Knowledge and Information Systems
https://doi.org/10.1007/s10115-023-02059-2

SURVEY PAPER

Entity linking for English and other languages: a survey

Imane Guellil[1] · Antonio Garcia-Dominguez[2] · Peter R. Lewis[3] · Shakeel Hussain[4] · Geoffrey Smith[4]

Received: 9 September 2021 / Revised: 23 December 2023 / Accepted: 25 December 2023

Abstract
Extracting named entities text forms the basis for many crucial tasks such as information retrieval and extraction, machine translation, opinion mining, sentiment analysis and question answering. This paper presents a survey of the research literature on named entity linking, including named entity recognition and disambiguation. We present 200 works by focusing on 43 papers (5 surveys and 38 research works). We also describe and classify 56 resources, including 25 tools and 31 corpora. We focus on the most recent papers, where more than 95% of the described research works are after 2015. To show the efficiency of our construction methodology and the importance of this state of the art, we compare it to other surveys presented in the research literature, which were based on different criteria (such as the domain, novelty and presented models and resources). We also present a set of open issues (including the dominance of the English language in the proposed studies and the frequent use of NER rather than the end-to-end systems proposing NED and EL) related to entity linking based on the research questions that this survey aims to answer.

Keywords Entity linking · Named entity recognition · Named entity disambiguation · English entity linking approaches · Multilingual entity linking approaches

1 Introduction

The huge volume of data exchanged every day via social media, text messaging and chat services has led to an increase in the number of works on natural language processing (NLP) during the last decades. Different NLP areas require extracting meaningful information from the text (which is unstructured in the majority of cases) automatically and quickly. Some of the most important and studied areas of NLP are information extraction (IE), machine translation

✉ Imane Guellil
imane.guellil@ed.ac.uk

1 University of Edinburgh, Edinburgh, UK

2 University of York, York, UK

3 Ontario Tech University, Oshawa, ON, Canada

4 Folding Space, Birmingham, UK

(MT), opinion mining (OM) and question-answering (QA). These areas share one important aspect: they all need to identify proper nouns and classify them into the appropriate type of named entity [63, 136]. Named entities could be the names of persons (such as David, or Satoshi), locations (e.g. Tokyo, or Canada), or organisations (e.g. Stanford University, or Amazon). For example, let us consider this sentence: *"Mr Brown is living in California, and he is working at Amazon"*. A **named entity recognition** (NER) system should recognise *"Brown"* as the name of a person, *"California"* as the name of a location and *"Amazon"* as the name of an organisation. Based on this example, it seems that a simple dictionary of names combined with a set of regular expressions could solve this problem.

However, NER is not as simple as it appears. For instance, if we had *"Mr. Brown is living in California, and he is working at Amazon. He would like to buy a brown jacket. He also would like to visit the Amazon rainforest"*; it can be seen that both *"Brown"* and *"Amazon"* appear in two circumstances. *"Brown"* appears as the name of a person and as a colour (which is not a named entity). *"Amazon"* appears as the name of an organisation and the name of a location. These are evident ambiguities in the system. Hence, a simple system relying on dictionaries or regular expressions would fail in recognising the correct named entities.

Solving this problem involves another research area: **named entity disambiguation** (NED). *"Mr. Brown"* could be *Chris Brown* the singer or *Joseph Brown*, a Systems Development Engineer at Amazon. To answer this question, we need to link *"Brown"* to a knowledge base and select the best candidate based on the context where this entity appears. In this case, the sub-sentence *"is working at Amazon"* is crucial for affirming that *"Mr. Brown"* is an engineer from Amazon and not a famous singer. Also, due to these entities, *"Amazon"* is easily recognised as an organisation rather than a rainforest. The combined end-to-end process of finding the mention of the entity *Mr. Brown* in the text, and disambiguating it to *Chris Brown* in a knowledge base is known as **entity linking** (EL).

During the last decade, various works have addressed named entity linking, recognition, and disambiguation. For the works focusing on disambiguation, transformers and contextual embeddings are also mainly used, where these models provide state-of-the-art results regarding different NLP tasks. Based on the aforementioned example, distinguishing between the person's name *"Brown"* and the colour *"brown"* requires, in the majority of the cases, an individual word embedding for each *"brown"* for disambiguating. Because of the challenges related to this disambiguation, the majority of the works focus on the NER task, and only a few of them propose systems dedicated to the whole entity linking pipeline.

One of the major challenges related to entity linking resides in the scarcity of datasets. Almost all the datasets are constructed manually, which is effort- and time-consuming, leading to corpora including only thousands of items (documents, sentences, reviews, or comments). While automatic corpora construction is starting to be adopted by researchers, manual approaches are more accurate and provide better results. However, those corpora are domain-centric and only useful for a single research purpose. Most of those resources cannot be adapted to real-life scenarios, as they are poorly performed when applied to another domain. Also, the majority of the resources focus on English, leading to a lack of research on other languages (which translates to less entity linking (EL) tooling for those languages).

The main goal of this survey is to highlight the most recent studies, directions, challenges and limitations that have been proposed for entity linking. For this purpose, this survey is organised as follows. We start with Sect. 2 presenting some generalities about entity linking, recognition and disambiguation. Then, we present in Sect. 3 the most recent previous surveys that we analysed. Section 4 presents the methodology that we followed to construct this survey and the research questions that we aim to answer. We divided the surveyed works into two categories: Section 5 focuses on research on the English language, and Sect. 6 presents

multilingual research. In Sect. 7, we focus on the resources that have been made available. We synthesise the studied works and resources in Sect. 8. We compare our survey to others in Sect. 9. The paper ends with a discussion of open issues and perspectives for future works in Sect. 10.

2 Entity linking: background

Consider as an example, the mention of *"Ford"*. This mention could be associated with the *Ford Motor Company* American multinational automaker, or *Henry Ford* the founder or the *Ford Foundation*. Only the context could indicate which entity is linked to the mention (*"Ford"*) in a knowledge base.[1] From the example mentioned above, two tasks could be highlighted: 1) extracting the different mentions from a given text, and 2) disambiguation, corresponding to the extraction of the link or association from the initial mention to the right entity in a specific context.

First, the mention *"Ford"* is extracted from a given text/document. Afterwards, this mention is associated with the motor company, Henry, depending on the context. We observe that depending on the context, the mention *"Ford"* could be the name of a person, of an organisation, or a location (if we referred to Ford Island[2]). The task of extracting *"Ford"* from the document by determining its category (name of a person, organisation, location, etc.) is known under several terms: most commonly, **named entity recognition**, **named entity resolution**, or **named entity extraction**. The task of linking the extracted mention *"Ford"* to its entity in a given knowledge base is known as **named entity disambiguation**. Both research areas are presented in detail below.

2.1 Named entity recognition (NER)

Depending on the domain of interest, the named entities to consider are different. For example, in the biomedical domain, genes are the entities of interest. In the general domain, recognising the names of persons, organisations, and locations is crucial. In addition to these, dates, insurance numbers, and postal codes are important for companies handling customer data.

In some cases, producing labels from a small set of entity classes is not enough, especially where other details are required for each entity. For example, in addition to the name of a person, its role (doctor, engineer, director, soldier, terrorist, etc.) in a given organisation is also needed [107]. For the location, it could also be interesting to detect whether the extracted location represents a city, a country, a mountain, a park, etc. Proposing this classification of named entity recognition in different categories and subcategories is known as **fine-grained named entity recognition (FGER)**. To distinguish between the two research areas, the first area is usually named as **coarse-grained named entity recognition (CGER)**, because it uses a more general classification.

[1] A knowledge-base contains a set of entities and a collection of texts in which a set of mentions are identified in advance [157].

[2] An islet in Oahu, USA.

Jeffrey Preston Bezos (born January 12, 1964) is an American businessman who is the founder of Amazon, the world's largest e-commerce and cloud computing company. With a net worth of US$160 billion as of September 2023, Bezos is the third-wealthiest person in the world and was the wealthiest from 2017 to 2021, according to both the Bloomberg Billionaires Index and Forbes.

Fig. 1 Named entity recognition example

Fig. 2 Named entity disambiguation example

2.2 Named entity disambiguation (NED)

The principal characteristic of NED systems is that they focus on the task of disambiguation of a given entity, independently from the NER task [50]. To disambiguate the extracted entities, two directions are considered [29]: 1) focusing on each entity locally, independently from the other entities and by relying only on the surrounding text, 2) focusing on all the entities on the document globally or collaboratively (at the same time) to ensure coherence. Some systems are solely dedicated to disambiguation [60, 76, 144]. Other systems are end-to-end, covering both NER and NED [92, 141]. An end-to-end NED system is equivalent to **entity linking** (EL) because it considers both NER and NED.

Figure 2.2 illustrates the major idea of NED where both *Paris* and *France* are linked to their respective Wikipedia pages.

In order to give the reader a global overview of the described works through this survey paper, we present Fig. 1. The figure classifies all the works that are described in more detail in the following sections. The works are classified into three different categories: survey work (presenting a state-of-the-art), research work (presenting an approach or a methodology), and datasets, tools and knowledge bases.

3 Previous surveys

To present this paper, we follow four main steps: (1) Gathering surveys on EL, NER, and NED. (2) Analysing the gathered surveys. (3) Extracting a set of issues related to the studied surveys. (4) Proposing an approach for constructing a survey presenting the most recent works and resolving the issues of the other surveys. For gathering all the pertinent surveys on the research literature, we ran searches on Google Scholar using the queries *"survey entity linking"*, *"survey named entity recognition/resolution/extraction"*, *"survey named entity disambiguation"*, *"state of the art entity linking"*, *"state of the art named entity recognition/resolution/extraction"*, and *"state-of-the-art named entity disambiguation"*. We filter by year to have only the most recent surveys that have been published from 2016.

This technique found five survey papers recently presented on entity linking (only one is from 2007, but most of the research literature cites it). The first one from Balog [17] considered entity linking in general, including both NER and NED. The others were principally dedicated

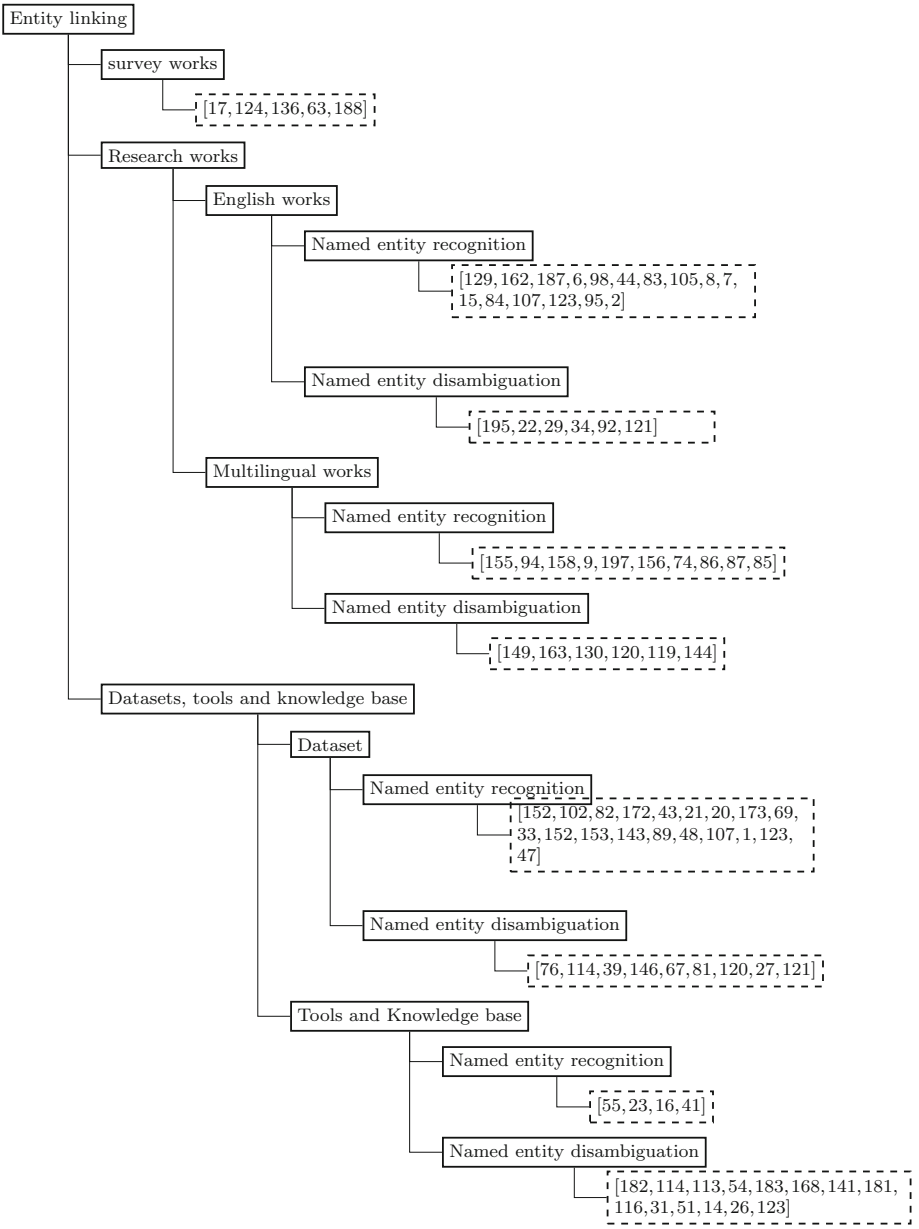

Fig. 3 Summary of the presented works

to NER [63, 124, 136, 188]. To the best of our knowledge, no surveys have been done exclusively on NED.

