# FEEL: Framework for the integration of Entity Extraction and Linking systems

Julio Hernandez<sup>a</sup>, Jose L. Martinez-Rodriguez<sup>a,d,\*</sup>, Ivan Lopez-Arevalo<sup>a</sup>, Ana B. Rios-Alvarado<sup>b</sup>, Edwin Aldana-Bobadilla<sup>c</sup>

 <sup>a</sup> Cinvestav-Tamaulipas Victoria, Mexico
<sup>b</sup> Facultad de Ingeniería y Ciencias Universidad Autónoma de Tamaulipas Victoria, Mexico
<sup>c</sup> Conacyt-Cinvestav Victoria, Mexico
<sup>d</sup> Unidad Académica Multidisciplinaria Reynosa-RODHE Universidad Autónoma de Tamaulipas Reynosa, Mexico

# Abstract

Entity extraction and linking (EEL) is an important task of the Semantic Web that allows us to identify real-world objects from text and associate them with their respective resources from a Knowledge Base. Thus, one purpose of the EEL task is to extract knowledge from text. In recent years, several systems have been proposed for addressing such a task in several domains, languages, and knowledge bases. In this sense, some systems that combine the benefits of varied EEL systems have been proposed in a kind of ensemble system (like in Machine Learning) to provide better performance and extractions than using a single system. However, there are no clear indications for the selection, configuration, and result integration of EEL systems in an ensemble setting. This paper proposes a framework for the integration of EEL systems by providing recommendations for the selection of systems, the configuration of input parameters, the execution of systems, and the final integration of results through a filtering strategy that measures the occurrence of entities and detects the overlapping of entities. Based on the proposed framework, we implemented a system using existing EEL systems (through publicly available APIs). The experiments were performed through the GERBIL framework. Our results demonstrate an improvement of the micro/macroprecision and recall of the implemented system regarding the selected individual EEL systems over seven datasets.

Keywords: named entity disambiguation, entity linking, entity joining, ensemble extractor

# 1. Introduction

The Semantic Web has an important goal of providing a formal data representation in order to share and reuse information by people and applications [1]. This goal is being addressed by varied

Preprint submitted to Web Semantics

standards and protocols such as the Resource Description Framework (RDF) and the Linked Open Data (LOD) principles [2], from which the data is organized into a graph structure, where nodes correspond to information resources (such as realworld objects) and edges to descriptions (adhered to formal vocabularies) between such resources<sup>1</sup>. However, due to 1) the large scale and heterogene-

<sup>\*</sup>Corresponding author

Email addresses: julio.hernandez@cinvestav.mx (Julio Hernandez), lazaro.martinez@cinvestav.mx (Jose L. Martinez-Rodriguez), ilopez@cinvestav.mx (Ivan Lopez-Arevalo), arios@docentes.uat.edu.mx (Ana B. Rios-Alvarado), edwyn.aldana@cinvestav.mx (Edwin Aldana-Bobadilla)

 $<sup>^{1}</sup>$ In LOD, every resource (including edges) must be denoted by a URI identifier, which is used to provide more information of the resource through the Internet (as is done in the traditional Web).

ity of the data in sources such as the Web, and 2) linguistic problems such as the detection of synonymy and ambiguity, the extraction (and formal representation) of resources from unstructured text is a challenging task, where real-world objects such as people, organizations, locations, among others (known as named entities) must be extracted and associated with an existing resource from a Knowledge Base (KB). This process is known as *Entity* Extraction and Linking (EEL), whose underlying purpose is to provide the identification, description, and interconnection of information on the Semantic Web. In recent years, several EEL approaches have been proposed [3, 4, 5, 6], in which KBs such as DBpedia [7] or YAGO [8] are commonly used for linking mentions of named entities<sup>2</sup> from text.

Existing EEL approaches provide support to a diversity of domains, heterogeneity of languages and availability of different KBs. However, this imposes restrictions to cover such elements within a single approach. In this sense, and based on the idea of ensemble systems (like in Machine Learning), some systems [11, 12] propose a combination of processes and/or results provided by existing EEL systems with the purpose of taking advantage of the benefits offered by individual systems. The general idea of an ensemble learning system is to provide better performance compared to the single classifiers. Following this idea, in an ensemble EEL system, the output of two or more systems is unified in a single result, providing better performance than a single EEL system. Moreover, as EEL systems provide a distinct number of extracted named entities, the idea is also to retrieve a concise and unified result, often with more identified entities than a single system.

However, ensemble EEL systems involve the selection of systems, parameter settings, and integration decisions in different stages to get systems implemented with a homogeneous result. First, the selection and execution of EEL systems, where several aspects may get involved, such as the domain, resource requirements, and the execution environment. Second, regarding the importance of the parameter configuration, the output of EEL systems can be directly affected by the selection of the input parameters. For example, the *confidence parameter* controls the degree of matching between mentions of named entities and URIs from a KB (when higher, the precision of systems is assumed better at the cost of low recall). Finally, combining the output of different EEL systems may produce duplicated and/or partially overlapped entity tuples (i.e., named entity mentions that share a fragment from the input text). Hence, it is desired a scenario that provides recommendations for the selection, configuration, and integration of EEL systems in an ensemble EEL system.

This paper proposes a framework that suggests criteria to assist with the implementation of ensemble EEL systems. Although individual EEL systems can be integrated at different stages of the EEL process (e.g., for the detection of named entity mentions or linking with resources), the main focus of the proposed framework is regarding the integration of the final output provided by individual EEL systems. Our approach differs from other existing frameworks (e.g., such as NERD [13]) in the sense that we suggest recommendations for the selection, configuration, invocation, and output integration of EEL systems. Regarding the evaluation, we implemented an ensemble EEL system (called FEELink) by following the stages of the proposed framework. FEELink takes into account the output integration of four EEL systems available via REST service invocations (DBpedia Spotlight [5], TagMe [14], WAT [6], and Babelfy [15]). Moreover, it was configured and tested through the GERBIL framework [16] over seven distinct datasets from the state of the art. The obtained results demonstrate that FEELink can be configured to outperform individual EEL systems performance in terms of the micro-recall, micro-precision and, in some cases, to the micro-F1 measures. From this, the contributions of this paper are as follows:

- A framework with adaptable recommendations and steps to select, configure, implement, and integrate EEL systems in an ensemble setting.
- A strategy for the configuration of input parameters (particularly the confidence value) for the selected EEL systems according to experiments performed over the GERBIL framework.
- A named entity filtering strategy to integrate the output of different EEL systems. This

<sup>&</sup>lt;sup>2</sup>Although named entities commonly refer to instances of classes such as person, location and organization [9], EEL approaches also cover "entities" or things that belong to other classes or generic objects (e.g., car, chair) [10]. To avoid confusions, we handle the term *named entities* to refer to both things throughout this document.

strategy measures the occurrence of entities and also detects overlapping entities.

The remainder of this paper is organized as follows: Section 2 presents the related work about EEL and ensemble EEL systems. The description and implementation details of the proposed framework are presented in Section 3. The experiments are presented in Section 4. Finally, Section 5 presents the conclusions.

#### 2. Related work

One goal of the Semantic Web includes the formal representation of textual contents (originally from the Web) following specifications such as the RDF data model and the Linked Data principles. Such specifications provide guidelines for the interconnection and sharing of information through resources linked to some KB. In this sense, resources refer to abstract representations of elements called *named entities*, that is to say, real-world things such as persons, organizations, locations, among others, that can be denoted with a proper name (e.g., Tim Berners-Lee, IBM, Toyota). This section presents some concepts and the related work regarding the extraction of named entities by EEL systems and ensemble EEL systems.

#### 2.1. EEL

The Named Entity Recognition (NER) task is in charge of identifying mentions of named entities and their type from text. However, in the context of the Semantic Web, and following the Linked Data principles, named entities must be identified by a URI, where every named entity is referenced by a resource from a KB. Thus, the identification of mentions of named entities from text and their association with their respective resources from a KB are part of the EEL task.

The EEL task is commonly composed of three steps: *Spotting*, *Candidate collecting*, and *Disambiguation* [17]. The first step refers to finding mentions of named entities in text. The second step refers to collecting candidate resources from a KB that can be linked to the mentions found in the previous step. Finally, and given that more than one candidate can be collected from the previous step, the disambiguation step refers to the selection of the most likely resource that better describes every mention of a named entity. The EEL task has been addressed by diverse systems and approaches that cover features related to text characteristics, languages, and domains.

Text characteristics refer to the features of the input text in terms of writing and contextual information [18]. For example, text from web blogs or social networks often contains misspelled words or reduced content that would difficult the disambiguation of named entities. Although systems based on general purpose data commonly assume well-written text as input, other approaches such as [14, 19, 6] are configured for working with short or noisy texts. Existing studies provide a comparison of systems over diverse datasets that would be useful for selecting a system according to the features of the data, the performance of systems, and the trade-off between metrics (e.g., precision, recall) [16].

Regarding the language, the selection of an EEL system may depend on the requirements of one (or multiple) languages. Although English is prevalent in most of the approaches, other languages such as Spanish, French, Chinese, among others, or a combination of them are also addressed by existing approaches [20].

