipedia temporal information, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2015, p. 97106.

[20] G. Luo, X. Huang, C. yew Lin, Z. Nie, Joint named entity recognition and disambiguation, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015, pp. 879–888.

[21] N. Kolitsas, O.-E. Ganea, T. Hofmann, End-to-end neural entity linking, in: Proceedings of the 22nd Conference on Computational Natural Language Learning, Association for Computational Linguistics, 2018, pp. 519–529.

[22] Q. Wang, M. Iwaihara, Deep neural architectures for joint named entity recognition and disambiguation, in: 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE, 2019, pp. 1–4.

[23] P. H. Martins, Z. Marinho, A. F. Martins, Joint learning of named entity recognition and entity linking, arXiv preprint arXiv:1907.08243 (2019).

[24] G. D. Forney, The viterbi algorithm, Proceedings of the IEEE 61 (3) (1973) 268–278.

[25] P.-H. Li, R.-P. Dong, Y.-S. Wang, J.-C. Chou, W.-Y. Ma, Leveraging linguistic structures for named entity recognition with bidirectional recursive neural networks, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2017, pp. 2664–2669.

[26] C. Dyer, M. Ballesteros, W. Ling, A. Matthews, N. A. Smith, Transition-based dependency parsing with stack long short-term memory, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, 2015, pp. 334–343.

[27] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, C. Dyer, Neural architectures for named entity recognition, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, 2016, p. 260270.

[28] A. Moro, A. Raganato, R. Navigli, Entity linking meets word sense disambiguation: a unified approach, Transactions of the Association for Computational Linguistics 2 (2014) 231–244.

[29] R. Navigli, S. P. Ponzetto, Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network, Artificial Intelligence 193 (2012) 217–250.

[30] H. Tong, C. Faloutsos, J.-Y. Pan, Fast random walk with restart and its applications, in: Proceedings of the Sixth International Conference on Data Mining, IEEE Computer Society, 2006, pp. 613–622.

[31] M. Wick, S. Singh, H. Pandya, A. McCallum, A joint model for discovering and linking entities, in: Proceedings of the 2013 workshop on Automated knowledge base construction, 2013, pp. 67–72.

[32] S. Trani, D. Ceccarelli, C. Lucchese, S. Orlando, R. Perego, Sel: a unified algorithm for entity linking and saliency detection, in: Proceedings of the 2016 ACM Symposium on Document Engineering, Association for Computing Machinery, 2016, pp. 85–94.

[33] H. Chen, B. Wei, Y. Liu, Y. Li, J. Yu, W. Zhu, Bilinear joint learning of word and entity embeddings for entity linking, Neurocomputing 294 (2018) 12–18.

[34] X. Han, L. Sun, J. Zhao, Collective entity linking in web text: a graph-based method, in: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, Association for Computing Machinery, 2011, pp. 765–774.

[35] H. Huang, Y. Cao, X. Huang, H. Ji, C.-Y. Lin, Collective tweet wikification based on semi-supervised graph regularization, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2014, pp. 380–390.

[36] Z. Guo, D. Barbosa, Entity linking with a unified semantic representation, in: Proceedings of the 23rd International Conference on World Wide Web, Association for Computing Machinery, 2014, pp. 1305–1310.

[37] F. Kalloubi, O. E. Nfaoui, El Habib Beqaali, Microblog semantic context retrieval system based on linked open data and graph-based theory, Expert Systems with Applications 53 (C) (2016) 138–148.

[38] Y. Li, S. Tan, H. Sun, J. Han, D. Roth, X. Yan, Entity disambiguation with linkless knowledge bases, in: Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conference Steering Committee, 2016, pp. 1261–1270.

[39] O.-E. Ganea, M. Ganea, A. Lucchi, C. Eickhoff, T. Hofmann, Probabilistic bag-of-hyperlinks model for entity linking, in: Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conference Steering Committee, 2016, pp. 927–938.

[40] W.-H. Chong, E.-P. Lim, W. Cohen, Collective entity linking in tweets over space and time, in: European Conference on Information Retrieval, Springer, 2017, pp. 82–94.

[41] F. Wei, U. T. Nguyen, H. Jiang, Dual-fofe-net neural models for entity linking with pagerank, arXiv preprint arXiv:1907.12697 (2019).

[42] A. Parravicini, R. Patra, D. B. Bartolini, M. D. Santambrogio, Fast and accurate entity linking via graph embedding, in: Proceedings of the 2nd Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA), Association for Computing Machinery, 2019, p. 10.