The survey from Balog [17] in 2018 defined the problems of EL, NER and NED. The author concludes by presenting the general process of EL, which is only composed of both NER and NED. The authors affirm that the named entities have to be extracted before being

disambiguated. The author also focuses on both NER and NED by citing some pertinent research in both areas. For disambiguation, the authors also distinguish between local and global disambiguation. Afterwards, the authors describe some available entity linking systems (such as AIDA, DBpedia Spotlight, TagMe) and some publicly available datasets (such as MSNBC, AQUQINT and ACE2004). The authors conclude by presenting a set of challenges related to both NER and NED and by briefly highlighting the new tendencies related to entity linking, related to the semantic embedding and neural models that are recently used. In this context, four recent works were cited [57, 60, 174, 200].

The 2007 survey of Nadeau et al. [124] is the first survey dedicated to NER. This survey spanned from 1991 to 2006. These authors classify NER approaches into three main categories: (1) supervised approaches using maximum entropy models, decision trees, hidden Markov models (HMM) and conditional random fields (CRF); (2) semi-supervised approaches using "bootstrapping" to construct a corpus-based on a set of initial seeds; (3) unsupervised approaches using clustering. This survey also considers multilingual aspects, presenting works done on German, Japanese, Greek, Italian and other languages.

Goyal et al. published in 2007 a detailed survey about NER [63]. The authors classify approaches as rule-based (using hand-crafted features) or machine learning-based. Similarly to Nadeau's survey, the ML-based approaches were divided into three main categories: supervised, unsupervised, and semi-supervised or hybrid. For each category, the authors produced a summary table comparing their various features: target language/domain, the technique used, the dataset, and the results.

Yadav et al. [188] present the most recent survey on NER (published in 2019) by focusing on architecture using deep learning models. The presented survey aims to compare feature-engineered and neural network systems proposed for multi-domain and multilingual NER. The authors classify the NER system presented in the research literature into four categories: (1) knowledge-based systems that use a domain-specific lexicon, (2) unsupervised systems using bootstrapping, (3) supervised systems that use annotated data and hidden Markov model (HMM), support vector machine (SVM), etc., and (4) neural network systems. The authors focus on the last category, and they regroup the studied neural network system into four other categories: (1) word-level architectures using the set of words (embedding) composing a sentence as an input of a recurrent neural network (RNN), (2) character-level architectures using a set of characters as an input to the RNN, (3) character- + word-level architectures, where two models were dominant (the first one uses a combination between word embedding and a convolution over the characters composing the words and uses CRF for the decoding step, and the second one concatenates word embeddings and the character using an LSTM or Bi-LSTM layer), and (4) character + word + prefix/suffix model which also integrates the prefixes/suffixes features into the model. In addition to the systems, the authors also presented some NER datasets.

Finally, the survey from Patil et al. [136] on multilingual NER classified systems across two categories: (1) the systems for Indian languages (such as Hindi, Bengali and Punjabi) and (2) the systems dedicated to non-Indian languages (such as English, Spanish, Chinese and Arabic). Most of the approaches presented by the authors are statistical approaches using conditional random field (CRF) and maximum entropy (Maxent) in the majority of cases. The authors also highlighted the use of hybrid systems combining rule-based and statistical approaches, or combining more than one statistical algorithm (such as Maxent and HMM). The authors concluded that the most critical issues related to NER in Indian languages are the lack of annotated corpora, the Indian morphology and the variations in the writing style.

Our analysis of the available surveys in the literature identified the following issues:

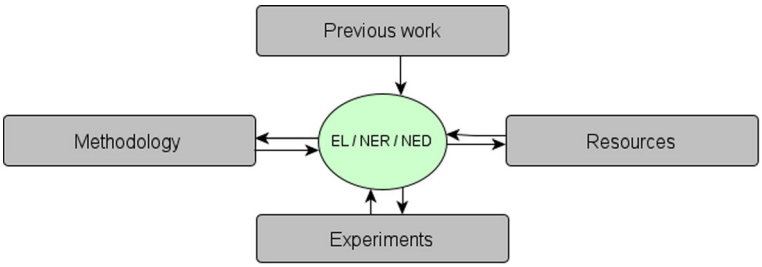

Fig. 4 Methodology for gathering research papers

- From the presented surveys, it can be concluded that only one paper focuses on NER + NED, and all the other surveys focus on NER only.
- No survey handles all of NER, NED, and multilingual aspects.
- Almost all the works presented by the studied surveys are old (before 2015).
- The surveys lack descriptions of the resources constructed, such as tools, API and datasets.
- The few datasets that were described did not include the necessary information to locate them online (e.g. the link to their website).

Our aim behind this survey is to present a recent paper focusing on the most recent works on multilingual NER and NED, resolving the issues cited above.

4 Survey methodology

We followed an incremental method for gathering the research works presented in this survey. Figure 4 outlines our approach.

We started by gathering the keywords related to EL, such as NER and NED. Then, we gathered only a few recent works on each field by targeting both English-specific and multilingual works. For each work, we focus on four main aspects: (1) previous works, to extract the added value of each studied work within the research literature, (2) methodology, to present to the community the most fundamental aspects of each methodology, (3) the used or constructed resources, to gather the publicly available APIs, tools, and datasets representing valuable resources for the community of research, and (4) the experiments and the parameters used with each model, their results and their comparisons with the previous works. Each one of these steps leads to gathering more work and more resources. From the collected works, we excluded the works handling **word disambiguation**.[3] We also excluded the research works presenting similar approaches using the same techniques, keeping one representative work in each family of approaches. This resulted in 167 research works about EL/NER/NED being selected, which we will discuss below.

Finally, in addition to presenting and analysing the works and resources focusing on EL/NER/NED, we also aim to answer the following research questions:

- Q1: What are the most recent methods/techniques used for entity linking (including named entity recognition and disambiguation)?
- Q2: What is the tendency related to the corpora? Did studies tend more to use a publicly available corpus, or did they prefer constructing their own corpora?

[3] Related to the disambiguation of all the words and not only the named entities.

– Q3: What are the main techniques proposed for constructing a corpus, and what are their main advantages and disadvantages?
– Q4: Which English-centric approaches produce the best results and performance?
– Q5: Which multilingual approaches produce the best results and performance?
– Q6: What are the open issues for entity linking?

To answer the above questions, we classify the works presented in the research literature into two main categories: the research works that have been done in English and the multilingual research works. For both categories, we split the works into NER and NED.

5 Research works on English

5.1 Named entity recognition

The most used approach for NER is "coarse-grained", which considers entities belonging to a small number of major classes (from one to ten). Most of the research studies focusing on a single class are dedicated to bio-medical NER: they extract the names of diseases, viruses, patients, etc. [98, 129]. The work focusing on three classes aims to extract person, location and organisation names [83]. The works focusing on four classes tend to focus on identifying person, location, organisation, and miscellaneous names [44, 162, 187]. Others focused on six [6] or ten categories [105]. The most important issue with the coarse-grained approach is that they focus on a small set of classes (up to 10). Fine-grained approaches aim to resolve this limitation, with some systems detecting more than 100 classes [107].

The coarse- and fine-grained classifications are correlated. For example, the coarse-grained class of person (PER) may contain the fine-grained classes of judge, lawyer and other person (plaintiffs, defendants, witnesses, appraisers, etc.). The location (LOC) includes the fine-grained classes of the country (LD: countries, states and city-states), city (ST: cities, villages and communities), street (STR: streets, squares, avenues, municipalities and attractions), and so on. The coarse-grained class "organization" (ORG) is divided into public, social, state and economic institutions, etc. [100]. We present, in the following, the works presented for both approaches (coarse-grained and fine-grained).

5.1.1 Coarse-grained NER approach

Different approaches were used, including rule-based, machine learning-based and clustering-based approaches. Rule-based approaches use pre-defined vocabularies that include complex logic. On the other hand, machine learning-based approaches employ statistical approaches (including support vector machines or SVM, and decision trees) [178]. More recently, deep learning methods have provided significant improvements in performance terms in multiple visual analysis tasks, such as object detection, classification and tracking. Deep learning models typically contain hundreds of thousands or even millions of trainable parameters, which give them their edge in terms of performance [133]. Finally, the goal of clustering is to discover the natural groupings of a set of objects. Many clustering algorithms are generic in the sense that they can be applied to any type of data that are equipped with a measure of distance between data points. Diverse types of clustering methods are available. The most popular clustering algorithm is k-means, which iteratively identifies k cluster centres (centroids), and each cell is assigned to the closest centroid [91].

In the context of rule-based approaches, Neelakantan et al. [129] propose an approach to automatically construct a dictionary for NER from Wikipedia, using a large corpus of unlabelled data and a few seed examples. This approach includes two steps: (1) collecting a list of candidate phrases from the unlabelled corpus for every named entity type using simple rules and (2) removing the noisy candidates from the list obtained to construct an accurate dictionary. To predict whether a candidate phrase represents a named entity, the lower-dimensional, real-valued canonical correlation analysis (CCA) embeddings of the candidate phrases are used as features, and the training is done using a small number of labelled examples and a binary SVM for classification. Two kinds of experiments were carried out: (1) using a dictionary-based tagger by relying on the four constructed dictionaries and the two used corpora (GENIA and NCBI) and (2) using the CRF-based tagger by considering the constructed dictionaries as features. First, the authors compare the different results obtained with the four dictionaries. For CRF, different regularisers[4] were used: 0.0001, 0.001, etc. The experiments using CRF were done using both CCA word and phrase embedding. The best F1-score obtained is up to 62.30 (on the GENIA corpus using CCA), up to 48.03 (on NCBI using manual construction), up to 79 (on GENIA using the CRF tagger and CCA-phrase), and up to 81 (on the NCBI corpus using the CRF tagger and CCA-phrase).

Only some studies use classic machine learning algorithms such as SVM [129]. The majority of the proposed studies rely on neural networks [6, 44, 98, 105, 187]:

- Xu et al. [187] proposed an approach that examines all possible fragments in the text (up to a certain length) one by one. It uses the FOFE method[5] to fully encode the fragment, its left context and right context into fixed-size representations, which are in turn fed to a FENN[6] to predict the entity mentions. This model is based on both character- and word-level models. In the evaluation phase, the authors also consider the nested entity (embedded names including others, for example, British Columbia or Western Canada).
- Aguilar et al. [6] propose a system, which embeds a sentence into a high-dimensional space (using CNN,[7] BiLSTM,[8] and dense encoders) to extract features. Afterward, the resulting vectors of each encoder are concatenated for performing multi-tasks. Finally, a CRF classifier uses the weights of the common dense layer to perform a sequential classification.
- Lee et al. [98] propose an approach relying on transfer learning and artificial neural networks to extract NER from patient note de-identification. Transfer learning is used to improve a learner from one domain by transferring information from a related domain.[9]

[4] Regularisation refers to techniques that are used to calibrate machine learning models in order to minimise the adjusted loss function and prevent overfitting or underfitting.

[5] Fixed-Size Ordinally-Forgetting Encoding Method for Neural Network Language Models can model the word order in a sequence using a simple ordinally forgetting mechanism according to the positions of words [198]

[6] Finite element neural network was obtained by embedding a finite-element model in a neural network architecture. It enables fast and accurate solutions of the forward problem [145]

[7] A convolutional neural network, which is one of the most popular deep neural networks. CNN has multiple layers, including a convolutional layer, pooling layer, nonlinearity layer and fully-connected layer. The CNN has an excellent performance in machine learning problems including natural language processing (NLP).