Regarding the domain, existing EEL systems have been proposed to address topics of diverse domains. This aspect depends largely on the dictionaries or KBs used for spotting and disambiguation, respectively. For example, the approaches Sieve [21] and Zheng *et al.* [22] extract and link entities from the medical domain (using the SNOMED-CT and Bioportal ontologies respectively), ELMDist [23] for the music domain (using MusicBrainz and annotated Last.fm data), and Pantel and Fuxman [24] for catalog products data (with data from a search engine). However, most of the systems rely on general purpose data provided by DBpedia or YAGO (extracted from Wikipedia data).

According to their performance [25, 16], some representative EEL systems are described as follows:

• TagMe [14]. It uses a Wikipedia based dictionary to perform the Spotting and Candidate collecting steps. The disambiguation step uses the anchor text of Wikipedia pages to perform two ranking measures: commoness and relatedness. The former measures the frequency that an anchor text is associated to a Wikipedia entity. And the second refers to a co-citation measure that indicates how frequently candidate entities for distinct mentions are associated to the same Wikipedia article.

- DBpedia SpotLight [5]. It employs NLP (Ling-Pipe [26]) for recognizing mentions of named entities and later employs Wikipedia data to create an Aho-Corasick index to perform keyword and string matchings that collects candidates ranked through a probabilistic model based on a TF-IDF variant (for disambiguation).
- WAT [6]. Succeeding to TagMe, WAT also uses a Wikipedia based dictionary. Thus, the mentions of named entities are supported by the anchor text and occurrence measures. However, WAT provides a PageRank-based disambiguation measure, which uses an entitymention graph that provides contextual features to the commonness function.
- Babelfy [15]. It performs the Spotting step through POS tagging. Candidates are collected using a string matching function over a personalized KB constructed from Wikipedia and WordNet. Named entities are later disambiguated through a function that finds the densest subgraph linking entities and mentions.

#### 2.2. Ensemble EEL

In Machine Learning, ensemble learning methods refer to algorithms that train classifiers and combine their functionalities and/or outputs. The idea is to classify an instance by taking a decision according to the prediction of some classifiers [27, 28]. Diverse ensemble learning approaches have been proposed, such as Bagging (multiple models trained on different sub-samples of the same dataset) and Boosting (multiple models trained sequentially) [27, 28].

Inspired by ensemble learning, different EEL systems can be combined for producing a unique and enriched result. Although the original idea of ensemble systems (from Machine Learning) is to integrate the output of distinct classifiers using a function (e.g., majority voting), in this case, the integration of the output provided by individual EEL systems is considered as an *ensemble approach*. Along these lines, an example of the EEL process is depicted in Figure 1, where the three EEL steps are performed. First, four mentions of named entities are obtained (e.g., Jon, The Walking Dead).



Figure 1: Example of the EEL process. The result of three hypothetical systems (A,B,C) is integrated into a final result (ABC) through an ensemble system.

Second, according to the obtained mentions, a list of resource candidates are obtained for linking from a KB (in this case from DBpedia) (e.g., dbr:Jon\_Bernthal<sup>3</sup>). For the disambiguation step, the result of three hypothetical systems is provided (A, B, C), where a resource is assigned for every mention. Note that different results are obtained, where every EEL system may obtain distinct mentions, candidates, and final resources. Finally, as a motivating example, the ABC output refers to the desired result integration of the three systems in an ensemble setting.

As shown in Figure 1, the combination of the three EEL systems is subject to aspects such as the selection of named entities (and additional decisions and configurations) to adapt a final output. The general intuition is to leverage features provided by individual systems. For example, the usage of data from diverse KBs, languages, particular domains, or probably providing trade-offs between the metrics applied for the evaluation (e.g., precision, recall, or F1).

Ensemble approaches have demonstrated their effectiveness in diverse subtasks of the EEL task. For example, for the Named Entity Recognition (NER) task, Speck and Ngomo [29] presented a thorough comparison of the performance of NER systems based on ensemble learning. Their results demonstrated an increased performance in terms of the

<sup>&</sup>lt;sup>3</sup>For space reasons, we use URI prefixes (namespaces) in accordance with the service hosted at http://prefix.cc, where dbr:Jon\_Bernthal represents a contraction for the URI http://dbpedia.org/resource/Jon\_Bernthal

F1 measure compared to the tools tested in their experiment.

Ensemble approaches have also been used in the entity disambiguation task. That is to say, the process involves a combination of approaches or algorithms for linking an already recognized named entity mention to its respective resource from a KB. In this sense, systems such as DEXTER [11] and BEL [30] have been proposed for such a task. DEXTER implements and tests a platform for comparing the results provided by diverse disambiguation strategies. However, the platform does not provide decisions for the integration of the diverse results. On the other side, BEL provides a strategy for retrieving candidates from YAGO [8] that are later disambiguated through a majority-voting algorithm that relies on various ranking classifiers and contexts (of mentions), at the cost of an increased complexity for the combination of algorithms.

Ensemble approaches have also been presented for the integration of the final output provided by more than one EEL system. The purpose is to produce a single (and homogeneous) result composed of tuples containing the named entity mention and its URI from a KB. For such purpose, various similar systems such as NERD [13], NTUNLP [31], Ruiz and Poibeau [32], and WESTLAB [3] have been proposed so far. A brief description of such approaches is mentioned as follows.

NERD (Named Entity Recognition and Disambiguation) [13] is a framework to integrate the output of 10 different EEL systems in a single output. The integration process is made through a manually created ontology which provides a set of mappings between the systems output. Additionally, the ontology describes 6 different types of named entities (person, organization, country, city, time and number) and provides equivalence relations between types of named entities defined by the KBs used by the tools. Such ontology is mainly used for the integration of entity types (for the NER task), where a set of statistics are presented (as result) for comparing the number of types discovered by each tool.

NTUNLP [31] is a system that combines a dictionary matching strategy (composed of Freebase [33] candidates) with the output provided by two existing EEL systems (DBpedia Spotlight [5] and TagMe [14]). NTUNLP merges the result of the EEL systems by including only non overlapped entities to those obtained in the matching strategy.

Ruiz and Poibeau [32] described an integration of

the output provided by five public open-source EEL systems. They proposed a way to define the confidence parameters used for the invocation of the EEL services according to experiments performed over four datasets with the support of the BAT framework<sup>4</sup> [34].

WESTLAB [3] combines the Stanford NER tool and four EEL systems for detecting mentions of named entities (without KB identifier) that are later merged by means of an algorithm that keeps the longest string matching with respect to the input text. The resulting mentions are used to retrieve candidates from DBpedia that are finally disambiguated through a centrality algorithm that leverages the corresponding Wikipedia structure of the candidates.

Although different ensemble EEL strategies have been proposed for different steps of the EEL task, this paper is focused on those approaches that consider the output integration provided by EEL systems. In this regard, there is a lack of recommendations for managing the required components for consuming and integrating the output of such systems. For example, for selecting the required EEL systems, for configuring its input parameters (particularly the confidence), and for the integration of named entities. This paper proposes such aspects by providing a framework for the integration of EEL systems.

# 3. Framework

This section presents a description of the proposed framework, it is called FEEL (*Framework* for the integration of Entity Extraction and Linking systems). It is focused on representing a structure or skeleton with recommendations, concepts, and tasks that must be followed to integrate EEL systems in an ensemble setting. The architecture of the proposed framework is presented in Figure 2, it is composed of three stages: Parameter Configuration, System Invocation, and Data Consolidation. A description of such stages is presented in subsequent subsections. Moreover, together with such a description, we provide the description of an ensemble EEL system implemented using four EEL systems<sup>5</sup> according to the stages of the pro-

<sup>&</sup>lt;sup>4</sup>https://github.com/marcocor/bat-framework. All URLs in this paper were last accessed on 2020/02/24.

 $<sup>^5\</sup>mathrm{A}$  demo of the implementation is available for testing purposes on <code>https://github.com/ragnarok85/FEEL</code>



Figure 2: General overview of the proposed framework to integrate EEL systems. It is composed of steps (solid boxes) and substeps (dashed boxes) that provide considerations in three stages.

posed framework. This implementation is called *FEELink*.

## 3.1. Parameter Configuration

The goal of this stage is the configuration of the resources required by the different EEL systems involved in the framework. Thus, this stage is composed of three main steps: System selection, Resource management, and Parameter tuning.

#### 3.1.1. System selection

This step provides aspects to be taken into account for the selection of EEL systems and/or approaches that will constitute the ensemble system. The selection of an EEL system involves diverse criteria required by users or applications. Thus, with respect to the input data, some criteria to take into account for the selection of an EEL system are the domain, language, and text characteristics (as presented in Section 2.1). Other considerations regarding the performance, usage, quality, or availability of systems should also be explored for the selection of a system as provided by Martinez-Rodriguez et al. [17].

In this regard, FEELink considers the use of EEL systems (through publicly available APIs) to be applied over general domain datasets. In this sense, four EEL systems were selected for the implementation: TagMe [14], DBpedia Spotlight [5], Babelfy [15], and WAT [6]. Although there exist several others EEL systems, such selection of systems was performed for practical reasons and according to the following criteria:

• Reported studies. The systems demonstrate a balance between precision and recall as demon-

strated in existing studies  $[25, 16]^6$ .

- Datasets. The systems deal with the domain, language, and features of several datasets. That is, Babelfy was evaluated over short text datasets and English articles [15], TagMe was evaluated over short text fragments obtained from the IITB dataset [14], WAT was evaluated over general webpages and news articles [6], and DBpedia Spotlight was evaluated over a news corpus [5].
- Availability. The systems offer free access/service for non commercial purposes.

