Using Knowledge Base Semantics in Context-Aware Entity Linking

Cheikh Brahim El Vaigh, François Goasdoué, Guillaume Gravier, Pascale Sébillot

Using Knowledge Base Semantics in Context-Aware Entity Linking

Cheikh Brahim El Vaigh
Inria, IRISA
Rennes, France
cheikh-brahim.el-vaigh@inria.fr

François Goasdoué
Univ. Rennes 1, IRISA
Lannion, France
fg@irisa.fr

Guillaume Gravier
CNRS, IRISA
Rennes, France
guig@irisa.fr

Pascale Sébillot
INSA Rennes, IRISA
Rennes, France
pascale.sebillot@irisa.fr

ABSTRACT

Entity linking is a core task in textual document processing, which consists in identifying the entities of a knowledge base (KB) that are mentioned in a text. Approaches in the literature consider either independent linking of individual mentions or collective linking of all mentions. Regardless of this distinction, most approaches rely on the Wikipedia encyclopedic KB in order to improve the linking quality, by exploiting its entity descriptions (web pages) or its entity interconnections (hyperlink graph of web pages). In this paper, we devise a novel collective linking technique which departs from most approaches in the literature by relying on a structured RDF KB. This allows exploiting the semantics of the interrelationships that candidate entities may have at disambiguation time rather than relying on raw structural approximation based on Wikipedia's hyperlink graph. The few approaches that also use an RDF KB simply rely on the existence of a relation between the candidate entities to which mentions may be linked. Instead, we weight such relations based on the RDF KB structure and propose an efficient decoding strategy for collective linking. Experiments on standard benchmarks show significant improvement over the state of the art.

KEYWORDS

Entity linking, collective entity linking, knowledge base semantics, entity relatedness measure, RDF

1 INTRODUCTION

Entity linking is a key natural language processing component for textual document engineering, with applications such as semantic search [2] and information extraction [18–20]. Entity linking consists in identifying the entities of a reference knowledge base (KB) that are mentioned in textual documents, thus linking documents to semantic entities. This is for instance a crucial step in the context of the iCODA project[1] that we are pursuing, where we seek to apply entity linking to a large archive of a regional newspaper using a KB

[1] http://project.inria.fr/icoda

of specific places, companies, public bodies and people. Linking entities mentioned in the archive to a reference KB enables semantic search and analytics within the archive and facilitates information extraction to complement the KB.

Entity linking is however a non-trivial task because entity mentions are usually ambiguous. For example, given the sentence "Steve Jobs is a co-founder of Apple.", the mention "Steve Jobs" refers to distinct entities within the Freebase KB [4]: a person, a book, a film, etc. Conversely, an entity may be mentioned in various forms where, for instance, the Freebase entity "Steve Jobs" may appear in a document as "Steve Jobs", "Steve Paul Jobs", "Steve P. Jobs", "Steve" and "Jobs", etc.

Automatic approaches for entity linking usually implement three well-established steps [34]. Named entity recognition is first performed to identify the entity mentions in a document. Candidate entities from the KB under consideration are then generated for each mention found. Finally, every mention is linked to one of its candidate entities in a so-called linking step. For this last step, two types of approaches can be found in the literature depending on whether the linking is performed *independently* for each individual mention, e.g., [12, 15, 26], or *collectively* for all mentions at once, e.g., [6, 13, 16, 21, 24, 32, 42]. In the first case, which we call *entity-by-entity linking*, every mention in a text is assumed to be independent from other mentions and is linked to a candidate entity on the sole base of some similarity between the mention and the candidate entities, so-called *local* scores. By contrast, for *collective entity linking*, entity mentions and the corresponding entities are not assumed independent one from another but somehow semantically related within a (coherent) document, i.e., mention-to-entity linking decisions are interdependent. In this case, the local mention-entity scores are complemented with *global* scores reflecting to which extent the candidate entities chosen for the mentions under consideration are related in the KB, according to a so-called entity *relatedness measure*.

Entity linking techniques usually make use of the information within the KB—e.g., entity names and descriptions whenever they are available—and of its structure. A large body of work use Wikipedia, e.g., [1, 5, 8, 13, 16, 22, 32], exploiting its entity descriptions (namely Wikipedia's web pages) and interconnections. The latter are implicitly captured by means of the hyperlink graph of Wikipedia

web pages. There is however only loose semantics behind these interconnections, which prevents from fully taking advantage of KBs reasoning capabilities to better exploit the relationship between entities. To skirt this issue, a few approaches use RDF structured KBs, such as BaseKB, DBpedia or YAGO, instead of Wikipedia [17, 29, 41]. They thus theoretically benefit from the use of the formal RDF data model for knowledge representation, in particular to exploit the precise semantics of the KB entities (e.g., types) and of their interrelationships (e.g., names and cardinalities), while Wikipedia can just inform about the existence of some unspecified relation between two entities through its hyperlink graph. Approaches that use the RDF KB structure in the context of collective entity linking however do not exploit the semantics of the relations within the KB, limiting themselves in the best case to a binary indicator of whether a relation exists or not between two entities [23]. They moreover use costly algorithms for collective disambiguation, which makes them poorly suitable for large scale document engineering.

In this paper, we address these two limits of collective entity linking—limited use of the KB semantics and computationally expensive collective disambiguation algorithms—with the following contributions:

(1) We devise a novel fine-grained *weighted semantic relatedness measure* (WSRM), which we design to better capture the description of the semantic interrelations between entities that are available in structured RDF KBs. In particular, this measure relies on the number of interrelations that pairs of candidate entities have in the RDF KB at hand, while KB-based measures in the literature just consider the existence of some relation between them.

(2) We provide a *lightweight* collective linking technique on top of the proposed relatedness measure. Notably, in contrast with most previous techniques in the literature, it avoids the manipulation of the so-called *mention-entity graph*, which constitute a known bottleneck of collective linking.

(3) We experimentally demonstrate *significant improvement over the state of the art* on standard benchmarks in collective entity linking, thoroughly benchmarking our approach against popular state-of-the-art methods available on the GERBIL platform [37] and comparing to results reported in the literature.

The paper is organized as follows. Sec. 2 discusses the state of the art, introducing concepts and notations. Sec. 3 introduces the basic notions required for the description of our method in Sec. 4. Experimental results are grouped in Sec. 5 before discussing the perspectives opened by the weighted semantic relatedness measure in Sec. 6.