[8] A bidirectional LSTM (long short-term memory), or biLSTM, is a sequence processing model that consists of two LSTMs: one taking the input in a forward direction, and the other in a backward direction. BiLSTMs effectively increase the amount of information available to the network.)

[9] For example, if two people would kike to learn to play the piano. A person having an extensive music background (even on another instrument such as a guitar) would be able to learn the piano in a more efficient manner by transferring previously learned music knowledge to the task of learning to play the piano than a person with no experience [185]

The proposed model includes six significant layers: 2 embedding layers (one for the token, one for the characters). Two LSTM layers (one for tokens one for characters)- a fully connected and a CRF layer. Two kinds of experiments were carried out: (1) Experiment on different sizes of the training (the target) to show how many labels are needed for the target dataset to achieve consistent performances with and without transfer learning. (2) Experiment by transferring different combinations of parameters used in the neural network rather than all to show the importance of each layer.

- Dernoncourt et al. [44] propose NeuroNER, which is a state-of-the-art NER based on neural network. The purpose of NeuroNER is to allow users to annotate entities using a graphical web-based user interface (BRAT)[10] [169]. The model contains three layers: (1) LSTM for character embedding, (2) LSTM for token embedding and 3) CRF.
- The purpose of the work [105] is to propose an approach for dealing with the noisy and colloquial nature of tweets using an LSTM to learn orthographic features automatically. The proposed approach includes three main components: (1) orthographic sentence generator (described in Sect.), (2) word representations as input vectors, (3) bidirectional LSTM. At the output layer, the CRF log-likelihood (likelihood of labelling the whole sentence correctly by modelling the interactions between two successive labels) is used.
- Finally, the purpose of the work [83] is to propose a hybrid system (using Python script) combining different freely available NER tools. Four freely available NER tools were used: NER, Spicy, LingPipe and NLTK. The proposed tool can recognise the basic three entity types: PERSON, LOCATION and ORGANISATION. The four tools were evaluated on WikiGold. As the Stanford NER gave the best results, the constructed corpus was firstly annotated by Stanford NER and reviewed manually. The hybrid system was evaluated on both constructed corpus history and infopedia. The results were compared to the results returned by Stanford NER and Spicy.

The current state-of-the-art results were recently obtained using models targeting more than one natural language processing (NLP) problem, including NER. In this context, Baevski et al. [15] propose a bidirectional transformer architecture predicting every token in the training corpus by using a cloze-style training objective (where humans were asked to guess omitted words in a sentence using its context, knowledge of syntax and other skills [176]). The proposed model aims to predict the centre word given right-to-left and left-to-right context representations. This model was used for many NLP tasks, including text classification, Question-Answering, parsing, and NER tasks. For NER and parsing, the authors rely on different architectures (using embedding models previously presented in the research literature [45, 139, 140]), using different language models as well as different learning rates. This model outperforms all the results presented in the research literature, with an F1 score up to 93.5. With an F1-score slightly below (up to 93.47), Jiang et al. [84] propose a neural architecture search (NAS) dedicated to both language modelling and NER. These authors were the first to integrate differentiable NAS. Differentiable NAS uses a continuous relaxation to architecture representation, making gradient descent straightforwardly applicable to search [84]. They use recurrent neural networks (RNNs), including a recurrent cell consisting of 8 nodes. The proposed system runs on 40 training epochs with a batch size of 256 and a learning rate of 20. For NER and in addition to the works presented in [45, 139], these authors also compare their results to results presented by Lample et al. [96] (F1-score up to 90.94). Their results were also compared to the results obtained by Akbik et al. [8] (F1-score up to 93.18).

However, no publicly available system is associated with both works presented by Baevski et al. [15] and Jiang et al. [84]. Akbik et al. [7, 8] obtained results slightly below the afore-

[10] https://github.com/nlplab/brat.

mentioned systems, but their proposed system (Flair) is publicly available.[11] Flair is based on contextualised word embedding, which associates each word with its context in a given sentence [8]. Afterwards, this work was improved by integrating a memory of all the embedding context for a given word [7]. All the generated vectors are then concatenated using a pooling operation. This work was also compared to different recent works proposed in the research literature [5, 36, 45, 96, 139].

Only one work was found to use clustering [162]. In this work, the authors propose an approach based on k-means clustering with three numbers of clusters: 100, 1000, and 5000. The authors also rely on the skip-gram (SG) model of `word2vec` for extracting word vectors and on the linear support vector classifier (SVC) for classification. Table 1 summarises all the works presented in this part by showing the constructed and used datasets and giving some details on the experimentation and results.

5.1.2 Fine-grained NER approach

FIGER [107] is one of the first fine-grained systems proposed for NER. The authors consider the fine-grained problem as a multi-class, multi-label classification problem. This system recognises 112 tags using a conditional random field (CRF)[12] tagger to find the candidates and the perceptron algorithm for the classification. For training the CRF, the authors opt for automatic construction of the corpus to generate a larger dataset allowing the classification of all the tags. For this, they exploit the anchor links in Wikipedia text to label entity segments with the appropriate tags automatically. To validate their system, the authors carry out two types of experiments: the first one is dedicated to NER and the second one to relation extraction (RE). For the first one, the authors compared their system to two other systems in the research literature, Stanford NER [55] and Illinois Named-Entity Linking (NEL) [146]. For RE, the authors use MultiR[13] [77], trained using distant supervision by heuristically matching relation instances.

FIGER was compared to a more recent system, TypeNet [123]. TypeNet aims to integrate hierarchical information into the embedding space of entities and types to improve entity linking and fine-grained entity typing tasks. Although the principal goal of TypeNet is fine-grained NER, it also integrated an entity linking model based on a combination of string similarity score and string cosine similarity. To propose the hierarchy of the embedding space (the links between types and entities), the authors use two state-of-the-art knowledge graph embedding models: real and complex bi-linear maps. The real model is equivalent to RESCAL (a single IS-A relation type proposed by Nickel et al. [131]). The complex model is based on the ComplExmodel (using complex-valued vectors for types, and complex diagonal matrices for relations) proposed by Trouillon et al. [180] For evaluating their system, the authors consider three kinds of experiments: (1) mention-level entity typing using FIGER, (2) entity-level typing using Wikipedia and TypeNet, and (3) entity linking using MedMentions. The authors compare their approach to different state-of-the-art systems [68, 107, 160, 161].

Abhishek et al. [2] focused on FIGER [107] and Typenet [123] to present the HAnDS (Heuristics Allied with Distant Supervision) framework to automatically construct a dataset suitable for fine-grained NER. HAnDS requires three inputs: a linked text corpus (e.g.

[11] https://github.com/flairNLP/flair.

[12] CRF are a class of statistical modelling methods often applied in machine learning and used for structured prediction. These kind of taggers also consider "neighbouring" samples. Hence, a CRF can take context into account.

[13] https://github.com/ajaynagesh/multir.

Table 1 Works on coarse-grained NER

Work	Dataset constructed	Dataset used	Results
[129]	(1) Virus list (3100). (2) Disease list (60,080). (3) Dictionaries for DL-CoTrain +CCA algorithms	(1) GENIA corpus. (2) NCBI disease corpus. (3) List of disease names	Best F1: 62.30 (on GENIA corpus using CCA), 48.03 (on NCBI using manual construction), 79 (on GENIA corpus using CRF tagger and CCA-phrase), 81 (on NCBI corpus using CRF tagger and CCA-phrase)
[162]	–	(1) CoNLL03. (2) RCV1	Word2vec was trained on respectively 1/4, 1/2, 3/4 and all RCV1. 3 dimensions of clusters: 100,1000,5000. F1 = 83.1 with 3/4 of RCV1 and cluster dimension= 1000
[187][1]	(1) Machine-labelled Wikipedia (WIKI). (2) In-house dataset	(1) CoNLL03. (2)KBP2015	90.85 for CoNLL03. 0.739 KBP2015 (Trilingual). 0.75 for English part of KBP2015 combined to in-house corpus
[6][2]	–	WNUT-2017	CNN: kernel size = 3, filters = 64. BLSTM: dropout = 0.5. All network: batch size = 500, epoch = 150, Optimizer = Adam [90] CRF: L-BFGS as a training algorithm, Regularisation L1 + L2. F1-score: 41.86 (the best score in WNUT task)
[98][3]	–	(1) MIMIC. (2) i2b2(2014 and 2016 datasets)	60% of the target dataset for having the same performance as transfer learning (F1-score: 97.97). Each layer is important and the results (F1-score varying from 90.12 to 97.97)
[44][4]	–	(1) CoNLL03. (2) i2b2 (2014)	(1) character_lstm. (2) token_lstm (glove for word embedding). (3) Random_initial_transition (True). F1-score: 90.5 for CoNLL03 and 97.7 for i2b2 (2014)

Table 1 continued

Work	Dataset constructed	Dataset used	Results
[83]	22 web pages annotated [6]	WikiGold	Two scoring protocols (total matching and partial matching). The hybrid system outperforms the results returned by both Stanford NER and Spacy. The best F1 for history is 0.89 and it is 0.87 for Infopedia (where the best results for NER Stanford are up to 0.86 for both history and Infopedia)
[105]	–	(1) WNUT2016. (2) Pre-trained word embeddings [62]	A vector of random [72] is used for the unseen words. BiLSTM with: stochastic gradient descent (SGD), mini-batch size = 50, L2 regularisation, dropout rate = 0.5. F1 scores of 52.41 (segmentation and categorisation) and 65.89 (segmentation only)
[15]	–	(1) CoNLL03	Best F1: 93.5, using a CNN model trained on a large corpus containing 330 million of parameters. This model is fined-tuned using classification/regression and prediction tasks
[84]	–	CoNLL03	Best F1: 93.47, using an improved differentiable architecture search considering all input edges (related to a given node) in a single softmax
[7, 8][5]	–	(1)CoNLL03. (2) CoNLL2003g. (3) CoNLL2002 (Dutch part). WNUT-17	Best F1: 93.09 [8] and 93.18 [7]. SGD model with 150 epochs followed by an LSTM (256 hidden state). Max-pooling was also used in [7] for improving the results

[1] https://github.com/xmb-cipher/fofe-ner
[2] https://github.com/gaguilar/NER-WNUT17
[3] https://github.com/Franck-Dernoncourt/NeuroNER
[4] https://github.com/Franck-Dernoncourt/NeuroNER/tree/master/neuroner/trained_models
[5] https://github.com/flairNLP/flair
[6] http://eresources.nlb.gov.sg/index.aspx

Table 2 Works on fine-grained NER

Work	Dataset constructed	Dataset used	Experiments and results
[107][1]	(1) FIGER[2]. (2) FIGER (GOLD)[2]	(1) NYT corpus	FIGER outperforms both Stanford NER and NEL with F1 up to 0.639. The F1 of the association of MultiR and FIGER is up to 40 which is higher than the F1 from using only MultiR (up to 20.7)
[123][5]	(1) TypeNet. (2) MedMentions	(1) W2M data. (2) FIGER (GOLD)	F1 macro: 78.3 (outperforms all state-of-the-art system, except for Shimaoka et al. [161], up to 78.9). F1 micro: 75.4 (outperforms all state-of-the-art systems) by using the Hierarchy model
[95]	(1) Twitter dataset which contains 1000 tweets	–	The best precision for SANE CG is 96.83, where it is a little higher for FINET (96.88). The best precision for SANE FG is 78.82 (79.80 for FINET). SANE results are near state-of-the-art results, without relying on knowledge bases for type extraction
[2][3]	(1) WikiFbF[4]. (2) WikiFbT[4]. (3) Wiki-NDS[4]. (4) FIGER [4]. (5) 1k-WFB-g[4]	(1) FIGER (GOLD)	The best F1 scores for Fine-ED/Fine-ET are both obtained using Wiki-FbF corpus. For Fine-ED F1 is up to 82.94 for FIGER (GOLD) and up to 85.75 for 1k-WFB-g. For Fine-ET, F1 is up to 70.70 (ma-F1) 68.23 (mi-F1) for FIGER corpus. F1 is up to 68.42 (ma-F1) 69.23 (mi-F1) for 1k-WFB-g