## 3.1.2. Resource management

This step considers the management of hardware and software resources required by every EEL system. While the hardware requirements involves common components (e.g., RAM, CPU, storage), the software involves configurations and applications required by the EEL systems. FEELink was implemented as a Java application. We used a computer with Intel Core i5 processor (2.6 GHz) and 8GB of RAM, running OSX (64-bit) as the operating system for the integration. We configured the selected EEL systems on a local (DBpedia Spotlight) and remote environments (Babelfy, TagMe, WAT). Although the former was installed by following the given instructions<sup>7</sup>, all the systems were invoked through the Apache HTTP client<sup>8</sup>. Note that the local system provides a high service availability at the cost of constant monitoring of resources, configuration, installation, among others.

#### 3.1.3. Parameter tuning

In general, EEL systems typically require some input parameters (to be stated at execution/invocation time) such as *confidence*, type of extraction, input text, language, output format, token-key, among others. This step involves the description and configuration of the most common input parameters, particularly the confidence.

1. Confidence. The *confidence value* refers to the degree of matching between a mention and the selected resource (candidate) from a KB. This

value is a threshold for limiting the number of named entities extracted from text, where a higher value supposes an increasing precision at the cost of recall (and vice versa). To the best of our knowledge, a way to select an appropriate confidence value is not often mentioned for every EEL system. Rather than empirically configuring the confidence parameter for every selected system and testing dataset, the proposed framework considers the configuration of such a parameter by testing different systems over a fixed number of datasets and varied confidence values through the support of the GERBIL framework<sup>9</sup>. The idea relies on selecting the confidence values that maximizes the micro-F1 measure for a group of datasets. Note that the micro-F1 is considered for the configuration of parameters because it provides a trade-off between precision and recall and because it measures the average performance of systems regarding all annotations in a dataset [34] (details of the measures are presented in Section 4.1). Hence, the proposed strategy to configure the confidence parameter for every EEL system is as follows:

- (a) Select the dataset(s) for the configuration. Note that this configuration is generalized for the domain of the datasets and thus, it may not apply for different or specialized domains (e.g., medical, biology).
- (b) Execute the EEL system over the set of selected datasets and obtain the micro-F1 results<sup>10</sup>. Since the value to be configured is the confidence, the system must be executed several times by fluctuating the confidence value. As mentioned, the GERBIL framework was used to support this step.
- (c) Obtain the confidence interval that produces the higher micro-F1 values for each tested dataset. In case of ties, the greatest confidence value is selected.
- (d) Obtain the median of the confidence interval. For example, if the higher micro-F1 values for some EEL system tested

 $<sup>^{6}\</sup>mathrm{Although}$  the AIDA system [35] meets our requirements, it was not available at the time of writing

<sup>&</sup>lt;sup>7</sup>https://github.com/dbpedia-spotlight/dbpediaspotlight

<sup>&</sup>lt;sup>8</sup>https://hc.apache.org

<sup>&</sup>lt;sup>9</sup>GERBIL [16] is a general Linked Data benchmarking framework that provides the option of evaluating diverse EEL systems over several datasets through different measures.

 $<sup>^{10} {\</sup>rm In}$  this case, only the micro-F1 is considered because we look for a trade-off between the (micro) precision and recall.

over three datasets are obtained through the confidence values 0.4, 0.6, 0.7 (respectively), then the median<sup>11</sup> of the confidence interval is 0.55 (considering also 0.5 within the confidence interval).

(e) The result is rounded up (optional). This is performed for practical reasons. However, the exact median of the confidence interval can be used instead.

In general, the strategy for selecting the confidence values relies on performing experiments (through the GERBIL framework) with the selected EEL systems (and confidence values) until the micro-F1 measure is maximized for the selected datasets. We present particular details of the confidence parameter tuning for the four EEL systems of FEELink in Section 4.2.

- 2. Type of extraction. EEL systems may offer the option to only extract named entities from text (without KB linking) or linking mentions of named entities to KB resources or both tasks. Additionally, the output can be restricted to only named entities or concepts (those things not identified by a name, e.g., chair, car, database). Hence, the extraction and linking of named entities are the focus of the proposed framework and thus, configured for every selected EEL system.
- 3. Input text. The text to be processed by the EEL system. It is assumed that the input is entered as plain text.
- 4. Language. EEL systems detect and cover different languages. In case that the language of the input text is not detected, this option must be specified. This framework is focused on the English language.
- Output format. The systems can be configured to return entities in a variety of formats. For example, FOX<sup>12</sup> provides annotations of named entities in formats based on Linked Data (e.g., Turtle, NQuads, RDF/XML) while

DBpedia Spotlight and TagMe provide a JSON output mainly. Thus, it is recommended to select a homogeneous format for the systems in the framework in order to facilitate the parsing of results. The four systems in FEELink provide JSON output and thus, it was selected.

6. Token-key. Remote systems often require a token-key to control the number of requests performed by the users, where a limit of the daily requests is different for every system. For example, Babelfy provides 1000 daily requests but TagMe does not declare a request limit. Note that this aspect may suppose a significant economic cost for private use in systems such as IBM Watson<sup>13</sup> or already mentioned EEL systems used for business/commercial purposes.

The selected EEL systems for FEELink were configured to extract and link entities to a KB, to process plain text in English, and to produce the output in JSON format. Moreover, WAT, TagMe, and Babelfy required a token-key obtained from their official websites under registration. However, the confidence parameter was tuned according to the strategy of the proposed framework (later depicted in the Experiments).

## 3.2. System Invocation

After the parameters and resources are configured, the next stage is the invocation/execution of the selected EEL systems. The aim is to obtain a set of named entity tuples<sup>14</sup> from a given text (even if there are duplicated/overlapped entities). This stage considers two steps: Request preparation and Field parser.

#### 3.2.1. Request preparation

This step refers to the execution/invocation of systems, where two aspects are considered: invocation and possible runtime exceptions. The invocation refers to the step where the systems are executed. Thus, it involves the construction of the request according to the input parameters defined in the previous stage. Regarding runtime exceptions,

 $<sup>^{11}\</sup>mathrm{We}$  use the median because our purpose is to get a balance in performance rather than using particular values for particular datasets. Moreover, the median is a robust central tendency measure able to deal with outliers.

<sup>&</sup>lt;sup>12</sup>http://fox-demo.aksw.org/#!/demo

 $<sup>^{13}</sup>$ https://www.ibm.com/watson/alchemy-api.html

<sup>&</sup>lt;sup>14</sup>We consider a named entity tuple as a structure of the form <**M**,**URI**,**s**,**o**>, containing the mention of the named entity (M), its URI from a KB resource, and the start (s) and end (e) offsets (containing the index of the named entity from the input text).

if a failure occurs at the execution of any EEL system (i.e., a request cannot be satisfied, connection refusal, response timeouts, empty results), the ensemble system should be able to continue with the execution of the remaining EEL systems in order to save as many results as possible from a given text.

In the particular case of FEELink, the invocation of the four selected EEL systems was performed through HTTP POST requests that include the input text, the token-key, the output format, and the language as configured before. Moreover, FEELink was able to handle exceptions such as timeouts, connection refusals, and empty results. Additionally, pauses of 3 seconds between requests were performed in order to avoid a denial of service; this time value was set for testing purposes but a shorter time would probably be accepted too.

## 3.2.2. Field parser

This step involves the gathering of named entities retrieved by the individual EEL systems. It refers to the identification of the diverse outputs, which must contain the elements of the entity tuples.

FEELink is focused on four main parameters: Mention, URI, Offset, and Confidence. The tags used by each EEL system are presented in Table 1. In this case, we indicate if the parameter is set as input, obtained from the output (I|O) or not applicable for the configuration (–). Thus, when the confidence threshold cannot be defined as input, then it is filtered from the output (given the entity score). We also included the parameters for the start/end offset but, in the particular case of Babelfy, the *charFragment* tag returns the start and end indexes of the mention and thus, the mention must be extracted from the original text.

Parameter	TagMe/ WAT	Babelfy	DBpedia Spotlight
Input text (I)	text	text	text
Confidence threshold (I)	_	_	confidence
Token-key (I)	gcube-token	key	_
Mention (O)	spot	charFragment	surfaceForm
URI (O)	entity	DBpediaURL	URI
Confidence (O)	rho	coherenceScore	similarityScore
Offset (O)	$\operatorname{start}/\operatorname{end}$	${\rm charFragment}({\rm start/end})$	offset/-

Table 1: Tag names used by EEL systems for Input and Output parameters.

# 3.3. Data Consolidation

This final stage covers the integration of results provided by the selected EEL systems.