2 RELATED WORK

Entity linking (EL) has been widely investigated in the literature, as reported in the recent survey [34]. In this section, we mostly restrict the discussion to the ranking of candidate entities for mentions, since this is the step that mainly concerns our contributions. We focus on entity-by-entity linking in Sec. 2.1 before considering collective linking in Sec. 2.2. Limits of these approaches are discussed in Sec. 2.3.

2.1 Entity-By-Entity Linking

Assuming the mentions in a text to be independent one from another, entity-by-entity methods are based on the similarity between some entity mention and its candidate entities in a KB. Basically, they search the candidate entity that maximizes a so-called *local* similarity measure between a mention and its candidate entities. In Fig. 1 which depicts mentions and candidate entities within a document, only the relationship between a mention and its entity are considered.

This can be formalized as

$$\hat{e} = \arg\max_{e_i} \phi(m, e_i) \tag{1}$$

where e_i is a candidate entity, m is an entity mention, and ϕ is the local score function.

Entity-by-entity linking is studied in several publications [8–10, 12, 15, 26]. Words and entities are embedded in vectors in [8, 9, 26] and a cosine similarity is used to compute the local score, while [10] and [12] use *Wikipedia popularity*. This popularity, a.k.a. *commonness*,

$$\text{pop}(m, e) = \frac{n(m, e)}{\sum_{e' \in W} n(m, e')} \tag{2}$$

can be defined as the probability that a mention m is used as the text (anchor) of a hyperlink referring to an entity e, with W the set of all Wikipedia pages and $n(m, e)$ the number of times m occurs as an anchor for e in some Wikipedia page. [12] further combines popularity with a convolution neural network (CNN) to extract topic vectors from both the context of the mention (words surrounding it) and the context of an entity (its Wikipedia page description). Finally [15] jointly embeds entities and mentions in the same space leveraging mention context, entities description and fine-grained types (Person, Politician, Governor, Organization, Location, Country, City, etc.).

2.2 Collective Entity Linking

Collective entity linking (CEL) improves on entity-by-entity linking by not only considering the similarity between an individual mention and its candidate entities, but also taking into account the intricate interrelationships that candidate entities of the different mentions may have. To this aim, the local score function $\phi()$ is complemented with a so-called *global* score function $\psi()$ reflecting *entity relatedness*, i.e., affinity between entities. The CEL problem can be thus formalized as

$$(\hat{e}_1, ..., \hat{e}_n) = \arg\max_{e_{1,...,n}} \left(\sum_{i=1}^{n} \phi(m_i, e_i) + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \psi(e_i, e_j) \right) \tag{3}$$

where n is the total number of mentions in a text.

Popular entity relatedness measures in the literature are the cosine similarity using entity embeddings, the Wikipedia hyperlink-based measure (WLM) [28]

$$WLM(e_i, e_j) = 1 - \frac{\log\left(\frac{\max(|IN_{e_i}|, |IN_{e_j}|)}{|IN_{e_i} \cap IN_{e_j}|}\right)}{\log\left(\frac{|W|}{\min(|IN_{e_i}|, |IN_{e_j}|)}\right)} , \tag{4}$$
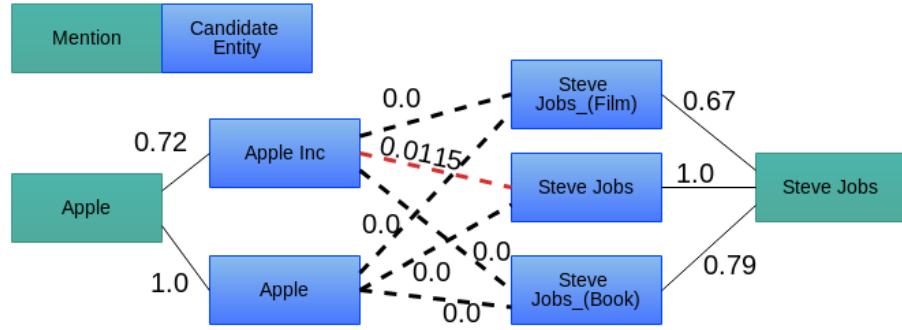
Figure 1: Illustration of a mention-entity graph within a document: the two mentions (Apple and Steve Jobs) are linked to candidate entities with some local score (solid lines) while the entity relatedness appears as weighted dotted lines. In this toy example, the strong relation between the entities Steve Jobs and Apple Inc helps in jointly selecting those two.

the Normalized Jaccard similarity (NJS) [32]

$$\text{NJS}(e_i, e_j) = \frac{\log\left(|\text{IN}_{e_i}| \cap |\text{IN}_{e_j}| + 1\right)}{\log\left(|\text{IN}_{e_i} \cup \text{IN}_{e_j}| + 1\right)}, \tag{5}$$

and the reference binary indicator (Ref) [1, 23]

$$\text{Ref}(e_i, e_j) = \begin{cases} 1 & \exists r, (e_i, r, e_j) \in KB \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

In these equations, IN_{e_i} is the set of incoming hyperlinks in the Wikipedia page for e_i, and r is a relationship between e_i and e_j in a broad sense (a hyperlink for the Wikipedia KB or a relation for an RDF KB). For example, *WLM* is used in [16, 23, 28], *Ref* in [1, 23], cosine similarity between entities in [24, 25, 29, 42], while a combination of the three is used in [32].

Beyond the entity relatedness measure, collective linking techniques also differ on their linking strategy, either relying on the direct optimization of Eq. 3 [21, 23, 24, 29, 42] or on a *mention-entity graph* [1, 6, 13, 16, 22, 25, 28, 32, 41] such as the one illustrated in Fig. 1.

Optimization-based techniques aim at computing the solution to Eq. 3, which is known to be NP-hard. They therefore compute an approximation of the solution. [29, 42] jointly embed words and entities in the same space and train a classifier with local and global scores to select the best solution. [24] reduces the problem to a sub-matrix search and solves the optimization of Eq. 3 using a gradient descent. [23] first selects for each mention, the best candidate entity based on an entity type scheme like Eq. 2, then solves the optimization of Eq. 3 by using only the previous selected candidate entities in the computation of global scores. Finally, [21] advocates for models that jointly solve named entity recognition and entity linking using only words and characters embedding [14].