[1] https://github.com/xiaoling/figer
[2] https://drive.google.com/file/d/0B52yRXcdpG6MMnRNV3dTdGdYQ2M/view
[3] https://github.com/abhipec/HAnDS
[4] https://drive.google.com/drive/folders/1LvVk7-ygqWT1VT-5BZ4HiMVP77KhgFvk?usp=sharing
[5] https://github.com/iesl/TypeNet
[6] https://github.com/Tsinghua-PhD/APE
[7] https://wiki.dbpedia.org/downloads-2016-10
[8] https://www.tensorflow.org/

Wikipedia), a knowledge base (capturing concepts, their properties, and inter-concept properties: e.g. Freebase) and a type hierarchy (a hierarchical organisation of various entity types, e.g. FIGER and TypeNet). To reduce the false-positive and the false-negative, HAnDS follows three stages: (1) link categorisation and processing for removing the incorrect anchor detected as entity mention, (2) inference of additional links, by linking the correct referential name of the entity mention to the correct concept in the knowledge base, and (3) sentence selection, allowing high-quality annotations by using a POS tagger and other features. For evaluation, the authors consider two sub-tasks: Fine-ED, a sequence labelling problem and Fine-ET, a multi-label classification problem. for Fine-ED, LSTM-CNN-CRF model[14] is used [112]. For Fine-ET, an LSTM-based

Lal et al. [95] present SANE, a system using Wikipedia categories to recognise fine-grained entities. The authors focused on named entity typing (NET), where they associate a semantic type to a given entity. SANE is based on Stanford NER for the extraction of named entities and on a pattern-based matching for fine-grained NER. The best categories are chosen using a selection model from Word2vec. The selected categories in the lookup-based extraction phase are mapped to appropriate WordNet types. A 3-class (Person, Organization, Location) NER classifier is used to find coarse-grained (CG) named entities. Afterwards, the identified entities are processed using SANE. SANE is compared to FINET. The results of both systems (i.e. SANE and FINET) were manually labelled by two independent annotators. The inter-annotator agreement (Kappa) is 0.72 for FINET and 0.86 for SANE.

Table 2 summarises all the works presented in this part by showing the constructed and used datasets and giving some details on the experimentation and results.

5.2 Named entity disambiguation

The first proposed NED approach focused on local disambiguation, which resolves the entity mentions independently and uses various hand-designed features and heuristics (specific to each mention) [34]. This approach suffers from two major limitations: it overlooks the topical coherence among the target entities, and unseen words/features are not recognised (data sparseness) [29]. To resolve these issues, global disambiguation (where all entity mentions are disambiguated simultaneously) was proposed [22, 29, 195]. We present below the identified works for both local and global disambiguation approaches.

5.2.1 Local disambiguation

Local disambiguation approaches disambiguate each mention in a document separately, utilizing clues including the textual similarity between the document and each candidate to disambiguate [146].

In this context, Chisholm et al. [34] propose an entity disambiguation system (named named entity linking (NEL)) to compare Wikipedia and Wikilinks. For extracting features, the authors apply three approaches: (1) entity prior, corresponding to the probability of a link pointing to a given entity, (2) name probability, corresponding to the relationship between a name and an entity, and (3) textual context, by using the surroundings words. Two techniques were used: Bag Of Words (BOW) context, and Distributional BOW, where a word embedding vector of dimension equal to 300 is used. To perform disambiguation, an SVM classifier was used. The authors carry out many experiments to compare Wikipedia and Wikilinks,

[14] https://github.com/jayavardhanr/End-to-end-Sequence-Labeling-via-Bi-directional-LSTM-CNNs-CRF-Tutorial.

showing the impact of combining them and the impact of the corpus size on the results, and then compare their NEL system to four other systems in the literature [11, 73, 76, 78].

5.2.2 Global disambiguation

The global optimisation problem is an NP-hard problem where approximations are required. For example, for Wikipedia, the common approach is to utilise the Wikipedia link graph to obtain an estimate of pairwise relatedness between titles in order to efficiently generate a disambiguation context [146].

Yang et al. [195] propose the structured gradient tree boosting (SGTB) learning model for named entity disambiguation. The constructed framework is built by using the SGTB model [194], by employing a conditional random field (CRF) objective. To compute the partition function (normalisation term) for training and inference, beam search is used. Moreover, Bidirectional Beam Search with a Gold path (BiBSG) is used for reducing the model variance and considering both past and future information in the prediction step. Two experiments are conducted: in-domain (using AIDA-CoNLL corpus) and cross-domain (using all the other datasets). Cao et al. [29] present NCEL, a neural model for collective entity linking. NCEL includes three main compounds: candidate generation, feature extraction and neural model. Firstly, the generation of the different candidates is based on Wikipedia page titles, and a dictionary derived from a large web corpus and Yet Another Great Ontology (YAGO). Secondly, Neural Collective Entity Linking (NCEL) uses both local contextual features (based on similarity) and global coherence information (based on a window size for defining neighbour adjacency). Finally, NCEL incorporates graph convolutional networks into a deep neural network to utilise structured graph information for collective feature abstraction. NCEL was compared to 16 other systems focusing on EL [34, 59–61, 68, 73, 76, 93, 117, 122, 135, 141, 177, 181, 193, 199]. The results were compared using the Gerbil[15] benchmark. The authors also detail the analysis of two corpora, the less complex one (TAC2010) and the most complex one (WIKI and CWeb).

Bhatia et al. [22] present a simple, fast, and accurate probabilistic entity-linking algorithm used in the enterprise. To do this, the authors rely on automatically constructed domain-specific knowledge graphs. The idea of this approach is to first extract the named entities from the query (using publicly available systems such as Apache OpenNLP[16] or Stanford NER[17]). Afterwards, a list of target entities is generated by retrieving all entities from the graph containing the extracted tokens. For each result, the entity and text context are computed using the naive Bayes algorithm. The role of the entity context component is to compute the probability of observing the entities forming the context after observing the target entities. The role of the text context component is to compute the probability of observing the query terms after observing the target entities. Finally, the scores for all the target entities are combined to produce a final ranked list. The authors compare their approach to 5 other works in the research literature [4, 30, 75, 76, 113]. Their knowledge base has 2,261 candidates per mention to disambiguate, which is high compared to manually cleaned knowledge bases such as DBpedia.

Kolitsas et al. present an end-to-end system to perform the task of entity linking [92], inspired by the most recent models of Lee and al. [99] and Ganea et al. [60]. The purpose of this system is to generate all possible spans/mentions to select the top candidates

[15] https://github.com/dice-group/gerbil.

[16] http://opennlp.apache.org/.

[17] https://nlp.stanford.edu/software/CRF-NER.html.

Table 3 Works on NED

Work	Dataset constructed	Dataset used	Results
[195][1]	–	(1) AIDA-CoNLL. (2) AQUAINT. (3) MSNBC. (4) ACE2004. (5) WIKI. (6) CWeb	Accuracy up to 95.9 for in-domain experiments(outperforms all other systems). Accuracy up to 90.5 for AQUAINT corpus, up to 92.6 for MSNBC corpus (up to 93.7 in [60]), up to 89.2 for ACE corpus, up to 81.8 for CWeb corpus and up to 79.2 for WIKI (up to 84.5 in [67]) for cross-domain
[22]	(1) A semantic graph of Wikipedia [8]	(1) The KORE50	knowledge base has 2261 candidates. Precision up to 0.74, which is better than the precision obtained by the compared approach. The response time by the query was up to 125 ms
[29][5]	–	(1) AIDA-CoNLL. (2) TAC2010. (3) ACE2004. (4) AQUAINT. (5) WIKI and CWeb	Two layers and 1 hidden unit in MLP encoder. NCEL outperforms all the other systems (except with AIDA-CoNLL corpus and ACE2004. The best F1 micro for AIDA-CoNLL was obtained in [59] and in [67] for ACE2004). Best F1 micro/F1 macro are respectively 0.91/0.92

Table 3 continued

Work	Dataset constructed	Dataset used	Results
[34][6]	–	(1) AIDA-CoNLL. (2) TAC2010	Combining Wikipedia and Wikilink gives the best results on both corpora (88.7 for CoNLL and 80.7 for TAC10). NEL outperforms almost all the other compared systems for TAC10 (where the best accuracy is up to 81.0 and was obtained by [73])
[92][7]	–	(1) Wikipedia 2014. (2) Gerbil datasets. (3) AIDA/CoNLL	LSTM, with a dimension of 300, for constructing both character and word embedding vectors. The best results were obtained with the model combining the proposed system with the Stanford NER system (with an F1 up to 66.9)
[121]	–	(1) T-REx	300-dimensional word embeddings on Wikipedia 2014 and Gigaword from Glove [138]. Bi-LSTM has one layer and 256 hidden units. Best F1-score obtained is up to 71.3

[1] https://github.com/bloomberg/sgtb
[2] http://opennlp.apache.org/
[3] https://nlp.stanford.edu/software/CRF-NER.html
[4] https://www.ibm.com/watson/services/natural-language-understanding/
[5] https://github.com/TaoMiner/NCEL
[6] https://github.com/wikilinks/nel
[7] https://github.com/dalab/end2end_neural_el
[8] The graph includes30millions of entities and 192 million distinct relationships

referred by each mention. The best candidates are selected using an empirical probabilistic entity/map built by Ganea et al. [60] and based on Wikipedia hyperlinks, Crosswikis [167] and YAGO. To disambiguate the generated candidates, the authors compute a similarity score using embedding dot products (of the different word embedding vectors constructed for the mentions and their context). For extending their model from local disambiguation to global disambiguation, the authors added a layer to their neural network model. However, for global disambiguation, they only consider the candidate with the highest local score. The proposed system was compared to many state-of-the-art systems included in Gerbil.

Mulang et al. [121] present Arjun, a context-aware entity linking approach, including 3 subtasks: (1) surface form extraction identifying all the surface forms associated with the entities, (2) entity mapping (or candidate generation) where the surface forms are mapped to a list of candidate entities from the local knowledge graph, and (3) entity disambiguation, where the most appropriate candidate entity for each surface form is selected. For both subtasks (1) and (3), the authors extended the attentive neural model proposed by Luong et al. [109]. In contrast to Luong et al., the authors use a bidirectional long short-term memory (Bi-LSTM) model for the encoder and a one-directional LSTM model for the decoder. For creating the local knowledge graph, the authors follow the same methodology described by Sakor et al. [151] where each entity label is extended with its aliases from Wikidata. Arjun was compared to OpenTapioca [42], which is an end-to-end EL approach released for Wikipedia.

Table 3 summarises all the works presented in this part by showing the constructed and used datasets and giving some details on the experimentation and results.

6 Multilingual research works

6.1 Named entity recognition

The purpose of the paper of Seyler et al. [155] is to show the importance of external knowledge for performing NER. The authors present a novel modular framework that divides the knowledge into four categories: (1) knowledge-agnostic, including local features extracted directly from the text, (2) name-based knowledge that identifies patterns in names and exploits the fact that the set of distinct names is limited, (3) knowledge base-based knowledge extracted from an entity annotated corpus, and (4) entity-based knowledge by encoding document-specific knowledge about the entities found in the text. The extracted features were used to train a linear-chain CRF. The experimentation shows the impact of incrementally adding external knowledge. The system was also applied to two additional languages, namely German and Spanish.

Kuru et al. [94] present CharNER, a character-level tagger for language-independent Named Entity Recognition (NER). CharNER operates at a character level, where the characters belonging to the same word are annotated with the same tag. The system architecture is composed of a 5-layer bidirectional LSTM network, connected to an output layer (a softmax layer). Finally, a Viterbi decoder takes the sequence of character tag probabilities produced by the softmax layer and produces word-level tags. The presented results are close to those of the literature, without using any manually generated features.

Shen et al. [158] present a deep active learning architecture to extract NER with a small training corpus. To reduce the computational complexity, CNN was used as a character-level and word-level encoder and LSTM as a tag decoder. For active learning, the authors

explore the uncertainty-based sampling strategy [101]. They use several algorithms: (1) least confidence (LC, for sorting examples in ascending order according to the probability assigned by the model), (2) maximum normalised Log-Probability (MNLP, normalising LC for concentrating on both long and short sequences, in contrast to LC which concentrates only on long sequences), (3) interpreting the variability of the predictions over successive forward passes due to dropout as a measure of the model's uncertainty, and (4) other sampling strategies (OSS, by maximising the representatives of the label set without querying a similar example).