# 3.3.1. Filtering

The results obtained through the invocation of the EEL systems often contain overlapped (and/or duplicated) entities that should be filtered in order to get a unified result. Hence, four cases are considered for filtering, where entities can be partially or totally overlapped:

- 1. Entity frequency. This case measures how often a named entity is retrieved by the EEL systems. Named entities below a threshold f (between one and the max number of EEL systems) are removed.
- 2. Duplicated entity mentions. This case occurs when two or more named entity tuples have the same text as mention (surface form), but different identifier (URI). A first option to tackle this problem is (as described by ensemble learning systems) to follow a majority voting strategy [30], where the most frequent entity retrieved by the systems is selected. In case of a tie (same frequency), a selection decision is based on a relevance/ranking function. Such a ranking is commonly defined according to the accuracy results of experiments performed by individual systems over specific datasets.
- 3. Duplicated tuples. It refers to tuples that share the same mention and URI. For this case, duplicated tuples are removed, preserving the first of the tuples returned by either of the invoked tools.
- 4. Partial entity overlapping. This case refers to pairs of named entities that share part of the mention but may or not share the same URI. In other words, it occurs when a mention is partially contained or overlapped within another mention. For example, as presented in Figure 1, the entity Dead(dbr:Dead) is partially contained (subsumed) within the entity The Walking Dead (dbr:The\_Walking\_Dead\_(TV\_series)). We suggest to keep the longest matching named entity with respect to the input text; as pro-

posed by EEL approaches such as NTUNLP [31], ADEL [12], AGDISTIS [36], HERD [37], and KORE [38].

For the case 1, a relevance function is involved in the filtering for providing decisions in the selection of named entities. In this sense, such a function is configured to deal with the entity overlapping case, where the longest matching entity would be included with their subsumed named entities. For the case 3, considering the example of Figure 1, the entity **Dead** would also be selected (allowing overlapping named entities) in order to provide additional information or context. Although this problem has been studied within the Named Entity Recognition task [39], to the best of our knowledge, it has not been addressed for the EEL task. However, defining such a function is not within the focus of this work.

The filtering process is described in the Algorithm 1. The algorithm takes as input a list of EEL tuples and a frequency value f. A first filter is performed (line 1) to remove those named entities with a frequency below a threshold (f). A list of duplicated named entities by mention and offset (rEEL) are obtained and removed from the EEL list (lines 2 and 3). Later (line 4), the named entities are deduplicated (through a majority voting and/or ranking function). The final list of named entities is obtained in line 5 (as a filtered set), which is initialized with the first entity (substracted) from the EEL list. This is performed to iterate over existing named entities in the EEL list and in a list of filtered named entities (lines 6 and 7). Overlapped named entities are found by offset and by the length (LEN) of the mention (named entities with the same mention were previously filtered) as presented in lines 9 and 10. If the entity in the EEL list is longer than the entity in the filtered list, then it is replaced with the longest matching entity (line 11). Otherwise, if no overlapping is found, the entity is added to the filtering list (line 14).

The entities provided by the four selected EEL systems in FEELink were filtered according to a majority voting strategy. In case of a tie, the entity from the best-ranked EEL system was selected. Such ranking of systems was manually configured according to the micro-F1 results presented in Table 3 (Section 4), where WAT was the best-ranked system and Babelfy obtained the lowest results of the study (ranked at the last position). Note that we only handle the output provided by EEL systems. However, more robust functions (such as Bagging and Boosting) can be implemented if the EEL systems are trained and reconfigured from scratch.

Algorithm 1: Filtering of overlapped named
entities
<b>Data:</b> EEL tuples $<$ M,URI, $s$ , $e$ >, $f$
<b>Result:</b> FDOE, the set of non-duplicated and
non-overlapped named entities
1 EEL $\leftarrow$ EEL $-$ FreqThreshold(EEL, $f$ );
2 $rEEL \leftarrow \text{GetDuplicated}(\text{EEL});$
<b>3</b> EEL $\leftarrow$ EEL $- rEEL;$
4 $\text{EEL} \leftarrow \text{EEL} \cup \text{Ranking}(rEEL);$
<b>5</b> FDOE $\leftarrow$ GetFirst(EEL);
6 forall $eel \in \operatorname{EEL}$ do /* Iterate over named
entities */
7   forall $fe \in FDOE$ do /* Iterate over
filtered named entities */
$\mathbf{s} \qquad    \mathbf{if} \ eel.s \geqslant fe.s \ \mathbf{and} \ eel.s < fe.e \parallel$
9 $fe.s \ge eel.s$ and $fe.s < eel.e$ then
10   if $\text{LEN}(eel.M) > \text{LEN}(fe.M)$
then
11 FDOE $\leftarrow$ {FDOE $-fe$ } $\cup$ eel;
12 end
13 else
14 FDOE $\leftarrow$ FDOE $\cup eel;$
15 end
16 end
17 end

#### 3.3.2. Output preparation

Once the named entities are filtered, the next step is to prepare the final output according to the user or application needs. EEL is a key component of the Semantic Web, the output of the system should be presented as RDF (e.g., RDF/XML, Turtle, N-Quads). Thus, the output format selected is based on RDF Turtle<sup>15</sup>. Moreover, given that the configuration of the system is performed through the GERBIL framework, the entities must be annotated with the NIF vocabulary<sup>16</sup>.

## 4. Experiments

This section presents the experiments performed to test the proposed framework and its implementation. We divided the experiments into parameter tuning, performance evaluation, and comparison. The parameter tuning provides details of the experiments performed for tuning the confidence value of

<sup>&</sup>lt;sup>15</sup>https://www.w3.org/TR/turtle/

<sup>&</sup>lt;sup>16</sup>http://persistence.uni-leipzig.org/nlp2rdf/

each EEL system. Next, in the performance evaluation, FEELink is evaluated against the four EEL systems comprising it. Finally, it is presented a discussion of the FEELink results compared to a state of the art ensemble EEL system.

# 4.1. Settings

Details of the used dataset and environment are presented in this section.

Datasets. Seven gold-standard datasets (available in GERBIL) were selected for the experiments of FEELink: Derczynski [40], IITB [41], MSNBC [42], N3-Reuters-128 [43], ACE2004 [44], KORE50 [38], and OKE 2018 Task 2 [45]. The first four datasets are used for configuration and the last three for testing. The general intuition is to check the performance of FEELink with datasets not used in the tuning step. Additionally, two datasets were also used in the comparison: AQUAINT [46], and AIDA/CONLL B [35]. A brief description of the datasets is as follows:

- Derczynski. It consists of 182 English tweets extracted randomly from a streaming API with annotations about people, locations, and organizations (mainly).
- IITB. This dataset contains 107 English webpages with mixed domains related to sports, entertainment, science, technology, and health.
- MSNBC. This dataset contains 20 English news articles extracted from 10 different categories (e.g., politics, entertainment, sports).
- N3-Reuters-128. It contains 128 economic news articles (in English) extracted from the Reuters agency.
- ACE2004. This dataset contains 57 news articles with co-reference annotations in English.
- OKE 2018 Task2. It contains 56 documents in English with diverse topics related to sports, artwork, news, among others.
- KORE 50. This dataset is composed of 50 sentences in English from diverse domains such as music, celebrities, and business.
- AQUAINT. It contains 50 documents with news topics with annotations of common named entities (e.g., person, location, organization).

• AIDA/CONLL. It consists of 231 documents with topics related to news and sports.

The purpose of selecting such datasets is to configure and test FEELink over different topics. Note that NIL (Not In Lexicon) named entities are not considered for the experiments (those that are relevant but do not contain an association with a KB resource). A summary of features describing the datasets used for tuning and evaluation is presented in Table 2; where *Topic* denotes the domain or type of documents of the dataset, *Doc.* refers to the number of documents, *Ann.* is the number of named entities linked to a KB, and *Avg. En./Doc.* denotes the average number of named entities per document.

Dataset	Topic	Doc.	Ann.	Avg. En./Doc.
Derczynski	Tweets	182	286	1.57
IITB	Mixed	107	11250	109.22
MSNBC	News	20	747	37.35
N3-Reuters-128	News	128	621	4.85
KORE 50	Mixed	50	144	2.88
OKE	Mixed	56	423	7.55
ACE2004	News	57	306	5.37
AQUAINT	News	50	727	14.54
AIDA/CONLL	News	231	4485	19.41

Table 2: Description of the datasets used for parameter tuning and in the evaluation.

*Metrics.* To configure the confidence parameter of FEELink (and further experiments), micro and macro measures based on the traditional Information Extraction measures (e.g., precision, recall, F1) were used. Micro and macro measures take as base the traditional formula of precision or recall but, the parameters represent the combination of results from each document, or the entire dataset. Macro-precision considers the average precision over individual documents or sentences, while micro-precision considers the entire gold standard as one test without distinguishing the individual documents or sentences. The measures precision (P), recall (R) and their micro (mic) and macro (mac) versions are presented in Formulas 1 and 2, respectively.

$$P = \frac{tp}{tp + fp} \quad micP = \frac{\sum_{i}^{n} tp_{i}}{\sum_{i}^{n} (tp_{i} + fp_{i})} \quad (1)$$
$$macP = \frac{\sum_{i}^{n} P_{i}}{n}$$

$$R = \frac{tp}{tp + fn} \quad micR = \frac{\sum_{i}^{n} tp_{i}}{\sum_{i}^{n} (tp_{i} + fn_{i})},$$
$$macR = \frac{\sum_{i}^{n} R_{i}}{n}$$
(2)

where tp refers to true positive cases, fn to false negative cases and n to the number of documents/sentences.

Finally, the F1 measure was used for combining the values of precision and recall in one metric as presented in Formula 3, where micro and macro variants can also be used for obtaining it.

$$F1 = \frac{2PR}{P+R} \tag{3}$$

*Environment.* FEELink was evaluated through the GERBIL framework using the parameters presented in Section 3.1.3 (and Section 4.2 for the confidence configuration). The evaluation applies two entity matching configurations: strong and weak. For strong matching, a named entity is considered as correct if the extracted tuple (the mention and its URI) exactly matches with the tuple annotated in the dataset. On the other hand, for weak matching, the extracted tuple is considered as correct if its URI exactly matches the URI of the tuple in the dataset but the extracted mention partially or exactly matches with the mention in the dataset. Although the parameter tuning of the selected EEL systems was performed through experiments under a strong matching configuration (and thus the focus of the experiments), we also tested with a weak matching configuration in order to verify the behavior of FEELink under a less strict comparison.