Mention-entity graph-based techniques build a graph whose nodes are entity mentions and candidate entities, and whose edges connect either some mention to one of its candidate entities or candidate entities of different mentions. Edges are weighted, in the first case with the local score, in the last case with the global score [16, 25]. A slightly different graph is used in [1]: nodes denote (mention,
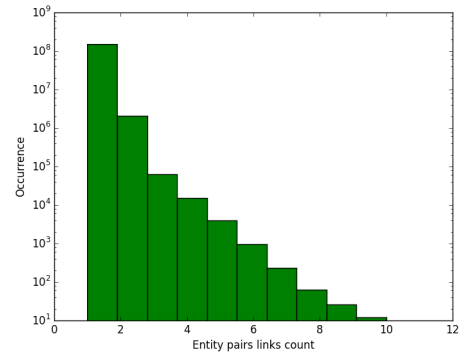

Figure 2: Distribution of entity pairs (Y-axis) per number of relations between entities (X-axis) in the BaseKB RDF KB.

candidate entity) pairs and edges connect nodes with similar candidate entities, weighted with a global score. Also, [41] constructs a graph leveraging mentions coreferences and tries to match it with a candidate entity graph. [6] used a graph convolution network (GCN) to perform the linking using local and global scores. [13, 22] use a factor graph based on *loopy belief propagation* (LBP). Lastly [32] proposes to solve the CEL with analogy to the minimum spanning tree problem using cosine similarity, WLM and NJS. Graph-based methods are in general very accurate, at the price of the computationally expensive construction of mention-entity graphs.

2.3 Limits of Entity Linking Techniques

In this paper, we argue that collective entity linking, which already improves on entity-by-entity linking, can be further improved by capitalizing on the precise semantics of the relationships between entities that a structured RDF KB provides. In particular, entity relatedness measures used so far in CEL techniques either build on the loose semantics of connections between entities in the Wikipedia hyperlink graph (e.g., WLM, NJS or Ref), which only indicates that connected entities are *somehow* related, or simply on the *existence*

of some relation between entities in an RDF KB (e.g., Ref). When a RDF KB is used, we advocate that going beyond the simple existence of some relation between two entities, by taking into account by how many relations these entities are actually connected, is beneficial. To give an insight on the valuable extra information this might yield, Fig. 2 shows the distribution of entity pairs per number of relations between entities in the BaseKB RDF KB. For instance, there are approximately 10^8 pairs of entities connected by a single relation, and about 10 pairs connected by 10 relations. It is worth noting that all these pairs of entities are indistinguishable if we only consider the existence of a relation between them as *Ref* does. We incorporate this information into our relatedness measure (Sec. 4.3) to model that the more connected entities are in the KB, the more likely they are mentioned together in some text document. We then follow a lightweight supervised CEL technique (Sec. 4.4), whose global scores use this measure, and which does not require constructing the mention-entity graph.

3 PRELIMINARIES

We briefly introduce here the two central notions on which builds our collective entity linking technique: the *RDF standard* from the data management and knowledge representation & reasoning communities, which is the data model of the KBs we consider, and *word embeddings* which we use to compare entity mentions and candidate entities.

3.1 RDF Knowledge Bases

The resource description framework (RDF) is a popular graph data model by W3C [39] used to express both data and domain knowledge in a structured KB. RDF is the cornerstone of semantic web applications, whose emblematic incarnation is the linked open data cloud[2].

Formally, an RDF KB is a set of (s, p, o) *triples* [38], each describing its *subject s* with the *property p* whose value—called *object*—is *o*. An RDF KB can be seen as a graph made of $s \xrightarrow{p} o$ edges. Triples are used to state facts and domain knowledge using *special properties* from the RDF standard, as shown in Table 1 and exemplified next.

A fact is either a *class assertion* stating that some identifier has some *type*, or a *property assertion* stating a *relation* between two identifiers or between some identifier and some constant. For instance, a KB may state that i_1 is a person, whose name is Steve Jobs and who is employed by i_2 with the following triples:

$(i_1, \text{type}, \text{Person}), (i_1, \text{name}, \text{"Steve Jobs"}), (i_1, \text{emplBy}, i_2).$

On the other hand, domain knowledge is expressed by establishing so-called *ontological constraints* between the classes (types) and properties (relations) allowed to state facts. Such constraints are *subclass and subproperty relationships*, a.k.a. ISA constraints, as well *typing constraints for property attributes*: the first is called *domain* while the second is called *range*. For example, the above KB may additionally state that employees are persons, only employees are employed, only organizations are employers and that being

---
[2]https://lod-cloud.net

RDF fact	Triple notation
Class assertion	(s, type, o)
Property assertion	(s, p, o) with $p \notin \{\text{type}, \text{subClassOf}, \text{subPropertyOf}, \text{domain}, \text{range}\}$

RDF knowledge	Triple notation
Subclass	$(s, \text{subClassOf}, o)$
Subproperty	$(s, \text{subPropertyOf}, o)$
Domain typing	(s, domain, o)
Range typing	(s, range, o)

Table 1: RDF triples for facts and knowledge.

employed by is a particular case of working for:

(Employee, subClassOf, Person), (emplBy, domain, Employee), (emplBy, range, Organization), (emplBy, subPropertyOf, workFor).

Importantly, RDF KBs model both explicit and implicit data and knowledge triples. The explicit ones are those stored in the KB, e.g., the triples of the above example, while the implicit ones are those that can be derived from the explicit ones and *entailment rules* from the RDF standard [40]. For instance, above, from the fact i_1 *is employed by* i_2 and the constraint *only employees are employed*, the implicit fact i_1 *is an employee* is inferred, i.e., the implicit triple $(i_1, \text{type}, \text{Employee})$ holds in the KB although it is not explicitly stored there. Also, from the two constraints *only employees are employed* and *employees are persons*, the implicit constraint *only person are employed* is inferred, i.e., the implicit triple (emplBy, domain, Person) also holds in the KB.

In this paper, we exploit the *fine-grained semantic description of entities* that RDF KBs make available in terms of their types— e.g., the entity i_1 representing Steve Jobs is known as a Person and implicitly as an Employee in the KB—and of their relationships— e.g., i_1 is known to be employed by i_2 and implicitly to work for i_2. Crucially, to fully take advantage of the explicit *and* implicit triples of a KB, in particular for the collective entity linking technique proposed in this paper, we assume that the KB is *saturated*, i.e., *all* its implicit triples have been computed and explicitly added to the KB. KB saturation, a.k.a. *closure*, is a reasoning step defined in the RDF standard [40] that many RDF data management tools implement.