Al-Rfou et al. [9] propose Polyglot, a language-independent NER system. To automatically construct a system dedicated to 40 languages, the authors relied on Wikipedia, Freebase and neural word embeddings. The authors consider the NER task as a word-level classification problem (same as the model proposed by Collobert [37]). Polyglot includes two main stages: (1) encoding the semantic and syntactic features of words in each language and (2) automatically generating a corpus from Wikipedia and Freebase. The Polyglot embeddings [10] were used for each language: the model was trained on Wikipedia without any labelled data. The process of creating a NER corpus includes two steps as well: (1) linking the Wikipedia articles to the corresponding entities and (2) using the exact surface form matching to extend the annotation (oversampling). The authors compare their system to that of Nothman et al. [132].

Yu et al. [197] present Cog-Comp, a Character-level Language Model (CLM) that considers each letter as a word and each word as a sentence, to show its impact on multilingual NER. The authors focus on 8 languages, including English. They also propose two features (entity, and language, based on the original language of the named entities) for improving the results. The system proposed by the authors was compared to two state-of-the-art NER systems, Cog-CompNER [88] and LSTM-CRF [96].

Shao et al. [156] also investigate the impact of additional features and configurations on neural network-based models in the context of multilingual NER. The authors focused on three baseline models, including a standard Bi-LSTM, a feed-forward network, and a window-based Bi-LSTM. The authors consider many features such as CRF at the output layer, POS and character embedding layers, and 3 different activation functions (hard sigmoid, relu and tanh). The models were applied to three languages, including English, German and Arabic. The authors compare their models to many systems proposed in the research literature. For English, the models were compared to 4 state-of-the-art systems [35, 37, 56, 80]. Three state-of-the-art systems were compared for German [3, 70, 147]. For Arabic, the models were compared to the system of Benajiba et al. [19].

Halwe et al. [74] also focus on a low-resourced language (Arabic) by presenting a deep co-learning approach to extract the named entities. The authors first construct an algorithm classifying Arabic Wikipedia articles into one of the four categories, namely person, location, organisation and objects (for non-entities). Afterwards, the authors rely on the proposed classifier to automatically annotate a large corpus of Arabic Wikipedia articles (25,000 articles). The authors were able to partially annotate 66,156 sentences from the extracted corpus. Finally, the authors adopt the concept of co-training proposed by Blum et al. [24] to combine annotated corpora, with their partially annotated constructed corpus for the task of NER. As a deep neural network architecture, the authors use both LSTM and BiLSTM layers. They also combine Bi-LSTM and CRF.

More recently, Jin et al. [85–87] focused on approaches transforming entities in a knowledge base (KB) to an entity graph to apply graph-based algorithms to it. The idea of their first work [86] is to construct an entity graph by using links between entities. They also used both graph structure and entity features for fine-grained NER. They applied an attributed and

predictive network embedding model to construct entity features and structure the graph. Finally, they use multi-label classifiers to determine the entity class. The authors compare their approach to 8 state-of-the-art methods (FIGMENT [190], CUTE [186], MuLR [191], Global [128], Corpus [189], PTE [175], Planetoid [196] and ASNE [104]).

In their second work, Jin et al. [87] convert entities in the KB into three semantic graphs. Each graph represents a specific kind of correlation among entities. The first one (Aco) is dedicated to representing the co-occurrence relations among entities. The second one (Acat) represents the category-proximity between entities. The third one (Aprop) represents property proximity between entities. Afterwards, the authors propose hierarchical multi-graph convolutional networks (HMGCNs), representing a deep learning architecture combining Aco, Acat and Aprop. To handle relations between types, a recursive regularisation is adopted. The proposed approach was compared to 4 of the state-of-the-art systems mentioned above (FIGMENT, CUTE, MuLR and APE). Experiments show that the two approaches proposed by Jin et al. significantly outperform all the compared system.

Finally, in their most recent work [85], the same authors propose a multilingual transfer learning model combining a mixture-of-experts approach. Their model dynamically captures the relationship between the target language and each source language and generalises to predict types of unseen entities in new languages. They investigate the role of the similarity between the source and the target languages on performance. They focused on six languages: German, English, Dutch, Russian, Spanish and Chinese. The main idea of their model is to use multiple source languages as a mixture of experts to learn the metric related to the weight of the experts for different target examples. For extracting features, they rely on mBERT [46] being pre-trained on concatenated Wikipedia data in 104 languages. From their different experiments, the authors conclude that the more similar the source and the target languages are, the better the performance will be: a large set of source languages with a high deviation of similarity performs worse than one of its subsets whose members are more similar to the target than other sources. The best-obtained F1-score (0.636) was achieved using three languages (English, German and Spanish), where English was relatively more important.

Table 4 summarises all the works presented in this part by showing the constructed/used dataset and giving some details on the experimentation and results.

6.2 Named entity disambiguation

Rosales et al. [149] present VoxEL, a multilingual manually annotated dataset dedicated to entity linking. In addition to English, VoxEL includes German, Spanish, French and Italian. VoxEL is based on 15 news articles (94 sentences mostly dedicated to politics, particularly at a European level) sourced from VoxEurop. Two kinds of tagging are used: (1) strict tagging, based on three entities (person, location and organisation), and (2) relaxed tagging, using a knowledge base and considering any noun phrase mentioned in Wikipedia as a valid entity. 204 mentions were annotated by strict VoxEL and 674 by relaxed VoxEL (for each language). For validating this dataset, the authors compare it to other multilingual corpora dedicated to EL by using various state-of-the-art multilingual systems, including TagME [53], TDH [49], DBpedia Spotlight [113], Babelfy [117], and FREME [154]). To present a fair comparison, the authors carry out all the experiments on the Gerbil benchmark.

Sil et al. [163] present LIEL, a Language-Independent Entity Linking system, including two steps: (1) extraction of the different mentions related to named entities and (2) linking the extracted mentions to a knowledge base (Wikipedia). To extract mentions and perform coreference resolution, the authors use the IBM Statistical Information and Relation Extraction

Table 4 Work on multilingual NER

Work	Dataset constructed	Dataset used	Results
[155]	–	(1) CoNLL03. (2) MUC-7. (3) CoNLL2003g. (4) CoNLL2002	Experimentation on 3 languages; English, German and Spanish. The best results for English were obtained on CoNLL03 (up to 91.12, by adding all the external knowledge defined). The experiments on CoNLL2003g and CoNLL2002 slightly outperform the results presented in the literature (with a respective improvement of 1.56 and 1.98)
[94]	–	(1) CoNLL2002. (2) CoNLL2003. (3) The Turkish NER dataset. (4) The Czech NER dataset. (5) ANERCorp (Arabic) and different combinations related to the size of the network are used. The results are up to, 78.72 (Arabic), 72.19 (Czech), 79.36 (Dutch), 84.52 (English), 70.12 (German), 82.18 (Spanish), 91.30 (Turkish)	
[158]	–	(1) CoNLL03. (2) OntoNotes−5.0	LSTM word-level encoder/character-level and CNN word-level encoder/character-level. Best F1: 90.89 (CoNLL03) and 86.63 (OntoNotes−5.0). Achieving 99% performance of the best deep model trained on 24.9% of the training data on the English dataset and 30.1% on Chinese. Also, 12.0% and 16.9% of the training data were enough to surpass the performance of the shallow models [143]
[9][1]	Polyglot-NER dataset including 40 languages	(1) Polyglot embedding. (2) Wikipedia. (3) Freebase. (4) CoNLL dataset	11 languages are used. Polyglot was compared to other NER systems in the research and other NER tools (i.e. OpenNLP, Stanford NER, and NLTK). The results obtained in English and Spanish outperform the results in the literature (F1 up to 71.3 for English and 63.0 for Spanish). Polyglot also outperforms the results of OpenNLP and NLTK

Table 4 continued

Work	Dataset constructed	Dataset used	Results
[197][2]	–	(1) CoNLL 2003. A subset of LORELEI project annotated [171]. (2) Wikipedia	4 different kinds of language models were used, inspired by Peng et al. [137], N-gram was implemented using SRILM toolkit [170]. Introducing features improves the results (F1 up to 96.5 for English and up to 89.4 for the others)
[156]	–	(1) CoNLL 2003. (2) GermEval 2014 NER shared task. (3) ANERcop	40 epochs for the Feedforward and 80 epoch for LSTM. The best F1 score is up to 85.59 for English, 66.36 for German and 63.37 for Arabic. The proposed system is comparable to the best-performing NER system in English and German data sets, but it is behind the Arabic system because Arabic requires more pre-processing than German
[74]	An Arabic Wikipedia NER corpus partially annotated (66–156 sentences)	(1) ANERCorp. (2) NewsFANE_Gold [12]. (3) Different testing corpora (NEWS + Tweets dataset [40] and AQMAR dataset [115])	The best F1 on NEWS corpus is up to 0.74 (BiLSTM co-learning). The best F1 on AQMAR is up to 0.62 (BiLSTM co-learning). The best F1 on Tweets corpus is up to 0.59 (BiLSTM co-learning + CRF). The results obtained by the authors outperform all the results presented in the research literature on the same test corpora
[86][3]	(1) A large FGNER dataset based on DBpedia (214 types, 300,000 entities and 5,243,230 entities)	(1) DBpedia[7]	Best macro-F1:70.2. Best micro-F1: 71.1. It outperforms all the systems presented in the literature. The best results were obtained by APE on DBpedia(macro-F1: 76.1, micro-F1: 78.5) and were obtained by FIGMENT on FIGER (macro-F1: 78.5, micro-F1: 81.9)
[87][4]	–	(1) FIGER. (2) DBpedia	Best macro-F1:79.8, best micro-F1: 82.4 (DBpedia) and, best macro-F1:79.8, best micro-F1: 83.6 (DBpedia). It outperforms all the systems presented in the literature. The best results were obtained by CUTE (macro-F1: 67.3, micro-F1: 67.7)
[85][5]	–	(1) MVET dataset [192]	The best-obtained F1-score (0.636) was achieved using three languages, English, German and Spanish where English is relatively more important

[1] https://sites.google.com/site/rmyeid/projects/polylgot-ner
[2] https://cogcomp.org/page/publication_view/846
[3] https://github.com/Tsinghua-PhD/APE
[4] https://github.com/SIGKDD/HMGCN
[5] https://github.com/SIGKDD/CLET

(SIRE) tool.[18] For mention detection, the authors use a CRF model of IBM SIRE and use the maximum entropy clustering algorithm for co-reference resolution (where 53 entity types were identified). For the entity linking step, the authors search the best mention that would maximise the information extracted from the entire document. LIEL was compared to many systems for English ([32, 114, 159]), and for Chinese and Spanish it was compared to the systems of the shared tasks at TAC 2013[19] and TAC 2014.[20]

BENGAL [130] is the first automatic approach which uses structured data to produce entity-linking benchmarks. The first purpose of Bengal is to propose a gold standard to generate benchmarks in English and also in other languages, such as Brazilian Portuguese, and Spanish. BENGAL is based on an RDF[21] graph. BENGAL starts by selecting a set of seed resources from the graph using a given number of triples to use during the generation process: for example, if the number is 3, BENGAL focuses on the person, organisation and location triples. To extract the set of seeds, a SPARQL query is used. A set of sub-graphs is then extracted, describing the information of each entity, the relationship between entities, and other aspects. The last part of the approach consists of applying a verbalisation, which transforms the graph into a set of sentences (documents) by using a set of predefined predicates. Gerbil was used to evaluate the performance of BENGAL on English by comparing it to other datasets constructed manually. BENGAL was also used for evaluating the annotation performance in Brazilian Portuguese. In this case, an RDF verbaliser [118] was used. This verbaliser was extended to Spanish using an adaptation of SimpleNLG [165].

MAG is a multilingual, knowledge-base-agnostic and deterministic entity linking approach [120]. MAG consists of an offline phase and an online phase. During the offline phase, five indexes are generated: surface forms (all the labels related to an entity), person names (all the variations of person names across different languages), rare references (using the Stanford POS tagger [179] to extract adjectives related to the entities), acronyms (a handcrafted index from STANDS4[22]), and context (using Concise Bounded Description[23]). The online phase consists of two steps: candidate generation and disambiguation. To generate all the candidates, all mentions are preprocessed by separating the acronyms (each word containing 5 letters or less is considered an acronym) from the string mentions. The string mentions are normalised. Afterwards, the candidates are searched using three different techniques: (1) by acronym, if the mention is classified as an acronym, (2) by label, relying on the set of surface forms which were generated, and (3) by context, using the TF-IDF metric.[24] Finally, a disambiguation graph is constructed to extract the optimal candidate. This step is equivalent to the disambiguation approach of AGDISTIS [181] based on HITS[25] and PageRank.[26] MAG was recently extended to support 40 languages, including low-resourced languages such as Ukrainian, Greek, Hungarian, Croatian, Portuguese, Japanese and Korean [119]. This work also presents a demo relying on online web services which allows for easy access to the entity linking approaches previously proposed by the authors [120]. By using

18 http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/relationship-extraction.html.