#### 4.2. Parameter tuning

This subsection presents the experiments performed for the configuration of the confidence values described in Section 3.1.3. The selected EEL systems were executed over each of the four selected dataset for configuration, making successive variations of the confidence value in every execution (from 0 to 1 with intervals of 0.1). The result of the diverse executions is shown in Table 3, where the higher micro-F1 value for every system on every dataset is bolded. For example, according to the results of DBpedia Spotlight depicted in the Table 3, the range of confidence values that provided the highest micro-F1 values is from 0.4 (IITB dataset) to 0.9 (Derczynski). These values produced a median of 0.65, which was selected as the confidence value for such a system. The confidence configuration for the systems is summarized in Table 4, where *Exec.* refers to the type of system execution, *Conf.* to the confidence value (obtained by the median of confidence values from Table 3), and *Req. Key* indicates if the system requires a token-key. Note that such values were used for configuring the EEL systems in FEELink for the Performance evaluation of Section 4.3.

## 4.3. Performance evaluation

The purpose of these experiments is to evaluate the performance of FEELink (ensemble system) regarding the individual EEL systems of the ensemble.

System versions. The integration of named entities is provided mainly by the filtering process. Thus, we prepared two FEELink versions for the experiments. Considering the frequency (fr) of named entities extracted by the individual EEL systems, the first version is configured to filter entities with a frequency of one (fr1) and the second version with a frequency of two (fr2). Note that, confidence values from Table 4 are used for these experiments.

#### 4.3.1. Results

The results of the evaluation are shown in Tables 5 and 6 for the strong and weak matching configurations respectively. The scores for micro and macro variants for precision, recall, and F1 are presented for the individual EEL systems and the two system versions -FEELink(fr1) and FEELink(fr2)- over the seven datasets. Note that bold values indicate the best scores for the corresponding metric and dataset.

The results demonstrate the predominant performance of WAT, FEELink(fr1), and FEELink(fr2). It can be noticed that FEELink(fr2) outperforms all systems regarding the micro/macro versions of precision and partially regarding the F1. On the other hand, FEELink(fr1) demonstrated the best performance regarding the recall.

Regarding the experiments through a weak matching comparison (using the same system parameters), the results presented in Table 6 demonstrate an increase in the macro and micro measures concerning those shown in Table 5. This fact was expected due to the lenient matching evaluation of named entity mentions. Although

Micro-F1											
WAT											
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Derczynski	0.309	0.315	0.305	0.291	0.257	0.227	0.163	0.143	0.097	0.041	0
IITB	0.135	0.133	0.124	0.115	0.107	0.093	0.081	0.066	0.049	0.023	0
MSNBC	0.595	0.607	0.598	0.567	0.539	0.523	0.480	0.425	0.336	0.159	0
N3-Reuters-128	0.383	0.392	0.390	0.368	0.350	0.318	0.286	0.225	0.170	0.075	0
				DBpedi	aSpotligs	ht					
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Derczynski	0.126	0.126	0.127	0.129	0.254	0.271	0.294	0.292	0.305	0.305	0.161
IITB	0.240	0.240	0.242	0.242	0.258	0.227	0.209	0.198	0.183	0.172	0.061
MSNBC	0.153	0.153	0.152	0.154	0.345	0.381	0.388	0.402	0.406	0.412	0.181
N3-Reuters-128	0.069	0.069	0.070	0.071	0.179	0.216	0.244	0.260	0.261	0.264	0.088
				Ta	igMe						
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Derczynski	0.124	0.226	0.268	0.288	0.254	0.255	0.120	0.007	0	0	0
IITB	0.219	0.243	0.180	0.129	0.092	0.063	0.029	0.004	0	0	0
MSNBC	0.168	0.310	0.381	0.354	0.296	0.254	0.116	0.019	0	0	0
N3-Reuters-128	0.090	0.152	0.254	0.257	0.207	0.147	0.026	0.000	0	0	0
				Ba	belfy						
	0	0.1	0.2	0.3	0.4	<b>0.5</b>	0.6	0.7	0.8	0.9	1.0
Derczynski	0.238	0.253	0.213	0.226	0.151	0.155	0.053	0.064	0.047	0.040	0
IITB	0.093	0.075	0.033	0.029	0.016	0.015	0.004	0.007	0.004	0.002	0
MSNBC	0.334	0.357	0.196	0.119	0.097	0.062	0.046	0.046	0.041	0.041	0
N3-Reuters-128	0.207	0.252	0.115	0.089	0.081	0.074	0.020	0.059	0.036	0.018	0

Table 3: Micro-F1 results for configuring the EEL systems over Derczynsky, IITB, MSNBC and N3-Reuters-128 datasets.

EEL E System	xec.	Invocation	Conf.	Req. Key.?
TagMeRBabelfyRWATRDBpediaI	emote emote emote	Post Post Post Local,	0.2 0.1 0.1	Yes Yes Yes

Table 4: Configuration of parameters for the selected EEL systems.

the results are slightly better than those presented for a strong matching comparison, similar behavior is shown in favor of the micro/macro- precision and recall scores presented by FEELink(fr1)and FEELink(fr2) respectively. In consequence, all the results provided by FEELink(fr1) and FEELink(fr2) for strong and weak configurations, outperformed those provided by individual EEL systems regarding the micro/macro- recall and precision respectively. Although FEELink(fr2) got a slight reduction in the micro-recall (with respect to FEELink(fr1)), it is still better than the individual performance of systems for almost all the tested datasets. On the other hand, FEELink(fr2) outperformed the micro/macro-F1 presented by WAT on the Derczynski, IITB, OKE, and ACE2004 datasets.

One aspect involved in the performance showed by FEELink is related to the number of extracted named entities. In this sense, a comparison of elements extracted between the best individual EEL system in the above experiments for strong matching (WAT, according to the micro-F1 experiments presented in Table 5) and the two versions of FEELink (fr1 and fr2) is shown in Table 7, where E.E. = Extracted Entities, tp = true positive, fp =false positive, and fn = false negative extractions. Gray level background indicates the first, second and third higher counts of tp per column. The † and  $\ddagger$  symbols indicate the highest values for fn and fp columns respectively. In this sense, given that FEELink(fr1) accepts named entities that were extracted by at least one EEL system, it extracted the highest number of named entities (E.E.), but it also got the highest number of false positives (fp) in comparison to FEELink(fr2) and WAT. Hence, the results negatively affect the micro/macro-precision of FEELink. On the other hand, WAT extracted the lowest number of named entities and the highest number of false negatives (fn) on almost all the tested datasets, which negatively impacts its micro/macro-recall.

	Micro/Macro-precision									
	WAT	Spotlight	TagMe	Babelfy	FEELink(fr1)	FEELink(fr2)				
Derczynski	0.354/0.277	0.302/0.305	0.164/0.180	0.285/0.268	0.199/0.228	0.399/0.379				
IITB	0.459/0.415	0.551/0.513	0.409/0.359	0.481/0.237	0.469/0.469	0.644/0.626				
MSNBC	0.616/0.554	0.449/0.416	0.242/0.265	0.510/0.252	0.424/0.420	0.639/0.634				
N3-Reuters-128	0.377/0.332	0.283/0.215	0.104/0.106	0.337/0.139	0.231/0.208	0.409/0.366				
KORE50	0.563/0.562	0.508/0.348	0.476/0.300	0.500/0.242	0.482/0.486	0.626/0.510				
OKE	0.481/0.491	0.438/0.425	0.493/0.493	0.541/0.539	0.389/0.400	0.554/0.563				
ACE2004	0.091/0.109	0.088/0.099	0.094/0.128	0.124/0.139	0.072/0.089	0.118/0.155				
		Mi	icro/Macro-r	ecall						
	WAT	Spotlight	TagMe	Babelfy	FEELink(fr1)	FEELink(fr2)				
Derczynski	0.283/0.284	0.283/0.327	0.364/0.343	0.227/0.294	0.423/0.377	0.385/0.397				
IITB	0.078/0.097	0.121/0.122	0.173/0.168	0.041/0.066	0.188/0.197	0.112/0.122				
MSNBC	0.598/0.583	0.364/0.392	0.430/0.450	0.274/0.301	0.616/0.599	0.515/0.512				
N3-Reuters-128	0.408/0.410	0.241/0.223	0.285/0.288	0.201/0.185	<b>0.418</b> /0.399	0.349/0.337				
KORE50	0.528/0.498	0.222/0.199	0.278/0.251	0.1667/0.138	0.563/0.532	0.431/0.386				
OKE	0.426/0.417	0.348/0.324	0.312/0.305	0.284/0.294	0.466/0.464	0.409/0.417				
ACE2004	0.513/0.319	0.458/0.299	0.275/0.257	0.386/0.256	<b>0.546</b> /0.356	0.533/ <b>0.359</b>				
		Ν	/licro/Macro	-F1						
	WAT	Spotlight	TagMe	Babelfy	FEELink(fr1)	FEELink(fr2)				
Derczynski	0.315/0.274	0.292/0.302	0.226/0.221	0.253/0.273	0.271/0.268	0.392/0.375				
IITB	0.133/0.149	0.198/0.196	0.243/0.217	0.075/0.081	0.268/0.269	0.191/0.197				
MSNBC	0.607/0.566	0.402/0.405	0.310/0.304	0.357/0.325	0.503/0.484	0.570/0.548				
N3-Reuters-128	0.392/0.355	0.260/0.219	0.152/0.145	0.252/0.178	0.298/0.263	0.377/0.326				
KORE50	0.545/0.519	0.309/0.242	0.351/0.262	0.250/0.168	0.519/0.499	0.510/0.424				
OKE	0.452/0.436	0.387/0.355	0.382/0.364	0.372/0.358	0.424/0.419	0.471/0.467				
ACE2004	0.155/0.152	0.148/0.139	0.141/0.147	0.187/0.166	0.127/0.133	0.194/0.193				

Table 5: Results of micro/macro measures under **strong matching**. The first column corresponds to the dataset, the following four columns indicate individual EEL systems, and the last two columns indicate the two FEELink versions.