3.2 Word Embeddings

Word embeddings allow encoding words as points in a multidimensional space. There exist several embedding models like [3, 27, 31]; their basic intention is that semantically similar words in a text end up in near coordinates within the target multidimensional space. Thus the cosine similarity between two points in the multidimensional space at hand reflects their semantic proximity in text. In this work, we consider the skip-gram embedding model [27] which relies on the famous formula of Firth [11]: "You should know a word by the company it keeps". For skip-gram, word embeddings are learned to predict words that appear in their context, i.e., surrounding words, based on dot products between vector representation of words. More formally, given a large training corpus represented as a sequence of words $w_1, ..., w_N$, the objective of the skip-gram model is to

maximize

$$\sum_{t=1}^{N} \sum_{c \in C_t} \log P(w_c|w_t) \tag{7}$$

where the context C_t is the set of indices of words surrounding word w_t. A usual choice to define the probability of a context word using dot product is the softmax function

$$P(w_c|w_t) = \frac{\exp(v_c^T v_t)}{\sum_{i=1}^{N} \exp(v_c^T v_t)} \quad , \tag{8}$$

where v_i denotes the vector representation of word w_i. These vectors can embed efficiently words semantics, especially when the model is trained using a huge dataset like Wikipedia or Google News.

In this work, we classically use skip-gram-based word embeddings and cosine similarity as ingredients of our local score calculation to link entity mentions to candidate entities.

4 COLLECTIVE LINKING WITH KB SEMANTICS

Our main contribution to the collective linking task lies in the definition of a novel entity relatedness measure, designated as weighted semantic relatedness measure (WSRM), that takes advantage of the KB structure in the candidate selection stage of the standard EL pipeline. This measure is combined with local mention-entity similarity measures in a classification step whose goal is to predict whether a link should be made or not between a mention and a candidate entity. The classifier operates independently on each mention-entity pair, however taking as input features depicting the relatedness of the entity to candidates of other mentions, in addition to the local score. Taking into account the context in which a mention-entity pair appears enables a form of collective linking at a low computational cost, replacing global optimization with an ensemble of local classifications.

Following many previous studies, e.g., [6, 13, 29, 32, 42], we disregard the named entity mention detection stage, and assume that we know for each document the mentions and their type (person, organization, location, etc.). For the candidate generation stage, a standard architecture is chosen as in [34], relying on Wikipedia for practical reasons. The candidate selection stage accounts for the WSRM between candidate entities within the document in an efficient manner, relying here on BaseKB, one of the various RDF KB providing structured knowledge about Wikipedia entities. Note that while BaseKB is used here, we are conceptually not limited to this KB as the WSRM can account for any RDF KB. The choice of BaseKB was mostly driven by the fact that various baselines use the very same KB and that candidate generation relies at this stage on Wikipedia entities, thus requiring a mapping of the reference KB to these entities.

In the remainder of this section, we represent a document D by its set of entity mentions, $D = (m_1, ..., m_n)$ where n is the number of mentions. For each mention m_i, $C(m_i) = (e_{i1}, ..., e_{ik})$ denotes the set of its candidate entities. The presentation follows the logical steps of the whole process, namely candidate generation, local score definition, weighted semantic relatedness measure and the

classification-based actual linking, where the two last steps constitute the core of our contribution.

4.1 Candidate Entities Generation

To generate candidate entities from the KB for each mention in a document, we chose a simple yet efficient method exploiting Cross-Wiki [35]. Cross-Wiki is a dictionary computed from a Google crawl of the web that stores the frequency with which a mention links to a particular entity in Wikipedia. We used the same Cross-Wiki dictionary as in [15][3]. Each entry of the dictionary corresponds to a possible entity mention and provides a list of Wikipedia pages to which the mention points to along with popularity scores.

This list is directly used for candidate generation whenever a mention appears in the dictionary. The dictionary entries are normalized by removing all punctuation marks and converting to lowercases. For example, the generation of candidate entities for "Steve Jobs" with Cross-Wiki leads to

cross-wiki["stevejobs"] => [['7412236', 0.99],['5042765',0.01]]
where '7412236' is the id of the Wikipedia page "Steve Jobs", with a score of 0.99, and '5042765' is the id of the Wikipedia page "God" (sic!). For mentions absent from cross-wiki, we perform a request on Wikipedia using the text of the mention, and collect the resulting Wikipedia pages as the candidate entities.

4.2 Local Mention-Entity Score

The local relevance score $\phi(m_j, e_{ij})$ between a mention m_i and a candidate entity e_{ij} can be established between the mention and the candidate entity name only (name of the corresponding Wikipedia page), but can also consider the context (e.g., the surrounding words of the mention, and the description of the entity respectively). In this work, we consider two local score functions, namely the cosine similarity between m_i and the Wikipedia title of e_{ij} in an embedded space, and the popularity as defined in Eq. 2. The former reflects the geometric (and thus semantic) proximity in the embedded space of one mention and one entity. The latter is known to be a good estimation of the similarity between a mention and an entity.

4.3 Entities Relatedness

As highlighted in Sec. 2.2, a wide range of entity relatedness measures (a.k.a. coherence) has been defined and used in the literature to establish a score that reflects the relationship between two entities in a KB. But only a few reflect a real semantic similarity measure as defined by the precise interrelationships entities have in an RDF KB. One notable exception is the $\text{Ref}(e_i, e_j)$ measure of Eq. 6 which solely provides a binary relatedness measure.

To take full advantage of the KB semantics, we introduce a weighted semantic relatedness measure (WSRM) based on the total number of relations two entities share in the KB. The key idea in this measure is to not only express the existence of relations in the KB between the entities at hand, but also weight the relation between entities, where the more relations between the entities, the stronger their relationship. Formally, we define the relatedness between two

---

[3]https://drive.google.com/uc?id=0Bz-t37BfgoTuSEtXOTI1SEF3VnM&export=download

entities e_i and e_j as

$$\text{WSRM}(e_i, e_j) = \frac{n(e_i, e_j)}{\sum_{e' \in E} n(e_i, e')} \qquad (9)$$

where $n(e_i, e_j)$ is the total number of property assertions in the KB (see Tab. 1) that have e_i as subject and e_j as object, and E is the set of entities in the KB. We assume the KB to be saturated (all implicit triples are made explicit) and we disregard triples corresponding to class assertions and RDF knowledge (e.g., type, subPropertyOf, domain, range, sameAs).