19 https://tac.nist.gov/publications/2013/papers.html.

20 https://tac.nist.gov//2014/BiomedSumm/.

21 Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications.

22 https://www.abbreviations.com/.

23 https://www.w3.org/Submission/CBD/.

24 term frequency-inverse document frequency is a statistical measure that evaluates how relevant a word is to a document.

25 Hyperlink-Induced Topic Search is a link analysis algorithm that rates Web pages.

26 PageRank is an algorithm used by Google to rank web pages in their search engine results.

this demo, the user is also able to define a set of parameters such as the graph-based algorithm (choosing between HITS and PageRank), whether to use acronyms or not, etc. MAG is also used in a domain-specific problem using a knowledge base of music terms [134].

Raiman et al. [144] propose DeepType, a system associating with each entity a set of types (e.g. Person, Place, etc.) to disambiguate entities. The authors were inspired by the previous work of Ling et al. [106] showing an improvement in the performance of their system after integrating the types proposed in FIGER [107]. The type system is automatically designed using a set of relations from Wikipedia and Wikidata. To predict the type system, the authors propose an algorithm containing 2 steps: (1) stochastic optimisation or heuristic search, and (2) gradient descent to fit classifier parameters. The idea of stochastic optimisation is to use an objective proxy function to avoid training an entire entity prediction model for each evaluation of the objective function. A neural network classifier is then trained by incorporating the resulting type to label data in multiple languages. A bidirectional LSTM network [96] with the word, prefix, and suffix embeddings (as previously done by Andor et al. [13]) is used. DeepType was compared to three other state-of-the-art systems [53, 114, 193].

Table 8 summarises all the works presented in this part by showing the constructed/used datasets and giving some details on the experimentation and results. More details about the constructed/used datasets, tools and ontologies that were referenced in this section are presented in the following part (Sect. 7).

7 Datasets, tools and knowledge bases

7.1 Datasets

7.1.1 NER datasets

This part describes 22 datasets used in the research literature and that was proposed/used by the works presented in this survey. CoNLL03[27] consists of a set of newswires in English [152]. CoNLL03 is separate from the Reuters RCV1[28] corpus (RCV1 was constructed from August 1996 to August 1997) [102]. KBP2015 is a trilingual dataset[29] that consists of discussion forum posts and news articles that were published in recent years: all the documents are related, but they are not parallel across languages [82]. WNUT-2016[30] [172] is a corpus consisting of tweets which were manually annotated using BRAT[31]: the corpus distinguishes 10 different named entity types (i.e. person, company, facility, geo-loc, movie, music artist, other, product, sports team and TV show). WNUT-2017[32] [43] is a manually annotated corpus used in the 3rd Workshop on Noisy User-generated Text (W-NUT). The documents contain the types *person*, *location*, *corporation*, *product*, *creative-work* and *group*. This corpus was extracted using many sources such as YouTube, Twitter, etc. For the manual annotation, three annotators were assigned to each document. The GermEval[33] 2014 NER shared task

[27] https://github.com/synalp/NER/tree/master/corpus/CoNLL-2003.

[28] https://trec.nist.gov/data/reuters/reuters.html.

[29] LDC2015E42: TAC KBP 2015 Tri-Lingual Entity Discovery and Linking Knowledge Base.

[30] https://github.com/napsternxg/TwitterNER/tree/master/data.

[31] http://brat.nlplab.org/.

[32] https://noisy-text.github.io/2017/emerging-rare-entities.html.

[33] https://sites.google.com/site/germeval2014ner/data.

Table 5 Work on multilingual NED

Work	Dataset constructed	Dataset used	Results
[149]	VoxEL[1]	(1) SemEval 2015 Task 13. (2) DBpedia Abstracts. (3) MEANTIME	F1 up to 0.857 (TagME, English corpus, Relaxed corpus). F1 up to 0.805 (BabelfyR, Spanish corpus, strict corpus)
[163]	–	For English: (1) ACE. (2) MSNBC. (3) TAC2014. For Spanish: TAC 2013/ TAC 2014. For Chinese: TAC 2013	Bag-of-Titles (BOT) [114, 146] and F1 were used (MSNBC + ACE). The metrics B3+ F1 (used in TAC shared tasks).F1 is up to 0.862 (ACE/English). B3 +F1 is up 0.80 (TAC 2014/ Spanish). B3+ F1 is up to 0.60 (TAC 2013/ Chinese). LIEL outperforms the results presented by the research literature for English and Spanish
[130][2]	BANGEL[3]	(1) Gerbil datasets. (2) HAREM dataset[4] [58] (for Portuguese). (3) Voxel dataset (for Spanish)	13 datasets in English and 4 in Spanish was automatically generated. Best F1: 0.84, using DEXTER. BENGAL outperforms almost all the manually constructed datasets included in Gerbil

Table 5 continued

Work	Dataset constructed	Dataset used	Results
[120][5]	–	(1) English dataset: Gerbil dataset (ACE2004, AIDA/CoNLL, KORE50, etc.). (2) Multilingual dataset: N^3 news.de, DBpedia Abstracts	The best F1 (English) is up to 0.69, using the HITS algorithm. The best F1 (multilingual) is up to 0.80 using PageRank and HITS on both languages French and Italian. (3) The best F1 (Fine-grained, using an extension of Gerbil [184]) is up to 0.95 related to the entity person
[119][5,6]	–	(1) Wikipedia. (2) Wikidata. (3) DBpedia	MAG [120] system was extended for handling 40 languages. A system usability study (SUS)[7] was used to validate the design of the user interface, and 15 users were asked for the annotation. The SUS-Score is up to 86.3 (interface with the top 10%)
[144][8]	–	(1) Wikipedia/Wikidata. (2) WIKI-DISAMB30 [53]. (3) AIDA-CoNLL. (4) TAC-2010. (5) 1000 articles (English, French, German, and Spanish) from Wikipedia	Best F1-score: 94.33 (wiki English), 92.98 (wiki French), 98.68 (wiki German), 98.24 (wiki Spanish), 92.37 (WIKI-DISAMB30), 94.87 (AIDA-CoNLL), 90.85 (TAC-2010). Deeptype outperforms all the compared state-of-the-art systems

[1] https://users.dcc.uchile.cl/~hrosales/VoxEL.html
[2] https://github.com/dice-group/BENGAL
[3] https://hobbitdata.informatik.uni-leipzig.de/bengal/
[4] http://faturl.com/bengalpt
[5] http://agdistis.aksw.org/mag-demo
[6] https://github.com/dice-group/AGDISTIS_DEMO/tree/v2
[7] https://measuringu.com/sus/
[8] https://github.com/openai/deeptype

[21] dataset represents a collection of citations that were extracted from News Corpora and German Wikipedia. ANERcop[34] [20] was annotated using the same format of the CONLL 2002 corpora. ANERcop was manually collected and annotated by only one annotator to maintain coherence.

The MIMIC and i2b2 (2014 and 2016 datasets) [173][35] were used by Lee et al. [98] to show the utility of transfer learning for NER. MIMIC is a part of the MIMIC-III dataset.[36] WikiGold[37] [69] is an annotated corpus including a small sample of Wikipedia articles in CoNLL format (IOB) [83]. MUC-7 [38] is a set of New York Times articles [33]. CoNLL2003g[39] [152] was extracted from the German newspaper Frankfurter Rundshau, between September and December 1992. CoNLL2002[40] [153] is a collection of newswire articles from May 2000, made available by the Spanish EFE News Agency [153]. OntoNotes−5.0[41] is a large-scale corpus of annotation of multiple levels of the shallow semantic structure in text. OntoNotes−5.0 contains three languages: English (one million words + 200K words of the English translation), Chinese (one million words) and Arabic (300K words). OntoNotes−5.0 was extracted from newswire and magazine articles, broadcast news, broadcast conversations, web data and conversational speech data. The GENIA corpus[42] is a semantically annotated corpus dedicated to biological text mining. In the GENIA corpus, the articles are encoded in an XML-based mark-up scheme, and each article contains its MEDLINE ID, title and abstract: all the texts in the abstracts are segmented into sentences [89]. The NCBI Disease corpus[43] is a large-scale corpus consisting of 6,900 disease mentions in 793 PubMed citations.[44] This corpus was developed by a team of 12 annotators (two people per annotation) and covered all sentences in a PubMed abstract. Disease mentions are categorised into Specific Disease, Disease Class, Composite Mention and Modifier categories [48].

WikiFbF[45] is a corpus created automatically using Wikipedia, Freebase and the FIGER hierarchy [107]. WikiFbT[46] is a corpus also created automatically using Wikipedia, Freebase and the TypeNet hierarchy. Wiki-NDS[47] is a corpus created using the naive distant supervision approach with the same Wikipedia version used for creating both WikiFbF and WikiFbT. 1k-WFB-g[48] is a fine-grained annotated corpus, manually annotated by Ling et al. [1] to cover large typeset. The sentences used for the construction of this corpus were extracted from Wikipedia text. Typenet[49] [123] is a dataset of hierarchical entity types for fine-grained

[34] https://raw.githubusercontent.com/HassanAzzam/Arabic-NER/master/ANERCorp.

[35] https://www.i2b2.org/NLP/DataSets/.

[36] https://github.com/MIT-LCP/mimic-code.

[37] https://csee.essex.ac.uk/staff/poesio/Teach/807/Assignments/Ass2/.

[38] https://catalog.ldc.upenn.edu/docs/LDC2001T0.

[39] http://lcg-www.uia.ac.be/conll2003/ner/.

[40] https://github.com/teropa/nlp/blob/master/resources/corpora/conll2002/esp.testa.

[41] https://catalog.ldc.upenn.edu/LDC2013T19.

[42] http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA.

[43] http://www.ncbi.nlm.nih.gov/CBBresearch/Fellows/Dogan/disease.html.

[44] https://www.ncbi.nlm.nih.gov/pubmed.

[45] https://drive.google.com/drive/folders/1LvVk7-ygqWT1VT-5BZ4HiMVP77KhgFvk?usp=sharing.

[46] https://drive.google.com/drive/folders/1LvVk7-ygqWT1VT-5BZ4HiMVP77KhgFvk?usp=sharing.

[47] https://drive.google.com/drive/folders/1LvVk7-ygqWT1VT-5BZ4HiMVP77KhgFvk?usp=sharing.

[48] https://drive.google.com/drive/folders/1LvVk7-ygqWT1VT-5BZ4HiMVP77KhgFvk?usp=sharing.

[49] https://github.com/iesl/TypeNet.

entity typing. TypeNet was created by manually using Freebase types [26] and the synsets of the WordNet hierarchy [52]. Another fine-grained dataset is FEW-NERD [47] a large-scale, publicly available[50] human-annotated few-shot NER dataset with 8 coarse-grained and 66 fine-grained entity types. FEW-NERD includes 188,238 sentences from Wikipedia (corresponding to 4,601,160 words).

7.1.2 NED datasets

Nine datasets were proposed for NED. KORE50[51] [75] is a dataset containing highly ambiguous entity mentions. The research community widely uses this corpus, and it is considered among the most challenging datasets for entity disambiguation. On average, each sentence contains only 14 words, where 3 of them represent mentions [22]. AIDA-CoNLL[52] [76] is based on CoNLL 2003. The annotation of this corpus was done manually using YAGO2 (detailed in Sect. 7.4). Two students disambiguated each mention. In case of conflict, the authors chose the right one. AQUAINT[53] [114] is a randomly selected and manually linked subset. MSNBC[54] [39] represents a subset containing two stories for each of ten categories (business, U.S. politics, entertainment, health, sports, tech and science, travel, TV news, U.S. news, and world news): the corpus was extracted on January 2, 2007. ACE2004[55] [146] is a corpus manually annotated using Amazon Mechanical Turk (AMT[56]). The accuracy of annotations was approximately 85%, and the authors manually corrected the annotations to increase their precision. WIKI and CWeb[57] [67] are two corpora that respectively contain 345 and 320 files constructed by sampling large publicly annotated corpora such as ClueWeb and Wikipedia. For constructing these corpora, the authors collected many annotated documents and retained only the most ambiguous documents. The authors use many thresholds for the indicator of difficulty. They opt for a bracket where the accuracy is the highest. TAC[58] [81] is a corpus extracted from English Wikipedia in October 2008. This corpus includes three kinds of entities, person, organisation and geo-political entities. N^3 news.de[59] is a real-world data set collected from 2009 to 2011 [120]. DBpedia Abstracts[60] [27] is a large, multilingual corpus generated from enriched Wikipedia data of annotated Wikipedia abstracts. In addition to English, DBpedia Abstracts contain six languages: Dutch, French, Spanish, Italian, Japanese [120]. T-REx[61] is a dataset annotated using Wikidata triples. [121].