[43] C. Liu, F. Li, X. Sun, H. Han, Attention-based joint entity linking with entity embedding, Information 10 (2) (2019) 46.

[44] Z. Fang, Y. Cao, Q. Li, D. Zhang, Z. Zhang, Y. Liu, Joint entity linking with deep reinforcement learning, in: The World Wide Web Conference, Association for Computing Machinery, 2019, pp. 438–447.

[45] C. B. El Vaigh, F. Goasdoué, G. Gravier, P. Sébillot, Using knowledge base semantics in context-aware entity linking, in: Proceedings of the ACM Symposium on Document Engineering 2019, 2019, pp. 1–10.

[46] M. C. Phan, A. Sun, Y. Tay, J. Han, C. Li, Pair-linking for collective entity disambiguation: Two could be better than all, IEEE Transactions on Knowledge and Data Engineering 31 (7) (2019) 1383–1396.

[47] X. Yang, X. Gu, S. Lin, S. Tang, Y. Zhuang, F. Wu, Z. Chen, G. Hu, 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, pp. 271–281.

[48] E. Rama-Maneiro, J. C. Vidal, M. Lama, Collective disambiguation in entity linking based on topic coherence in semantic graphs, Knowledge-Based Systems 199 (2020) 105967.

[49] D. Milne, I. H. Witten, Learning to link with wikipedia, in: Proceedings of the 17th ACM conference on Information and knowledge management, Association for Computing Machinery, 2008, pp. 509–518.

[50] J. B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem, Proceedings of the American Mathematical society 7 (1) (1956) 48–50.

[51] S. Zhang, H. Jiang, M. Xu, J. Hou, L. Dai, The fixed-size ordinally-forgetting encoding method for neural network language models, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2015, pp. 495–500.

[52] W. Fang, J. Zhang, D. Wang, Z. Chen, M. Li, Entity disambiguation by knowledge and text jointly embedding, in: Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning, Association for Computational Linguistics, 2016, pp. 260–269.

[53] I. Yamada, H. Shindo, H. Takeda, Y. Takefuji, Joint learning of the embedding of words and entities for named entity disambiguation, in: Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning, Association for Computational Linguistics, 2016, pp. 250–259.

[54] J. G. Moreno, R. Besançon, R. Beaumont, E. Dhondt, A.-L. Ligozat, S. Rosset, X. Tannier, B. Grau, Combining word and entity embeddings for entity linking, in: The Semantic Web, Springer International Publishing, 2017, pp. 337–352.

[55] P. Le, I. Titov, Improving entity linking by modeling latent relations between mentions, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, 2018, pp. 1595–1604.

[56] G. Zhu, C. A. Iglesias, Exploiting semantic similarity for named entity disambiguation in knowledge graphs, Expert Systems with Applications 101 (2018) 8–24.

[57] D. Mueller, G. Durrett, Effective use of context in noisy entity linking, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2018, pp. 1024–1029.

[58] S. Chen, J. Wang, F. Jiang, C.-Y. Lin, Improving entity linking by modeling latent entity type information, arXiv preprint arXiv:2001.01447 (2020).

[59] P. Le, I. Titov, Distant learning for entity linking with automatic noise detection, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2019, pp. 4081–4090.

[60] Ö. Sevgili, A. Panchenko, C. Biemann, Improving neural entity disambiguation with graph embeddings, in: Proceedings of the 57th Conference of the Association for Computational Linguistics: Student Research Workshop, Association for Computational Linguistics, 2019, pp. 315–322.

[61] W. Shi, S. Zhang, Z. Zhang, H. Cheng, J. X. Yu, Joint embedding in named entity linking on sentence level, arXiv preprint arXiv:2002.04936 (2020).

[62] I. L. Oliveira, D. Moussallem, L. P. F. Garcia, R. Fileto, Optic: A deep neural network approach for entity linking using word and knowledge embeddings, in: Proceedings of the 22th International Conference on Enterprise Information Systems, 2020, pp. 315–326.

[63] Z. Wang, J. Zhang, J. Feng, Z. Chen, Knowledge graph and text jointly embedding, in: Proceedings of the 2014 conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2014, pp. 1591–1601.

[64] H. Zhong, J. Zhang, Z. Wang, H. Wan, Z. Chen, Aligning knowledge and text embeddings by entity descriptions, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 267–272.

[65] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (2013).

[66] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, Curran Associates Inc., 2013, pp. 3111–3119.

[67] J. H. Friedman, Greedy function approximation: a gradient boosting machine, Annals of statistics 29 (5) (2001) 1189–1232.