Because the directions of the relations are somewhat arbitrary in KBs, depending on how the relation vocabulary was designed (e.g., think about the *publishes* and *publishedBy* symmetric RDF properties), we use a symmetric version of WSRM defined as

$$\psi(e_i, e_j) = \frac{1}{2} \left(\text{WSRM}(e_i, e_j) + \text{WSRM}(e_j, e_i) \right) . \qquad (10)$$

Conceptually, $\text{WSRM}(e_i, e_j)$ is similar to popularity however applied to two entities rather than to a mention and an entity: it exploits the number of relations two entities e_i and e_j share in the KB. As we normalize over the KB, this measure gives a probability that e_i is related to e_j: a value of 1 means e_i is equal to e_j, while a value of 0 means the absence of connection in the KB. The higher the number of relations between two entities, the higher their probability to be mentioned together in a text. Thus selecting candidate entities that have a large number of relations in the KB is intuitively bound to maximize the CEL objective function of Eq. 3. Note finally that while Eq. 10 only considers direct relations between the two entities, i.e., there is an RDF triple linking e_i and e_j one way or another, it opens the door to many extensions featuring semantic reasoning to assess the relation between the two entities. For instance, a naive extension could consist in counting also the number of paths of up to some length n between e_i and e_j or e_j and e_i in the KB (recall an RDF KB is a graph database) instead of limiting ourselves to paths of length 1.

4.4 Supervised Collective Entity Linking

The last step is to decide which candidate should be retained for each mention within the document. To do so in collective linking and exploit the WSRM entity relatedness measure, any of the solutions mentioned in Sec. 2.2 could be used, exploiting the mention-entity graph one way or another. These approaches are however computationally heavy and their cost grows rapidly as the number of mentions and candidates increases. We thus adopt a supervised approach similar to [29, 42], where a binary classifier is trained to predict whether a mention and a candidate entity are related (1) or not (0). However, in the collective linking setting, the classifier relies on features related on the mention-entity pair as well as on contextual (global) features accounting for the relatedness of the entity with candidates from other entity mentions in the document.

For practical reasons, it is better to have contextual features of fixed size, which for a candidate entity e_{ij} requires aggregating the entity relatedness scores $\psi(e_{ij}, e)$ for all $e \in C(m_l), l \neq i$. The alternative would be to consider all scores up to a maximum, zeroing non existing scores, but aggregation appears much more simple and is experimentally shown efficient (see Sec. 5). We consider a

conjunction of simple aggregators such as the sum, e.g.,

$$\text{sum}(e_{ij}; m_1, .., m_n) = \sum_{l=1, k \neq i}^{n} \sum_{e \in C(m_l)} \psi(e_{ij}, e) , \qquad (11)$$

mimicking the global term in Eq. 3, or the k maximum values

$$\max_k(e_{ij}; m_1, .., m_n) = \max_{l=1, j \neq i}^{n} @\text{k} \max_{e' \in C(m_l)} \psi(e_i, e) \qquad (12)$$

where max @k is the kth highest value. Note that combining aggregators is beneficial: on the one hand, using the sole sum can indeed introduce noise since not all candidate entities are relevant for the linking; on the other hand, leveraging the sole max can be very drastic because only candidate entities with maximum relatedness score are kept. But, as evidenced in [32], entities mentioned in a document are not necessarily the most connected in the KB. Hence we retained the sum, max @1,max @2 and max @3 as global contextual features, which can be seen as a kind of flexibility in selecting and aggregating the relatedness scores.

Finally, a binary logistic regression classifier is trained to predict whether a mention and an entity match or not, taking as input features the two mention-entity scores (cosine similarity and popularity) and the 4 global contextual features. At linking time, this classifier is used independently for all mentions m_i, considering all pairs (m_i, e) with $e \in C(m_i)$ and retaining the best one as the entity corresponding to m_i, i.e., e_{ij} with $j = \arg\max_k \text{logreg}(m_i, e_{ik})$ where logreg() is the binary logistic regression classifier.

5 EXPERIMENTS

A thorough experimental evaluation of our contributions on standard benchmarks was conducted in order to prove the relevance of both our novel entity relatedness measure WSRM, and our lightweight collective linking technique that builds on it. We describe our experimental setup in Sec. 5.1. Then, Sec. 5.2 presents a first experiment that aims to show the intrinsic quality of our relatedness measure, i.e., its greater ability to select the most appropriate entity for each mention w.r.t. other measures that do not exploit an RDF KB. Sec. 5.3 proposes an ablation study of our method in order to demonstrate the impact of local and global scores. Finally Sec. 5.4 compares our CEL approach to state-of-the-art entity-by-entity and CEL competitors, showing significant improvement.

5.1 Experimental Setup

5.1.1 Knowledge Base. Entity linking requires the use of a KB. It can be an encyclopedic KB like Wikipedia or the RDF KBs BaseKB, DBpedia, YAGO, etc., or any ad hoc KB constructed for a particular need as would ultimately be the case of our regional newspaper archive. In this paper, we make use of Wikipedia in the candidate generation stage and of BaseKB[4], an RDF knowledge base derived from Freebase which contains over one billion facts (i.e., triples) about more than 40 millions subjects, for semantic relatedness measure. As previously mentioned, the crucial interest of such an RDF KB over Wikipedia resides in the fact that both its entities and their interrelationships bear a precise semantics.

A mapping between Wikipedia and BaseKB entities is used to enable taking advantage of Wikipedia in the early stages of the

---

[4]http://basekb.com/

Dataset	Nb. docs	Nb. mentions	Avg nb. mentions/doc
TAC-KBP 2016 eval	169	9231	54.6
TAC-KBP 2017 eval	167	6915	41.4
AIDA-train	846	18519	21.9
AIDA-valid	216	4784	22.1
AIDA-test	231	4479	19.4
Reuters128	128	881	6.9
RSS-500	500	1000	2

Table 2: Statistics on the datasets used.

process, in particular the fact that names in Wikipedia are meaningful unique identifiers, and of BaseKB semantics in the linking stages. To ensure consistency, we discarded from BaseKB entities with no corresponding Wikipedia entry, resulting in approx. 4M entities in the RDF KB.

5.1.2 Datasets. Experimental results are reported on four standard datasets with different characteristics, including both short and long documents as well as formal (news) and informal (forum) texts:

- CoNLL-AIDA is an entity annotated corpus of Reuters news documents introduced by Hoffart et al. [16]. It is much larger than most of the other existing EL datasets, making it an excellent evaluation target. Data is divided into three parts: Train, AIDA-A (used for validation) and AIDA-B (used for evaluation). The original target KB for CoNLL-AIDA was YAGO but a recent update allows linking to BaseKB.
- TAC-KBP Entity Discovery and Linking (EDL) 2016-2017 datasets are newswire and forum-discussion documents originally collected for the TAC Knowledge Base Population Entity Discovery and Linking 2016 and 2017 international evaluation campaigns [20]. We only used the gold-standard where entity mentions are already annotated.
- Reuters128 [33] is a small dataset that contains 128 economic news articles taken from the Reuters-21587 corpus.
- RSS500 [33] contains 500 documents created from RSS feeds including all major worldwide newspapers and a wide range of topics, e.g., World, U.S., Business, Science, etc.