3 4 1	/ 7. /	
Where	/ Wacro-	precision
1111010	/ macro	procibion

	WAT	Spotlight	TaaMe	Babelfu	<b>FEELink</b> (fr1)	<b>FEELink</b> (fr2)
Derczynski	0.389/0.299	0.343/0.317	0.272/0.283	0.325/0.301	0.215/0.232	0.442/0.412
IITB	0.525/0.529	0.597/0.511	0.633/0.518	0.527/0.369	0.491/0.493	0.688/0.679
MSNBC	0.709/0.685	0.509/0.493	0.478/0.543	0.542/0.439	0.423/0.424	0.668/0.665
N3-Reuters-128	0.459/0.417	0.357/0.310	0.326/0.316	0.423/0.292	0.256/0.227	0.481/0.447
KORE50	0.644/0.641	0.524/0.368	0.488/0.310	0.500/0.242	0.536/0.542	<b>0.647</b> /0.537
OKE	0.591/0.589	0.515/0.509	0.582/0.569	0.667/0.630	0.476/0.484	0.651 / 0.654
ACE2004	0.115/0.129	0.096/0.106	0.102/0.139	0.124/0.139	0.092/0.111	0.125/0.162
	,	Mie	cro/Macro-re	ecall		·
	WAT	Spotlight	TagMe	Babelfy	FEELink(fr1)	FEELink(fr2)
Derczynski	0.311/0.280	0.290/0.336	0.353/0.356	0.259/0.314	0.521/0.444	0.427/0.424
IITB	0.089/0.107	0.110/0.112	0.116/0.117	0.045/0.054	0.196/0.206	0.119/0.130
MSNBC	0.689/0.668	0.400/0.414	0.366/0.393	0.292/0.309	0.776/0.778	0.539/0.536
N3-Reuters-128	0.497/0.477	0.269/0.294	0.313/0.331	0.252/0.252	0.567/0.549	0.409/0.410
KORE50	0.604/0.574	0.229/0.206	0.285/0.258	0.167/0.138	0.625/0.593	0.444/0.397
OKE	0.523/0.507	0.409/0.393	0.369/0.361	0.349/0.354	0.569/0.562	0.479/0.482
ACE2004	0.644/0.386	0.497/0.326	0.297/0.279	0.386/0.256	0.699/0.440	0.562/0.378
		Μ	licro/Macro-	F1		
	WAT	Spotlight	TagMe	Babelfy	FEELink(fr1)	FEELink(fr2)
Derczynski	0.346/0.282	0.314/0.314	0.307/0.295	0.288/0.293	0.304/0.286	0.434/0.404
IITB	0.153/0.169	0.186/0.179	0.195/0.182	0.083/0.088	0.280/0.282	0.204/0.211
MSNBC	0.699/0.671	0.448/0.432	0.414/0.409	0.380/0.347	0.548/0.537	0.597/0.574
N3-Reuters-128	0.477/0.430	0.307/0.276	0.319/0.301	0.316/0.241	0.352/0.309	0.442/0.399
KORE50	0.624/0.597	0.319/0.252	0.359/0.269	0.250/0.168	0.577/0.558	0.527/0.439
OKE	<b>0.555</b> /0.529	0.456/0.427	0.452/0.424	0.459/0.428	0.519/0.509	0.552/0.538
ACE2004	0.195/0.181	0.161/0.150	0.152/0.162	0.187/0.166	0.162/0.166	0.204/0.201

Table 6: Results of micro/macro measures under weak matching. Columns are described in the same way as in Table 5.

	WAT					$\mathbf{FEELink}(fr1)$			FEELink( <i>fr</i> 2)			
	E.E.	$^{\mathrm{tp}}$	$\mathbf{fn}$	fp	E.E.	$^{\mathrm{tp}}$	$\mathbf{fn}$	fp	E.E.	$^{\mathrm{tp}}$	fn	fp
Derczynski	434	81	$^{+205}$	148	771	121	165	$^{1485}$	452	110	176	166
IITB	19989	1427	$^{\dagger 16881}$	1681	21513	3257	15051	\$3205	19441	2047	16261	1133
MSNBC	1026	447	300	279	1371	460	287	$\ddagger624$	965	385	+362	218
N3-Reuters-	1474	359	521	594	2105	368	512	±1225	1323	306	+574	443
128	1111	000	021	004	2100	000	012	+1220	1020	000	1014	110
KORE50	203	76	$^{+68}$	59	231	81	63	$^{+87}$	157	86	58	13
OKE	617	180	$^{+243}$	194	732	197	226	$^{309}$	524	211	212	101
ACE2004	1870	156	$^{+150}$	1564	2477	169	140	$\ddagger 2168$	1521	163	143	1215

Table 7: Comparison between the best individual system (WAT) and FEELink versions (fr1, fr2) considering the number of discovered entities.

Table 8: Pearson correlation values of the EEL systems performance (micro-F1) and the named entities per class.

	Babelfy	FEELink(fr1)	FEELink( <i>fr</i> 2)	$\mathbf{Spotlight}$	$\mathbf{TagMe}$	WAT
persons	0.481	0.593	0.610	0.510	0.654	0.639
organizations	0.016	0.532	0.268	0.376	0.412	0.260
locations	0.451	0.029	0.270	0.049	0.145	0.319
others	0.485	0.040	0.325	0.113	0.218	0.367

## 4.3.2. Qualitative analysis

In order to test the performance of FEELink regarding the type of extractions, we performed an analysis of the Pearson correlation between four types of named entities (persons, locations, organizations, and others) with respect to the micro-F1 performance of individual EEL systems and FEELink (fr1, fr2). The correlation values were obtained through the GERBIL framework, which takes into account the seven datasets used in the evaluation (under strong matching). The results are shown in Table 8. The different gray level background indicates the top three values. The best value in the class of persons was provided by TagMe, for organizations by FEELink(fr1), for locations by Babelfy, and for others by Babelfy. It can be noticed that FEELink(fr2) obtained a competitive performance because it obtained a balance in the distribution of named entities by class. We think this can be useful in situations where the preponderance of classes by entity is not known in advance. On the other hand, if such fact is known at some level, then particular systems could be better suited.

## 4.3.3. Discussion

As presented in the previous section, an implementation of the proposed framework was created (FEELink) as a proof of concept. However, many possible variants can be involved in such implementation. Thus, some aspects regarding the configuration and testing are presented as follows.

*Configuration.* For the configuration of an implemented system, different other domains and EEL systems can be used. Thus, the following aspects can be taken into account:

- Parameter tuning. The configuration is usually performed according to particular EEL systems and domains. Although varied experiments were performed for the selection of the confidence values, it should be noted that the experiments can be stopped when the used (interval) confidence value tends to decrease the (micro-F1) score for some configurations. For example, according to the results presented in Table 3, TagMe does not provide better results after the confidence value is greater than 0.5 for the Derczynski dataset and thus, for this particular case, experiments with posterior confidence values could be omitted.
- Trade-off measure. In this particular case, the micro-F1 measure was selected to find a tradeoff between precision and recall. However, users may find useful to select a distinct metric (precision or recall), according to its requirements about domain, dataset, tagged named entitites, and balance of types of named entities.
- Default configurations. EEL systems often provide default configurations. So that the

users can use them without any further adaptation. For example, DBpedia Spotlight [5] points a reference confidence value of 0.6 (as the best value for its experiments). In this sense, we got a slightly similar value (0.65) as appropriate. Similarly, although WAT and TagMe suggest confidence values, after our experiments we got specific thresholds. We continued the parameter tuning of such systems to obtain the specific values of configuration because the confidence values are sensitive to variations, where specific values are not always applicable to any dataset.

- EEL service availability. EEL systems may not always be available (e.g., the web service is under maintenance). Thus, users must be aware regarding this situation at the selection of EEL systems.
- Extensibility. Although a specific number of EEL systems is not defined in the proposed framework, the performance of the complete system may be decreased due to hardware limitations. Thus, a higher number of systems would require a modular architecture, which is out of the scope of this work.
- Number of systems. Although it is supposed that a greater number of systems (used for the EEL ensemble) would provide better extractions, the complexity also increases regarding all the above aspects. Thus, defining a specific number of EEL systems depends on the requirements of users, tasks, configuration tuning, response time, and availability of resources (hardware and software).