To sum up, Table 6 presents all the corpora mentioned above. Some metrics, including the size of the corpus, its language, and the entities detected, are included. Some of the research works using the datasets are mentioned as well.

50 https://ningding97.github.io/fewnerd/.

51 http://apps.yovisto.com/labs/ner-benchmarks/.

52 https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/aida/downloads/.

53 https://catalog.ldc.upenn.edu/LDC2002T31.

54 http://cogcomp.cs.illinois.edu/page/resource_view/4.

55 https://catalog.ldc.upenn.edu/LDC2005T09.

56 https://www.mturk.com/.

57 https://dataverse.library.ualberta.ca/dataset.xhtml?persistentId=doi:10.7939/DVN/10968.

58 http://nlp.cs.qc.cuny.edu/kbp/2010/.

59 https://github.com/dice-group/n3-collection.

60 http://downloads.dbpedia.org/2015-04/ext/nlp/abstracts/.

61 https://github.com/hadyelsahar/RE-NLG-Dataset.

Table 6 NER dataset statistics

Name	Size	Language	Category	Entities	Work using it
CoNLL03	1393 newswires	English	Coarse	4 (PER, ORG, LOC, MISC)	[9, 44, 94, 155, 156, 158, 162, 187, 197]
RCV1	810,000 Reuters articles	English	Coarse	4 (PER, ORG, LOC, MISC)	[162]
KBP2015	944 documents	English/ Chinese/ Spanish	Coarse	4 (PER, ORG, GEO-politics, Facilities)	[187]
WNUT-2016	3856 tweets	English	Coarse	10 (PER, Company,facilities, etc.)	[105]
WNUT-2017	3295 documents	English	Coarse	6 (PERS, LOC, Coorporation, etc.)	[6]
MIMIC	1635 patient notes	English	Coarse	18 (name, ID, address, etc.)	[98]
i2b2 (2014 + 2016)	2304 patient notes	English	Coarse	18 (name, ID, address, etc.)	[44, 98]
WikiGold	145 Wikipedia articles	English	Coarse	4 (CoNLL format)	[83]
MUC-7	400 New York times articles	English	Coarse	3(PER, ORG, LOC)	[155]
CoNLL2003g	900 articles	German	Coarse	4 (PER, ORG, LOC, MISC)	[9, 155]

Table 6 continued

Name	Size	Language	Category	Entities	Work using it
CoNLL2002	714,505 lines	Spanish/ Dutch	Coarse	4 (PERS, ORG, LOC, MISC)	[9, 94, 155]
OntoNotes−5.0	2.5 millions words	English/ Chinese/ Arabic	Coarse	18 (PER, ORG, LOC, etc.)	[158]
Genia	2,000 MEDLINE Article	English	Coarse	Term boundaries, term classification, etc.	[129]
NCBI disease	793 PubMED citations	English	Coarse	6900 disease mentions	[129]
GemEval 2014	31,000 sentences (news + Wikipedia)	German	Coarse	4 (PER, LOC, ORG, MISC)	[156]
ANERcop	300 articles	Arabic	Coarse	4 (PER, LOC, ORG, MISC)	[94, 156]
WikiFbF	38 million entities	English	FGER	118 types	[2]
WikiFbT	46 million entities	English	FGER	1,115 types	[2]
Wiki-NDS	38 + 46 million	English	FGER	118 +1,115 types	[2]
1k-WFB-g	2420 mentions	English	FGER	117 types	[2]
Typenet	380 links	English	FGER	1081 + 860 types	[123]
FEW-NERD	188,238 sentences (4,601,160 words) Wikipedia	English	FGER	8 coarse-grained and 66 fine-grained entity types	[71, 79, 103, 108, 110, 111]

Table 7 NED dataset statistics

Name	Size	Language	Category	Entities	Work using it
KORE50	50 short sentences	English	-	3 ambiguous mentions per sentence	[22, 120]
AIDA-CoNLL	1393 articles	English	–	34,956 mentions	[29, 34, 92, 120, 144, 195]
AQUAINT	100 wikipedia articles	English	–	9300 topics	[29, 92, 120, 195]
MSNBC	2 stories per MSNBC topics	English	–	756 surface forms	[163, 195]
ACE2004	348 documents	English/ Arabic/Chinese	–	7 (PERS, ORG, LOC, etc.). forms	[29, 120, 144, 163, 195]
TAC	352 Wikipedia docs	English	–	3 (person, organisation and geo-political)	[29, 34, 144, 163]
N^3 news.de	53 documents	German	–	627 mentions	[120]
DBpedia Abstract	11 million documents	Multilingual (6 languages)	–	r 97 million entities mentions	[119, 120, 149]
T-Rex	4.65 million Wikipedia links	English	-	938,642 entities	[121]

7.2 Tools and ontology

Tools are the different systems developed for NER and NED. These systems can be used with adequate Python or Java libraries in order to automatically detect and disambiguate names with a few lines of code. *Ontology addresses questions of how entities are grouped into categories and which of these entities exist on the most fundamental level. Ontologists often try to determine what the categories or highest kinds are and how they form a system of categories that encompasses the classification of all entities. Commonly proposed categories include substances, properties, relations, states of affairs and events.*[62]

7.2.1 NER tools

Stanford NER[63] [55], is a Java package based on linear chain Conditional Random Field. The models were trained on a mixture of CoNLL, MUC-6, MUC-7 and ACE-named entity corpora. The basic required output tags are "PERSON", "LOCATION" and "ORGANIZATION" [83].
spaCy[64] is implemented in Python. No detailed information is presented related to its model. The related output tags include "PERSON", "LOC", "ORG", "GPE" etc. [83].
LingPipe[65] is implemented in Java and supports both rule-based NER and supervised training of a statistical model or more direct method like dictionary matching. The NER model was trained on the MUC 6 corpus. It is relatively slow but with higher accuracy. The output entity types are PERSON, LOCATION, ORGANISATION [83].
Natural Language Toolkit (NLTK)[66] [23] is a Python NLP toolkit heavily used in the research community. NLTK's NER is based on a supervised machine learning algorithm (Maximum Entropy Classifier), and it is trained on the ACE corpus. The output entities include PERSON, LOCATION, ORGANISATION [83].
OpenNLP[67] is a machine learning-based toolkit (developed with Java) for the processing of natural language text. It supports the most common NLP tasks, including named entity recognition (NER) [16].
FINET[68] [41] is a FGNER system handling short text (such as tweets or sentences). FINET generates candidate types (explicitly and implicitly mentioned types) using a sequence of multiple extractors and selects the most appropriate using word-sense disambiguation approaches. FINET is an unsupervised system relying on Wordnet and another knowledge base to generate training data.

7.3 NED tools

In this section, we focus on Gerbil for presenting the NED tools. Indeed, Gerbil is a benchmark gathering different tools on NED in order to compare them.

62 https://en.wikipedia.org/wiki/Ontology.
63 https://nlp.stanford.edu/software/CRF-NER.shtml.
64 https://spacy.io/.
65 http://aliasi.com/lingpipe.
66 https://www.nltk.org/.
67 https://opennlp.apache.org/.
68 https://www.uni-mannheim.de/dws/research/resources/software/finet/.

Gerbil [69][70] [182] is a framework dedicated to evaluating the dataset and tools proposed for semantic entity annotation (including NER, NED and EL). The aim of Gerbil is to provide an easy way to compare the results between the different state-of-the-art EL approaches. Gerbil relies on the framework proposed by Cornolti et al. [38] in order to propose six kinds of experiments: (1) D2KB (mapping a set of given entity mentions), (2) A2KB (an extension of D2KB by integrating disambiguation of the extracted mentions), (3) Sa2KB (also an extension of D2KB by integrating the score of the mention during the evaluation process), (4) C2KB (detecting entities in a given document), (5) Sc2KB (an extension of C2KB where a subset of entities is returned), and (6) Rc2KB (also an extension of C2KB returning a sorted list of resources from the entity set).

Gerbil includes 9 NER/NED systems: (1) Wikipedia Miner[71] [114], based on prior probabilities, context relatedness and quality, combined to classification, (2) DBpedia Spotlight[72] [113], using DBpedia and based on a vector representation with cosine similarity, (3) TagMe 2[73] [54], which recognises named entities by using link texts from Wikipedia (for disambiguation, it uses a link graph), (4) NERD-ML[74] [183], based on machine learning classification and on a CRF in order to extract and for disambiguate entities, (5) KEA NER/NED[75] [168] (based on a predetermined context, an n-gram analysis and a lookup of all potential DBpedia candidate entities for each n-gram), (6) WAT[76] [141] (a set of graph-based algorithms and SVM linear models for collective entity linking), (7) AGDISTIS [77] [181] (based on string similarity measures, a set of heuristic for handling co-referencing and on the graph-based HITS algorithm[78]), (8) Babelfy[79] [116] (based on random walks and a sub-graph algorithm for multilingual disambiguation by using BabelNet [127]), and (9) Dexter[80] [31] (an open-source entity disambiguation framework with several state-of-the-art disambiguation methods).

Gerbil integrates all the datasets used in order to train and evaluate the aforementioned systems (i.e. Wikipedia Miner,[81] DBpedia Spotlight,[82] TagMe 2,[83]NERD-ML[84] WAT,[85]

69 https://github.com/dice-group/gerbil.

70 http://aksw.org/Projects/GERBIL.html.

71 http://wikipedia-miner.cms.waikato.ac.nz/.

72 https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki/Web-service.

73 http://tagme.di.unipi.it/.

74 http://nerd.eurecom.fr/.

75 https://s16a.org/kea.

76 https://github.com/nopper/wat.

77 https://github.com/dice-group/AGDISTIS.

78 HITS is a link analysis algorithm that rates Web pages, developed by Jon Kleinberg.

79 http://babelfy.org/.

80 http://dexter.isti.cnr.it/.

81 https://tac.nist.gov//data/data_desc.html#AQUAINT.

82 https://wiki.dbpedia.org/spotlight/isemantics2011/evaluation.

83 http://acube.di.unipi.it/tagme-dataset/.

84 microposts2014 [28].

85 https://github.com/nopper/wat.

AGDISTIS,[86] Babelfy,[87],[88],[89],[90],[91],[92]). In addition to the systems' datasets, Gerbil includes ACE2004, IITB[93] (containing 103 documents and 11,250 entities), Meij[94] [166] (containing 502 tweets and 814 entities), MSNBC and N^{3}Reuters-128[95] and N^3 RSS-500[96] [148] (respectively containing 128 news/880 entities and 500 RSS-feeds/1000 entities).

7.4 Ontology

YAGO[97] [51] is an extensible ontology derived from Wikipedia WordNet and GeoNames. YAGO contains more than 1 million entities and 5 million facts, and it covers both entities and relations. The facts have been automatically extracted from Wikipedia and unified with WordNet, using a combination of rule-based and heuristic methods. The resulting knowledge base represents an improved WordNet, by adding knowledge about individuals like persons, organisations, products, etc., with their semantic relationships. In its original version, YAGO contains more than 1 million entities (like persons, organisations, cities, etc.) and contains more than 5 million facts (was born, wrote music for, etc.) about these entities. Different versions of YAGO have been proposed: YAGO, YAGO2 and YAGO3.
DBpedia[98] [14] is a multilingual knowledge base. For constructing it, structured information from Wikipedia such as categorisation information, links to external Web pages and geo-coordinates were extracted. The English version of the DBpedia knowledge base currently contains about 4,233,000 entities.
Freebase[99] [26] is a large online knowledge base created by its community members (collaboratively). Freebase has a friendly interface allowing all the users to interact with it. Freebase contains data extracted from many sources such as Wikipedia, and it includes more than 43 million entities and 2.4 billion facts [157].
MedMentions [100] [123] is a large dataset identifying and linking entity mentioned in PubMed abstracts[101] to specific UMLS[102] concepts. 246,000 UMLS entities were manually annotated using 3,704 mentions extracted from PubMed abstracts. The average depth in the hierarchy of a concept is 14.4 and the maximum depth is 43.