Table 2 gathers key figures and statistics for each of the datasets. For experiments on the Reuters128 and RSS-500 corpora, training of the classifier is performed on the AIDA train dataset due to the small size of these two datasets.

5.1.3 Practical details. Word embedding of dimension 100 trained on a dump of Wikipedia from October 2018 are used for all the experiments. Only words appearing at least 5 times are retained in the embedding.

At classification time, for candidate entities generated with Cross-Wiki that are absent from Wikipedia, all the features were set to 0 as, apart from the popularity provided by Cross-Wiki, features for out-of-KB entities cannot be computed. We also limit classification to entity mentions that have a corresponding entity in the KB: in other words, we disregard mentions for which the ground truth points to entities not present in BaseKB. Note also that for efficiency reasons, the WSRM entity relatedness measure was pre-computed for all pairs of entities in BaseKB, thus limiting the computational cost at linking time.

	nDCG@1	nDCG@5	nDCG@10
WLM [17]	0.54	0.52	0.55
Cosine similarity [42]	0.59	0.56	0.59
WSRM	0.69	0.64	0.58

Table 3: Intrinsic performance of entity relatedness measure.

5.2 Intrinsic Quality of the Entity Relatedness Measure

A first experiment was conducted to evaluate the intrinsic quality of our semantic entity relatedness measure alone, that is, its ability to select the best candidate among the candidate entities of a mention, both the mention and the corresponding candidate entities being given. To this end, we use the same test set as in [7], specifically created from a subset of the CoNLL-AIDA dataset to evaluate the relevance of relatedness measures. This test set contains 3,314 mentions and their corresponding candidate entities, with 91 candidate entities per mention in average, among which the correct entity is known. Leveraging the weighted semantic relatedness measure of Eq. 10, a ranked order of the candidate entities was obtained for each mention. The ranking is evaluated with the normalized discounted cumulative gain (nDCG) score to evaluate the quality of the entity relatedness measure. Table 3 presents our results together with those reported in the literature with the same experimental setup for WLM [17] and for cosine similarity using joint word and entity embeddings [42]. The WSRM clearly outperforms the WLM and is better than or comparable to the cosine similarity of [42]. This indicates the higher relevance of semantic relatedness measures than that of Wikipedia-link overlap (WLM). Joint embeddings are good at incorporating relatedness between a mention or entity, and a list of candidate entities, but their performance decrease when the list of candidate entities becomes shorter, where our method still performs very well—see columns *nDCG*@10 where the 10 first ranked entities are considered versus *nDCG*@1 where only the best ranking entity is considered.

5.3 Ablation Study

As highlighted in Sec. 2, CEL is supposed to improve over entity-by-entity linking by taking into account the interrelationships between the candidate entities of the different mentions. Considering the two mention-entity scores and the four contextual features respectively retained and proposed in Sec. 4.4, the ablation study described here aims at demonstrating the benefit of considering together local and global scores in our CEL technique, also attesting again the relevance of our weighted entity relatedness measure.

Using TAC-KBP 2016 as training data, several versions of our classifier were thus trained, which consider either local scores only (i.e., an entity-by-entity linking version of our approach) or also some or all the global features that we proposed. Table 4 presents the performance of those different versions on TAC-KBP 2017 in terms of linking accuracy.[5]

---

[5]Note that the results reported in Table 4 for our approach are not directly comparable to those obtained by TAC KBP EDL 2017 systems [20] since we do not use an end-to-end system and omit the entity recognition stage.

Features	F1-score
popularity	0.723
popularity + cosine	0.729
popularity + cosine + sum	0.732
popularity + cosine + sum + max$_{1,2,3}$	0.750

Table 4: Linking accuracy (F1 score) on the TAC KBP-2017 dataset. Popularity and cosine similarity are the local mention-entity scores; sum and max, the global features, are defined in Eq. 11 and Eq. 12 respectively.

Combining all local and global features leads up to the best result (0.75). This can be further improved by using entity type information, filtering candidate entities by type. Given entity mentions in the documents together with their types (person, location, organization, etc.), we only retain candidate entities for which the BaseKB type corresponds. Taking into account type information, we achieved a F1 score of 0.79.

5.4 Comparison to State-of-the-Art Approaches

Finally, we compare our CEL approach to a series of EL systems which report state-of-the-art results on the AIDA-A, AIDA-B, Reuters128 and RSS500 datasets, namely:

- CNN approach to capturing semantic similarity for entity linking [12] where one CNN is used to extract contextual features (CNN);
- end-to-end neural attention-based entity linking [21] where entity recognition and linking are jointly solved using words and characters embedding (End-to-End);
- graph-based approaches operating on the mention-entity graph, resp. AIDA [16], AGDISTIS [36] and Babelfy [30];
- Neural collective entity linking where a graph convolution network is used to label mentions with entities from the KB [6]. Cosine similarity between entities is used as the relatedness measure (NCEL);
- Probabilistic bag-of-hyperlinks model for entity linking which uses counts of co-occurrences of entities along with a loopy belief propagation algorithm [13] (PBoH);
- VINCULUM [23] where both Ref (Eq. 6) and WLM (Eq. 4) are used as relatedness measure and a two-step approach is used for collective linking.

The list of EL systems above can be split into entity-by-entity approaches (CNN, End-to-End) and collective ones (AIDA, PBoH, AGDISTIS, Babelfy, VINCULUM, NCEL). For fair comparison, we point out that [12] does not use global score and both End-to-End and VINCULUM use an automatic entity recognition stage (where we do use the reference entity recognition) which is out of the scope of this paper. The scores for AIDA, PBoH, AGDISTIS and Babelfy were obtained from the online available platform GERBIL [37] and scores for [6, 12, 21, 23] are taken from the original papers, those methods being absent from the platform.