Testing. The results presented in Tables 5 and 6 for the FEELink evaluation (fr1 and fr2 versions)demonstrate an increased performance concerning micro-macro precision and recall (and F1 in some cases) in comparison with individual EEL systems. As presented in Table 7, the highest micro-recall for FEELink(fr1) was produced due to the greater number of extracted named entities at the cost of penalizing the micro-precision (high number of false positive extractions). On the other hand, FEELink(fr2) outperformed all individual systems regarding the micro-precision because it takes into account the fact that one entity must be identified by at least two EEL systems. Although this fact slightly reduced the recall, FEELink(fr2) got encouraging micro-F1 results in comparison to WAT. From this evaluation, some aspects of testing can be noted:

- Frequency configuration. Although we could configure a frequency value between one and the max number of EEL systems in the ensemble, we only tested two frequency values in the evaluation. This is because we noted that by increasing the frequency, the precision was improved at the cost of the recall. Thus, we kept those frequency values to maintain a trade-off between such measures.
- Outdated datasets. Although GERBIL contains a dictionary to map equivalences (owl:sameAs) between entities (e.g., equivalent entities from YAGO and DBpedia), sometimes the KBs change the identifiers of their resources, which is reflected in parsing errors [47].
- No annotations. False-positive cases are often obtained by systems because of the absence of annotations in the benchmark datasets. This may happen due to the previous aspect and/or to updated dictionaries used by the systems.
- Black box evaluation. To the best of our knowledge, a more fine grained evaluation has not been proposed to measure internal aspects of EEL systems such as the consumption of resources (RAM, CPU), time, complexity, and the appropriate domain. Thus, this kind of details would be useful for defining the number of systems that can be integrated into an ensemble approach.
- Correlation of annotators and features. The integration of EEL systems may suppose an improvement in detecting individual entity types (displayed by the correlation analysis). However, different factors such as metrics, datasets, and systems are involved in the analysis. For example, the Pearson correlation values in Table 8 consider the micro-F1 for the seven datasets in the experiments. Thus, the *true positive*, *false negative*, and *false positive* values play a key factor in determining such values. Therefore, integrating systems not performing well for certain entity types (slightly) degrades the correlation values of FEELink but

a balance is achieved in the distribution of recognized named entities by type.

• Filtering. We decided to implement a frequency-based/longest-matching solution to filter overlapped entities. Although it provides encouraging precision values, such a solution might suffer the problem of avoiding correct entities or including incorrect entities. Thus, the overlapping is still an open problem in ensemble EEL systems, where more filtering criteria functions need to be proposed.

## 4.4. Comparison

As presented in the Related Work (Section 2), a similar approach that provides an integration of EEL systems was proposed by Ruiz and Poibeau [32] (denoted as RP). This section provides a comparison between the features used by such an approach and FEELink. Although the main purpose of our work is to describe the proposed framework, this comparison aims to analyze the performance (according to recommendations of the proposed framework) of FEELink against a state of the art ensemble EEL system.

## 4.4.1. Settings

Details of the used datasets and environment are presented in this section.

*Datasets.* Four datasets were used in this comparative: IITB [41], MSNBC [42], AQUAINT [46], and AIDA/CONLL B [35].

Environment. We configured FEELink using the fr2 version (described in Section 4.3) and the evaluation was performed through the GERBIL framework. However, implementing the approach of RP [32] is complicated due to the availability of the used systems and configuration details. Thus, the configuration details are as follows:

- Systems. FEELink integrates four EEL systems (described in Section 3.1.1). On the other hand, RP integrates TagMe, DBpedia Spotlight, Wikipedia Miner [46], AIDA [35], and Babelfy.
- Evaluation. While FEELink uses GERBIL, RP uses a combination of BAT [34] and nelEval<sup>17</sup> for configuration and testing respectively.

Moreover, a strong matching comparison was performed for both systems.

- Filtering. FEELink uses a filtering strategy based on the frequency of named entities and ranking of systems. RP implements a weighted function that takes into account the ranking and precision of systems to filter overlapped entities.
- Confidence values. FEELink uses a strategy to select the confidence values (based on the median of best F1 scores per dataset and system) using the GERBIL framework. Similarly, RP applies evaluations (through BAT) for configuring their filtering function. However, they use specific values of confidence for each dataset at testing.
- Datasets. The parameters were tuned considering four datasets for FEELink and two datasets for RP.

A summary of the features used by the compared systems is shown in Table 9.

Table 9: Comparison of features used by FEELink and Ruiz and Poibeau  $[32]~(\rm RP)$ 

Feature	FEELink	RP
Systems	4	5
Evaluation	GERBIL	nelEval/ BAT
Filtering	frequency/ ranking	weighted/ ranking
Confidence Datasets	median 4	per dataset 2

# 4.4.2. Results

The results of the comparison are shown in Table 10, where P the micro-precision, R the micro-Recall, and F1 the micro-F1 measures. It is worth mentioning that the values of RP were obtained from the article of Ruiz and Poibeau [32]. Best result values for each dataset are marked in bold.

Table 10: Results for the comparison of FEELink and RP

		RP		FEELink			
	Р	R	$\mathbf{F1}$	Р	R	$\mathbf{F1}$	
ПТВ	0.593	0.447	0.500	0.644	0.112	0.191	
MSNBC	0.543	0.434	0.482	0.639	0.515	0.570	
AQUAINT	0.341	0.641	0.445	0.379	0.397	0.388	
AIDA CONLLB	0.648	0.617	0.619	0.662	0.542	0.596	

<sup>&</sup>lt;sup>17</sup>https://github.com/wikilinks/neleval/wiki

The results demonstrated that FEELink got the best micro-precision values for all datasets. Although the conditions for the comparison are different for the configuration of systems, FEELink demonstrated competitive performance in comparison with RP, outperforming the micro-F1 results in the MSNBC dataset.

# 4.4.3. Discussion

Making an evaluation against third party ensemble-systems is not a straightforward task because of diverse aspects such as implementation details, purpose, availability of EEL systems, and hardware/software configurations. First, details about the implementation of some ensemble systems are not always provided, where the same testing environment is not easily reproducible. Second, the purpose of the ensemble systems is not the same. For example, while FEELink focuses on the extraction of named entity tuples, studies such as NERD [13] are based on the extraction and typing of named entities. Third, configuration parameters and EEL systems (e.g., by approaches such as NERD [13] and RP) are not always available. Thus, replicating the same output becomes a complicated task. However, we compared FEELink against the approach presented by RP. Although FEELink uses one less EEL system than RP, and a filtering function based on frequency, it demonstrated competitive results regarding the micro-precision measure.

# 5. Conclusions

This paper presents a framework to integrate the output provided by EEL systems. The proposed framework begins with the selection of EEL systems until the extraction and presentation of named entity results. The framework is composed of three general stages: Parameter Configuration, System Invocation, and Data Consolidation. The first stage refers to the selection of the EEL systems (whose results will be integrated), the management of hardware and software resources (to invocate such systems) and their input parameters configuration. The second stage refers to the invocation and collection of outputs by the systems. Finally, the last stage is in charge of the unification of such outputs through a filtering strategy that takes into account duplicated and/or overlapped named entity results.

The proposed framework was tested through an implementation (called FEELink) that integrates

the output of four different state of the art EEL systems and seven gold-standard datasets. The results demonstrate that the implemented system improves the performance of individual EEL systems regarding the micro-precision, micro-recall and, in some cases, regarding the micro-F1. Moreover, FEELink demonstrated competitive performance regarding the micro-precision and micro-F1 compared with the results presented by an ensemble EEL system from the state of the art.

The EEL is a very active research task where varied approaches and techniques are periodically proposed to improve the system's performance. Although the precision (and number of extractions) provided by the EEL systems is not ideal, a system based on the proposed framework would be useful for supporting tasks such as question answering [48], semantic annotation [3], and information retrieval [49], to mention a few.

*Future work.* We plan to test a system implementation with diverse domains, languages, and additional EEL systems. Moreover, we plan to define a strategy for the selection of EEL systems based on a characterization of internal features and an evaluation scheme to measure the correlation of systems.

Acknowledgments. This work was funded in part by SEP-Cinvestav project number 229. We would like to thank Michael Röder for his help with GER-BIL.

# References

- G. Antoniou, F. van Hermelen, A Semantic Web Primer, Second Edition, The MIT Press, 2008.
- [2] S. Auer, V. Bryl, S. Tramp (Eds.), Linked Open Data -Creating Knowledge Out of Interlinked Data - Results of the LOD2 Project, Vol. 8661 of Lecture Notes in Computer Science, Springer, 2014.
- [3] M. Chabchoub, M. Gagnon, A. Zouaq, Collective disambiguation and semantic annotation for entity linking and typing, in: Semantic Web Challenges: Third SemWebEval Challenge at ESWC 2016, Springer International Publishing, 2016, pp. 33–47.
- [4] L. Haidar-Ahmad, L. Font, A. Zouaq, M. Gagnon, Entity typing and linking using SPARQL patterns and dbpedia, in: ESWC, Springer, 2016, pp. 61–75.
- [5] P. N. Mendes, M. Jakob, A. García-Silva, C. Bizer, DBpedia Spotlight: Shedding Light on the Web of Documents, in: International Conference on Semantic Systems, ACM, 2011, pp. 1–8.
- [6] F. Piccinno, P. Ferragina, From tagme to wat: A new entity annotator, in: Proceedings of the First International Workshop on Entity Recognition & Disambiguation, ACM, 2014, pp. 55–62.