[68] O. Levy, Y. Goldberg, Dependency-based word embeddings, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics, 2014, pp. 302–308.

[69] Y. Eshel, N. Cohen, K. Radinsky, S. Markovitch, I. Yamada, O. Levy, Named entity disambiguation for noisy text, in: Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017), Association for Computational Linguistics, 2017, pp. 58–68.

[70] J. Devlin, M.-W. Chang, K. Lee, 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, Volume 1 (Long and Short Papers), Association for Computational Linguistics, 2019, pp. 4171–4186.

[71] O.-E. Ganea, T. Hofmann, Deep joint entity disambiguation with local neural attention, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2017, pp. 2619–2629.

[72] Q. Le, T. Mikolov, Distributed representations of sentences and documents, in: Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32, JMLR.org, 2014, pp. 1188–1196.

[73] B. Perozzi, R. Al-Rfou, S. Skiena, Deepwalk: Online learning of social representations, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, 2014, pp. 701–710.

[74] T. G. Dietterich, R. H. Lathrop, T. Lozano-Pérez, Solving the multiple instance problem with axis-parallel rectangles, Artificial intelligence 89 (1-2) (1997) 31–71.

[75] S. Riedel, L. Yao, A. McCallum, Modeling relations and their mentions without labeled text, in: Joint European Conference on Machine Learning

and Knowledge Discovery in Databases, Springer Berlin Heidelberg, 2010, pp. 148–163.

[76] Y. Fu, Y. Ma, Graph embedding for pattern analysis, Springer Science & Business Media, 2012.

[77] S. Ruder, I. Vulić, A. Søgaard, A survey of cross-lingual word embedding models, Journal of Artificial Intelligence Research 65 (1) (2019).

[78] P. Cui, X. Wang, J. Pei, W. Zhu, A survey on network embedding, IEEE Transactions on Knowledge & Data Engineering PP (01) (2018) 1–1.

[79] P. Goyal, E. Ferrara, Graph embedding techniques, applications, and performance: A survey, Knowledge-Based Systems 151 (2018) 78–94.

[80] R. Usbeck, M. Röder, A.-C. Ngonga Ngomo, C. Baron, A. Both, M. Brümmer, D. Ceccarelli, M. Cornolti, D. Cherix, B. Eickmann, et al., Gerbil: general entity annotator benchmarking framework, in: Proceedings of the 24th international conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2015, pp. 1133–1143.

[81] Z. Guo, D. Barbosa, Robust named entity disambiguation with random walks, Semantic Web 9 (4) (2018) 459–479.

[82] E. Meij, W. Weerkamp, M. De Rijke, Adding semantics to microblog posts, Wsdm 2012 (2012) 563.
URL http://dl.acm.org/citation.cfm?doid=2124295.2124364

[83] Y. Li, C. Wang, F. Han, J. Han, D. Roth, X. Yan, Mining evidences for named entity disambiguation, in: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 2013, pp. 1070–1078.

[84] M. Pershina, Y. He, R. Grishman, Personalized page rank for named entity disambiguation, in: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015, pp. 238–243.

[85] M. Nickel, K. Murphy, V. Tresp, E. Gabrilovich, A review of relational machine learning for knowledge graphs, Proceedings of the IEEE 104 (1) (2016) 11–33.

[86] Q. Wang, Z. Mao, B. Wang, L. Guo, Knowledge graph embedding: A survey of approaches and applications, IEEE Transactions on Knowledge and Data Engineering 29 (12) (2017) 2724–2743.

[87] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, D. Sontag, Learning a health knowledge graph from electronic medical records, Scientific reports 7 (1) (2017) 5994.

[88] A. Grover, J. Leskovec, node2vec: Scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, Association for Computing Machinery, 2016, pp. 855–864.

[89] A. Joulin, E. Grave, P. Bojanowski, M. Nickel, T. Mikolov, Fast linear model for knowledge graph embeddings, arXiv preprint arXiv:1710.10881 (2017).

[90] F. Abel, Q. Gao, G.-J. Houben, K. Tao, Analyzing user modeling on twitter for personalized news recommendations, in: User Modeling, Adaption and Personalization, Springer Berlin Heidelberg, 2011, pp. 1–12.

[91] F. Abel, Q. Gao, G.-J. Houben, K. Tao, Semantic enrichment of twitter posts for user profile construction on the social web, in: The Semantic Web: Research and Applications, Springer, Berlin, Heidelberg, 2011, pp. 375–389.