Results of the entity linking process evaluated in terms of micro-averaged F1 classification scores are reported in Tab. 5 except for VINCULUM which reports a different metric. On all four datasets,

Approach	AIDA-A	AIDA-B	Reuters128	RSS-500
CNN [12]	-	85.5	-	-
End-to-End [21]	89.4	82.4	54.6	42.2
NCEL [6]	79.0	80.0	-	-
AGDISTIS [36]	57.5	57.8	68.9	54.2
Babelfy [30]	71.9	75.5	54.8	64.1
AIDA [16]	74.3	76.5	56.6	65.5
PBoH [13]	79.4	80.0	68.3	55.3
WSRM	90.6	87.7	79.9	79.3

Table 5: Micro-averaged F1 score for different methods on the four datasets.

Approach	AIDA-A	AIDA-B
VINCULUM (Ref) [23]	69.1	66.4
VINCULUM (WLM) [23]	69.5	67.7
VINCULUM (both) [23]	69.4	67.5
WSRM	87.8	83.6

Table 6: F1 score for VINCULUM (as reported in [23]) and WSRM on the AIDA datasets.

the proposed WSRM with the logistic regression classifier does out-perform the AIDA, PBoH and NCEL collective linking approaches by a large margin (see below for statistical significance). The method is also competitive with respect to the end-to-end approaches. One interesting point to note is that AIDA and PBoH perform very differently on Reuters128 and RSS-500. This can be explained by the low density of mentions in the RSS-500 dataset with an average of 2 mentions per document. In these conditions, optimization-based approaches like PBoH do not perform well on short text unlike graph-based approaches. On the contrary, classification seems to be little affected by those drastic statistics.

Comparison with VINCULUM is reported in Tab. 6, using the macro averaged F1 classification as reported in [23]. While not directly comparable because VINCULUM does rely on automatic named entity recognition, the difference in macro averaged F1 scores of approx. 17 point is unlikely to be solely explained by entity recognition errors. As most CEL methods do use the same local scores, i.e., popularity and cosine similarity based on the skip-gram model, we can conclude that the improvement that we observe is for the most part attributable to the weighted semantic relatedness measure. The fact that contrary to AIDA, PBoH and NCEL we do not construct the mention entity graph explicitly but rather rely a set of independent contextual decisions—in other words, we use local optimization instead of global optimization of Eq. 3—also confirms this conclusion.

Statistical significance of the differences observed in Tab. 5 for methods present in GERBIL was assessed by means of a Student test for the Reuters 128 and RSS500 datasets. To this end, we built for each of these datasets 20 subsets of 20 randomly sampled documents. Micro-averaged error rates for each of the methods on each of the subsets are used in a paired t-test to compare methods, testing the equality of mean over two populations representing two CEL methods. For practical format reasons due to the GERBIL platform,

	Reuters128	RSS-500
WSRM vs AIDA	13.05	5.75
WSRM vs PBoH	4.27	14.33
PBoH vs AIDA	7.12	-6.78

Table 7: t-values for the statistical significance test of A/B pairs using the micro-averaged F1 scores. Rejection region at a risk $\alpha = 5\%$ for the equality of mean between A and B is $T > 2.539$ for the one-tail t-test and $|T| > 2.093$ for the two-tail t-test.

statistical significance for the AIDA datasets could not be tested. Table 7 reports the test statistics T values for A *vs.* B combinations of methods: for an alternative hypothesis "A is better than B" (one-tail t-test), statistical significance is achieved if $T > 1.73$ with a risk $\alpha = 5\%$, with $T > 2.539$ with a risk $\alpha = 1\%$; for an alternative hypothesis "A and B are different" (two-tail t-test), statistical significance is achieved if $|T| > 2.093$ and $|T| > 2.861$ for $\alpha = 5\%$ and $\alpha = 1\%$ respectively. Values reported in Tab. 7 consistently demonstrate significant gains of the WSRM-based CEL method over AIDA and PBoH. PBoH and AIDA are also significantly different one from another, however with different conclusions depending on the dataset as explained above.

6 CONCLUSION AND PERSPECTIVES

In summary, we demonstrated through experimental validation the benefit of better taking into account the semantic interrelations between the entities in a structured RDF KB for collective entity linking, improving over the state of the art. Conventional methods addressed this problem leveraging entity relatedness measures based on Wikipedia hyperlinks counts between two pages and on the presence of one relation between two entities in the KB. The drawback of those methods is for the first ones a lack of a true semantics and for the last one a lack of weighting of the relatedness between entities. The new entity relatedness measure that we proposed bears semantics since it uses relations between entities in BaseKB, a semantically structured KB, and attributes a weight to each couple of entities so that entities with a large total number of relations get large probability to be jointly mentioned in a document. This relatedness measure, combined with popularity and cosine similarity are the main ingredients we used to define a lightweight collective entity linking algorithm that was shown to compete with and outperform the state of the art in collective entity linking.

These results open new perspectives for taking into account the richness and expressiveness of structured KBs for entity linking, yet maintaining scalability. In particular, working on RDF KB allows not only to benefit from the KB semantics, but also opens the door to numerous extensions featuring semantic reasoning. For instance, reasoning can be leveraged to assess the relation between two entities enabling to define a robust semantics entity relatedness measure. In this work, we limited ourselves to direct relations between two entities assuming a saturated KB with the idea of demonstrating the interest of structured KB. The same idea could easily be extended to indirect relations, e.g. in the simplest form, paths of length two between two entities in the KB. But RDF KBs offer much more than that if the semantics of the relations are exploited. Typically, we expect semantic reasoning mechanisms to provide new features

for the linking stage, indicating to what extent and why two entities are related or not. On a general note, reasoning with RDF KBs in the context of collective entity linking is also bound to offer ways to explain the linking choices that were made. For instance, in the context of our iCODA project targeting the seamless collaborative exploration of collections of newspaper articles and of structured data and knowledge by journalists and librarians, justifying and explaining machine-based decisions is paramount to facilitate the validation of the machine-based links and their acceptance by users.

REFERENCES

[1] Ayman Alhelbawy and Robert Gaizauskas. 2014. Graph ranking for collective named entity disambiguation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Vol. 2. 75–80.

[2] Roi Blanco, Giuseppe Ottaviano, and Edgar Meij. 2015. Fast and space-efficient entity linking for queries. In *Proceedings of the 8th ACM International Conference on Web Search and Data Mining*. 179–188.

[3] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5 (2017), 135–146.

[4] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. 1247–1250.

[5] Razvan Bunescu and Marius Paşca. 2006. Using encyclopedic knowledge for named entity disambiguation. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*. 9–16.

[6] Yixin Cao, Lei Hou, Juanzi Li, and Zhiyuan Liu. 2018. Neural collective entity linking. In *Proceedings of the 27th International Conference on Computational Linguistics*. 675–686.