- [7] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, C. Bizer, Dbpedia - A large-scale, multilingual knowledge base extracted from wikipedia, Semantic Web 6 (2) (2015) 167–195.
- [8] J. Hoffart, F. M. Suchanek, K. Berberich, G. Weikum, YAGO2: A spatially and temporally enhanced knowledge base from wikipedia, Artif. Intell. 194 (2013) 28– 61.
- [9] R. Grishman, B. Sundheim, Message understanding conference- 6: A brief history, in: 16th International Conference on Computational Linguistics, Proceedings of the Conference, COLING 1996, Center for Sprogteknologi, Copenhagen, Denmark, August 5-9, 1996, 1996, pp. 466–471.
- [10] H. Rosales-Méndez, B. Poblete, A. Hogan, What should entity linking link?, in: Proceedings of the 12th Alberto Mendelzon International Workshop on Foundations of Data Management, Cali, Colombia, May 21-25, 2018., CEUR-WS.org, 2018, pp. 1–5.
- [11] D. Ceccarelli, C. Lucchese, S. Orlando, R. Perego, S. Trani, Dexter 2.0 - an open source tool for semantically enriching data, in: ISWC-PD, 2014, pp. 417–420.
- J. Plu, G. Rizzo, R. Troncy, Enhancing entity linking by combining NER models, in: Semantic Web Challenges
  Third SemWebEval Challenge at ESWC, Springer, 2016, pp. 17–32.
- [13] G. Rizzo, R. Troncy, NERD: A framework for unifying named entity recognition and disambiguation extraction tools, in: European Chapter of the (ACL), The Association for Computer Linguistics, 2012, pp. 73–76.
- [14] P. Ferragina, U. Scaiella, Tagme: On-the-fly annotation of short text fragments (by Wikipedia entities), in: Information and Knowledge Management (CIKM), ACM, 2010, pp. 1625–1628.
- [15] 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.
- [16] M. Röder, R. Usbeck, A. N. Ngomo, GERBIL benchmarking named entity recognition and linking consistently, Semantic Web 9 (5) (2018) 605–625.
- [17] J. L. Martinez-Rodriguez, A. Hogan, I. Lopez-Arevalo, Information extraction meets the semantic web: A survey, Semantic Web 11 (2) (2020) 255–335. doi:10.3233/SW-180333.
- [18] S. A. Salloum, M. Al-Emran, A. A. Monem, K. Shaalan, A survey of text mining in social media: facebook and twitter perspectives, Advances in Science, Technology and Engineering Systems Journal.
- [19] S. Perera, P. N. Mendes, A. Alex, A. P. Sheth, K. Thirunarayan, Implicit entity linking in tweets, in: ESWC, Springer, 2016, pp. 118–132.
- [20] H. Rosales-Méndez, B. Poblete, A. Hogan, Multilingual entity linking: Comparing english and spanish, in: Int. Workshop on Linked Data for Information Extraction, (ISWC 2017)., CEUR-WS.org, 2017, pp. 62–73.
- [21] J. D'Souza, V. Ng, Sieve-Based Entity Linking for the Biomedical Domain, in: Association for Computational Linguistics: Short Papers, ACL, 2015, pp. 297–302.
- [22] J. Zheng, D. Howsmon, B. Zhang, J. Hahn, D. L. McGuinness, J. A. Hendler, H. Ji, Entity linking for biomedical literature, BMC Med. Inf. & Decision Making 15 (S-1) (2015) S4.
- [23] L. E. Anke, S. Oramas, H. Saggion, X. Serra, Elmdist:

A vector space model with words and musicbrainz entities, in: The Semantic Web: ESWC 2017, Springer, 2017, pp. 355–366.

- [24] P. Pantel, A. Fuxman, Jigs and lures: Associating web queries with structured entities, in: Annual Meeting of the ACL: Human Language Technologies, ACL, 2011, pp. 83–92.
- [25] A. Gangemi, A comparison of knowledge extraction tools for the semantic web, in: ESWC, Springer, 2013, pp. 351–366.
- [26] B. Carpenter, B. Baldwin, Text Analysis with LingPipe 4, LingPipe Publishing, 2011. URL http://alias-i.com/lingpipe-book/lingpipebook-0.5.pdf
- [27] T. G. Dietterich, Ensemble methods in machine learning, in: Multiple Classifier Systems, First International Workshop, MCS 2000, Cagliari, Italy, June 21-23, 2000, Proceedings, Springer, 2000, pp. 1–15.
- [28] Z.-H. Zhou, Ensemble Methods: Foundations and Algorithms, Chapman and Hall/CRC, 2012.
- [29] R. Speck, A. N. Ngomo, Ensemble learning for named entity recognition, in: International Semantic Web Conference, Springer, 2014, pp. 519–534.
- [30] Z. Zuo, G. Kasneci, T. Grütze, F. Naumann, BEL: bagging for entity linking, in: International Conference on Computational Linguistics (COLING), ACL, 2014, pp. 2075–2086.
- [31] Y. Chiu, Y. Shih, Y. Lee, C. Shao, M. Cai, S. Wei, H. Chen, NTUNLP approaches to recognizing and disambiguating entities in long and short text at the ERD challenge 2014, in: International Workshop on Entity Recognition & Disambiguation (ERD), ACM, 2014, pp. 3–12.
- [32] P. Ruiz, T. Poibeau, Combining open source annotators for entity linking through weighted voting, in: Lexical and Computational Semantics, 2015, pp. 211–215.
- [33] K. D. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor, Freebase: a collaboratively created graph database for structuring human knowledge, in: International Conference on Management of Data, ACM, 2008, pp. 1247–1250.
- [34] M. Cornolti, P. Ferragina, M. Ciaramita, A framework for benchmarking entity-annotation systems, in: World Wide Web Conference, ACM, 2013, pp. 249–260.
- [35] J. Hoffart, M. A. Yosef, I. Bordino, H. Fürstenau, M. Pinkal, M. Spaniol, B. Taneva, S. Thater, G. Weikum, Robust disambiguation of named entities in text, in: Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2011, pp. 782–792.
- [36] R. Usbeck, A.-C. N. Ngomo, M. Röder, D. Gerber, S. A. Coelho, S. Auer, A. Both, AGDISTIS – Graph-based disambiguation of Named Entities using Linked Data, in: International Semantic Web Conference (ISWC), Springer, 2014, pp. 457–471.
- [37] A. Södergren, P. Nugues, A multilingual entity linker using pagerank and semantic graphs, in: Nordic Conference on Computational Linguistics, NODALIDA, Association for Computational Linguistics, 2017, pp. 87– 95.
- [38] J. Hoffart, S. Seufert, D. B. Nguyen, M. Theobald, G. Weikum, KORE: keyphrase overlap relatedness for entity disambiguation, in: International Conference on Information and Knowledge Management, ACM, 2012, pp. 545–554.

- [39] J. R. Finkel, C. D. Manning, Nested named entity recognition, in: Empirical Methods in Natural Language Processing (EMNLP), ACL, 2009, pp. 141–150.
- [40] 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.
- [41] S. Kulkarni, A. Singh, G. Ramakrishnan, S. Chakrabarti, Collective annotation of wikipedia entities in web text, in: International Conference on Knowledge Discovery and Data Mining, ACM, 2009, pp. 457–466.
- [42] S. Cucerzan, Large-scale named entity disambiguation based on wikipedia data, in: Empirical Methods in Natural Language Processing and Computational Natural Language Learning, ACL, 2007, pp. 708–716.
- [43] M. Röder, R. Usbeck, S. Hellmann, D. Gerber, A. Both, N<sup>3</sup> - A collection of datasets for named entity recognition and disambiguation in the NLP interchange format, in: Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014., (ELRA), 2014, pp. 3529–3533.
- [44] L. Ratinov, D. Roth, D. Downey, M. Anderson, Local and global algorithms for disambiguation to wikipedia, in: The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA, The Association for Computer Linguistics, 2011, pp. 1375–1384.
- [45] R. Speck, M. Röder, F. Conrads, H. Rebba, C. C. Romiyo, G. Salakki, R. Suryawanshi, D. Ahmed, N. Srivastava, M. Mahajan, A. N. Ngomo, Open knowledge extraction challenge 2018, in: Semantic Web Challenges - 5th SemWebEval Challenge at ESWC 2018, Heraklion, Greece, June 3-7, 2018, Revised Selected Papers, Springer, 2018, pp. 39–51.
- [46] D. N. Milne, I. H. Witten, Learning to link with wikipedia, in: Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, USA, October 26-30, 2008, ACM, 2008, pp. 509–518.
- [47] K. Jha, M. Röder, A. N. Ngomo, All that glitters is not gold - rule-based curation of reference datasets for named entity recognition and entity linking, in: ESWC, 2017, pp. 305–320.
- [48] C. Unger, A. Freitas, P. Cimiano, An introduction to question answering over linked data, in: Reasoning on the Web in the Big Data Era, Springer, 2014, pp. 100– 140.
- [49] J. Waitelonis, C. Exeler, H. Sack, Linked data enabled generalized vector space model to improve document retrieval, in: NLP & DBpedia Workshop in conjunction with ISWC 2015, CEUR, 2015, pp. 1–12.