[7] Diego Ceccarelli, Claudio Lucchese, Salvatore Orlando, Raffaele Perego, and Salvatore Trani. 2013. Learning relatedness measures for entity linking. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*. 139–148.

[8] Silviu Cucerzan. 2007. Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. 708–716.

[9] Mark Dredze, Paul McNamee, Delip Rao, Adam Gerber, and Tim Finin. 2010. Entity disambiguation for knowledge base population. In *Proceedings of the 23rd International Conference on Computational Linguistics*. 277–285.

[10] Greg Durrett and Dan Klein. 2014. A joint model for entity analysis: coreference, typing, and linking. *Transactions of the Association for Computational Linguistics* 2 (2014), 477–490.

[11] John R Firth. 1957. A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis* (1957).

[12] Matthew Francis-Landau, Greg Durrett, and Dan Klein. 2016. Capturing semantic similarity for entity linking with convolutional neural networks. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 1256–1261.

[13] Octavian-Eugen Ganea, Marina Ganea, Aurelien Lucchi, Carsten Eickhoff, and Thomas Hofmann. 2016. Probabilistic bag-of-hyperlinks model for entity linking. In *Proceedings of the 25th International Conference on World Wide Web*. 927–938.

[14] Octavian-Eugen Ganea and Thomas Hofmann. 2017. Deep joint entity disambiguation with local neural attention. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2619–2629.

[15] Nitish Gupta, Sameer Singh, and Dan Roth. 2017. Entity linking via joint encoding of types, descriptions, and context. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2681–2690.

[16] Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. 2011. Robust disambiguation of named entities in text. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. 782–792.

[17] Hongzhao Huang, Larry Heck, and Heng Ji. 2015. Leveraging deep neural networks and knowledge graphs for entity disambiguation. *arXiv preprint arXiv:1504.07678* (2015).

[18] Heng Ji and Nothman. 2016. Overview of TAC-KBP2016 tri-lingual EDL and its impact on end-to-end cold-start KBP. *Proceedings of the 2016 Text Analysis Conference* (2016).

[19] Heng Ji, Joel Nothman, Ben Hachey, and Radu Florian. 2015. Overview of TAC-KBP2015 tri-lingual entity discovery and linking. In *Proceedings of the 2015 Text Analysis Conference*.

[20] Heng Ji, Xiaoman Pan, Boliang Zhang, Joel Nothman, James Mayfield, Paul McNamee, and Cash Costello. 2017. Overview of TAC-KBP2017 13 languages entity discovery and linking. In *Proceedings of the 2017 Text Analysis Conference*.

[21] Nikolaos Kolitsas, Octavian-Eugen Ganea, and Thomas Hofmann. 2018. End-to-end neural entity linking. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*. 519–529.

[22] Phong Le and Ivan Titov. 2018. Improving entity linking by modeling latent relations between mentions. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. 1595–1604.

[23] Xiao Ling, Sameer Singh, and Daniel S Weld. 2015. Design challenges for entity linking. *Transactions of the Association for Computational Linguistics* 3 (2015), 315–328.

[24] Ming Liu, Gu Gong, Bing Qin, and Ting Liu. 2019. A multi-view-based collective entity linking method. *ACM Transactions on Information Systems.* 37, 2, Article 23, 29 pages. https://doi.org/10.1145/3300197

[25] Weiming Lu, Yangfan Zhou, Haijiao Lu, Pengkun Ma, Zhenyu Zhang, and Baogang Wei. 2017. Boosting collective entity linking via type-guided semantic embedding. In *Proceedings of the National CCF Conference on Natural Language Processing and Chinese Computing*. 541–553.

[26] Pablo N. Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. DBpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems*. 1–8.

[27] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of Advances in Neural Information Processing Systems*. 3111–3119.

[28] David Milne and Ian H Witten. 2008. Learning to link with Wikipedia. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*. 509–518.

[29] Jose G. Moreno, Romaric Besançon, Romain Beaumont, Eva D'hondt, Anne-Laure Ligozat, Sophie Rosset, Xavier Tannier, and Brigitte Grau. 2017. Combining word and entity embeddings for entity linking. In *Proceedings of European Semantic Web Conference*. 337–352.

[30] Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics* 2 (2014), 231–244.

[31] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. 1532–1543.

[32] Minh C. Phan, Aixin Sun, Yi Tay, Jialong Han, and Chenliang Li. 2018. Pair-linking for collective entity disambiguation: Two could be better than all. *IEEE Transactions on Knowledge and Data Engineering*.

[33] Michael Röder, Ricardo Usbeck, Sebastian Hellmann, Daniel Gerber, and Andreas Both. 2014. N^3- A collection of datasets for named entity recognition and disambiguation in the NLP interchange format. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*. 3529–3533.

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

[35] Valentin I. Spitkovsky and Angel X. Chang. 2012. A cross-lingual dictionary for English Wikipedia concepts. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*. 3168–3175.

[36] Ricardo Usbeck, Axel-Cyrille Ngonga Ngomo, Michael Röder, Daniel Gerber, Sandro Athaide Coelho, Sören Auer, and Andreas Both. 2014. AGDISTIS - Graph-based disambiguation of named entities using linked data. In *Proceedings of the International Semantic Web Conference*. 457–471.

[37] Ricardo Usbeck, Michael Röder, Axel-Cyrille Ngonga Ngomo, Ciro Baron, Andreas Both, Martin Brümmer, Diego Ceccarelli, Marco Cornolti, Didier Cherix, Bernd Eickmann, et al. 2015. GERBIL: General entity annotator benchmarking framework. In *Proceedings of the 24th International Conference on World Wide Web*. 1133–1143.

[38] W3C. 2014. RDF 1.1 Concepts and Abstract Syntax. https://www.w3.org/TR/rdf11-concepts.

[39] W3C. 2014. RDF 1.1 Primer. https://www.w3.org/TR/rdf11-primer.

[40] W3C. 2014. RDF 1.1 Semantics. https://www.w3.org/TR/rdf11-mt.

[41] Han Wang, Jin Guang Zheng, Xiaogang Ma, Peter Fox, and Heng Ji. 2015. Language and domain independent entity linking with quantified collective validation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 695–704.

[42] Ikuya Yamada, Hiroyuki Shindo, Hideaki Takeda, and Yoshiyasu Takefuji. 2016. 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*. 250–259. https://doi.org/10.18653/v1/k16-1025