86 https://github.com/dice-group/n3-collection.

87 AIDA-CoNLL.

88 KORE 50.

89 SemEval-2007 task 17 [142].

90 SemEval-2007 Task 07 [126].

91 Semeval-2013 task 12 [125].

92 SENSEVAL-3 [164].

93 https://www.cse.iitb.ac.in/~soumen/doc/CSAW/.

94 http://nlp.uned.es/~damiano/datasets/entityProfiling_ORM_Twitter.html

95 https://old.datahub.io/dataset/reuters-128-nif-ner-corpus/resource/8d2ce805-e713-4010-8496-ea643ae07860.

96 http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html

97 https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/.

98 https://wiki.dbpedia.org/.

99 http://www.freebase.com/.

100 https://github.com/chanzuckerberg/MedMentions.

101 https://www.ncbi.nlm.nih.gov/pubmed/.

102 Unified Medical Language System (UMLS)[103] [25] is a hierarchical ontology containing over 3.5 million concepts.

Table 8 NER and NED tools

Tool	Type	Corpus used	Entities
Stanford NER[1] [55]	NER	CoNLL, MUC-6, MUC-7 and ACE	"PERSON", "LOCATION" and "ORGANIZATION"
spaCy[2]	NER	Not provided	“PERSON”, "LOC", "ORG", "GPE", etc.
LingPipe[3]	NER	MUC 6	"PERSON", "LOCATION", "ORGANISATION"
Natural Language Toolkit (NLTK)[4] [23]	NER	ACE	"PERSON", "LOCATION", "ORGANISATION"
OpenNLP[5]	NER	Not provided	"PERSON", "LOCATION", "ORGANISATION", "MISC"
FINET[6] [41]	NER	Wordnet	(16k types of organisations, persons and locations
Gerbil[7] [182]	NED	Wikipedia, DBpedia Spotlight, TagMe, NERD-ML, KEA NER/NED, WAT, AGDISTIS, Babelfy, BabelNet, Dexter	Not mentioned

[1] https://nlp.stanford.edu/software/CRF-NER.shtml
[2] https://spacy.io/
[3] http://aliasi.com/lingpipe
[4] https://www.nltk.org/
[5] https://opennlp.apache.org/
[6] https://www.uni-mannheim.de/dws/research/resources/software/finet/
[7] http://aksw.org/Projects/GERBIL.html

8 Synthesis and analysis

8.1 Review metrics

To sum up, we analysed 177 works in the presented survey. Ninety-seven works (including 42 research works and 56 resource descriptions) are described in detail. Of the 42 research papers, 5 are surveys, and 37 are solution works. Of the 32 studied research works, 24 focus on English and 13 on other languages, most of the time, including English, such as German and Spanish. Also, from the 24 works on English, 18 focus on NER while 6 focus on NED. From the 13 multilingual works, 7 focus on NER and 6 on NER. Hence, from the total number of the presented works (33 works), 20 works are on NER while 13 are on NED. Almost all studied works are recent (published up to 2023), and 35 of them are from 2015.

Concerning the resources, from the 56 resources described, 31 are corpora, and 25 are APIs and tools. From the 30 corpora, 21 works focus on NER, and 9 works focus on NED. Of the 25 APIs and tools, 6 focus on NER and 19 focus on NED. Of the 22 corpora dedicated to NER, 16 corpora are dedicated to English, 2 are multilingual (including English), and 4 corpora are dedicated to other languages such as German, Spanish, Arabic, etc. Of the 9 corpora dedicated to NED, 6 are in English, 2 are multilingual (including English), and the last one is dedicated to German. The APIs and tools are language-independent: they are usually trained on corpora in English or German, but users can train its model using a language of their choice.

Finally, we observe that the new tendency for extracting named entities and disambiguating them is to use neural networks. Of the 33 studied works, 16 use neural networks. 10 works from the 16 (in total) are in English, and 6 works are multilingual. The other works are using standard machine learning algorithms such as SVM, CRF (is the most dominant used algorithm), etc. For neural networks, almost all the works are based on CNN, LSTM, or Bi-LSTM. Other works rely on existing tools such as CoreNLP, OpenNLP and Gerbil.

8.2 Analytical synthesis

This section aims to answer the research questions presented in Sect. 4. Answering these questions allows us to conclude this survey by giving a general picture of the current situation of the research related to EL, NER and NED. Answering these questions will also highlight open issues related to the field of EL that require further research.

R1: After analysing the presented works, we conclude that the majority of the presented works aim to extract/recognise entities only (NER task). Disambiguation is only proposed by the most recent studies (the majority of them in 2018). Also, only a few systems can be considered to do entity linking (also called end-to-end systems), such as those of Cao et al. [29] or Bhatia et al. [22]. An EL system is a system providing both NER and NED. Almost all the proposed systems focused on one task only, by starting with a set of predefined named entities (for NED). Concerning the used algorithms, the tendency is to rely on neural networks (NNs). The originality of each work is related to the used architecture (word/character level/word + character level) and the type of NN used in each level (such as CNN, LSTM, or Bi-LSTM). We also observed that CRF is usually used with an NN architecture, either as a feature extractor or as a layer (replacing the softmax[104] function for generating the labels).

[104] It is a function that takes as input a vector of K real numbers and normalises it into a probability distribution consisting of K probabilities proportional to the exponential of the input numbers.

R2: It can be seen that almost all the constructed resources are publicly available. However, some of them (such as MUC-6) are published under the LDC non-free licence. However, the majority of them are free for research purposes. It is the main reason why almost all the recent works rely on the publicly available corpora. Only some recent research works focused on constructing their corpora. Among the freely available corpora, almost all the works presented in the research literature focus on the CoNLL corpora (including all its variants: CoNLL2003 for English, CoNLL2002 for Spanish and Dutch, and CoNLL2003 for German) for NER and CoNLL/AIDA for NED. The main problems with these corpora are their limited size (only 1,393 articles for CoNLL2003) and that they cannot be used in all domains. They were extracted from newspapers, so they cannot be used in medical or technical domains. Also, these corpora include only 4 entity types (PER, ORG, LOC and MISC), limiting them to coarse-grained NER.

R3: There are two types of approaches to constructing the annotated corpora: manual and automatic. Almost all the works on coarse recognition use corpora, which were constructed manually. Automatic construction is the most appropriate for fine-grained recognition, where the number of recognised entities is up to 118. Manual construction is time- and effort-consuming, but it produces corpora that give more accurate results. The corpora that are constructed automatically can cover more entities and more domains with less effort, but they suffer from a lack of precision. Almost all the corpora that were constructed automatically rely on ontologies such as Wikipedia or YAGO. Finally, a lack of real-life scenarios can also be observed: almost all the research using these corpora uses them in an intrinsic way where the training and the testing corpus are different parts of the same corpus. In practice, the trained tools have to be used on data outside the corpus.

R4: It can be concluded that the results presented for NER are more promising than those presented for NED. This is perfectly understandable, since NER consists of only extracting the different entities without disambiguating them. Also, the corpora used for NER are less challenging than those used for NED, where an entity could correspond to different types. The results related to coarse recognition are much better than those related to fine-grained recognition. It is also understandable that extracting four entities is less challenging than extracting 118. The approaches relying on neural networks give promising results compared to the works using classical machine learning approaches. These results are better where the NN models are associated with the word embedding used for feature extraction. However, almost all the presented works compare themselves to up to three other works. In most of the cases, the comparison is not made using a benchmark, which could compare the approach to many other systems on many corpora.

R5: Two main tendencies emerged from the multilingual approaches: works that propose language-independent approaches, and works that create parallel corpora across all the studied languages. For constructing these corpora, the presented research usually relies on ontologies. These corpora represent new resources, but the real added value is behind the language-independent approach, where the proposed system could be applied to different languages. We also conclude that almost all the language-independent systems work at the character level using NNs. However, the experiments show that in almost all cases, the results obtained on English are better than the results obtained in other languages.

R6: By answering to all the previous research questions, we provided a general overview of the research works recently proposed for EL/NER/NED. However, each one of the presented answers also raised a set of issues (open issues). From the R1, we conclude that end-to-end EL systems are rare: the majority of the works are focusing on NER or NED, but not both. Also, as almost all the systems focused on coarse-grained NER, more work is needed on fine-grained NER. A stronger focus on these kinds of systems would undoubtedly improve

the quality of the proposed systems. From R2, we conclude that the research literature should focus more on the construction of annotated corpora rather than using the same ones for all the proposed studies. As it has a lack of resources dedicated to fine-grained NER, constructing more resources is undoubtedly an open issue. Furthermore, considering the heterogeneity of the used corpora for training and testing would certainly present more realistic results. It is also essential to focus more on unstructured data provided from social media: the results on W-NUT (corpora constructed from Twitter) are low compared to the results obtained on CoNLL (which were constructed from newspapers). From R3, we conclude that both manual and automatic resource construction present advantages and disadvantages. Some research works on semi-automatic construction integrating deep active learning are ongoing. Focusing more on the semi-automatic construction where the corpus would be annotated automatically and reviewed manually presents an open line of work which could resolve the disadvantages of both techniques. It would be less time- and effort-consuming since the corpus is first constructed automatically, and achieve more precision because the corpus is reviewed manually.

The results of R4 confirm our previous assumptions. However, another important aspect is highlighted in this answer: lack of reliability in the comparison of the results. Some recent works are relying on benchmarks (such as **Gerbil**) to provide a fair comparison between the proposed approach results and the results presented in the research literature, but more works should rely on them. Another related open issue is the lack of benchmarks: **Gerbil** provides a reliable comparison framework, but it includes only a few systems. It is possible to add more systems into **Gerbil**, but the integrated systems need web APIs to be integrated. Finally from R5, we conclude that even by proposing a language-independent system, the research literature should focus more on the characteristics of each studied language where the results in English are better than the results in the other languages.

9 Comparison with the other surveys presented in the research literature

From 2007 to 2019, five other surveys related to entity linking were proposed. However, only one of them focused on both NER and NED ([17]). All the other ones were dedicated to NER. The presented paper focuses on all the works, resources, and tools that are related to entity linking by handling both NER and NED. Some statistics comparing our survey to the proposed surveys in the research literature are presented in Table 9. From this table, we first observe that the most recent survey was proposed in 2019. However, only 23 recent works from 2015 were described in this survey, which focuses on neural network models proposed for NER. Also, 15 datasets were presented without being described in detail. No information about the size, the construction technique, or the location of the presented datasets was given.

Concerning the number of the described works, almost all the surveys presented more works than ours. Nadeau et al. [124] described 60 works, Goyal et al. [63] described 48 and Patil et al. [136] described 43. However, almost all the described works in these surveys are not recent. No work after 2015 was presented by Nadeau et al. [124] and Patil et al. [136]. The main aim of our survey is to cover the most recent works proposed for entity linking. As almost all the works presented before 2015 were covered in the previous surveys, we saw no reason to survey them again. Mainly due to this reason, we described and detailed only 37 research works, where 31 (95%) of them are after 2015. In addition to the works that were detailed and classified, we also present some comparisons with works presented before 2015.

The most valuable resources behind each natural language processing problem (including NER and NED) are the datasets, tools and APIs. However, Table 9 shows that the previous surveys described only a few resources. In addition to Yadav et al. [188] who cite a set of corpora without providing any details, Goyal et al. [63] presented a table associating each presented work to the used dataset. However, these authors only give some statistics about the datasets, without detailing their construction approaches, or even the classes that the datasets are dealing with (person, location, etc.). The presented surveys also neglected the tools and APIs, where only 11 tools were presented by Balog et al. [17] and only 6 by Patil et al. [136]. In contrast to the above surveys, our work gives particular attention to the available resources (tools, API and datasets): we described, detailed and classified 55 resources (25 tools/APIs, and 30 datasets). In addition to describing these resources and classifying them, we also provide the location of each resource (as a Web link).

Same as almost all the other surveys, we classify the presented works, in the research literature, by distinguishing those dedicated to English (only) from those focusing on other languages. For both NER and NED, we present the works on English and multilingual works (focusing on more than one language). We also present in Table 7 the language of each constructed dataset to highlight the resources which were constructed in other languages than English. Finally, it can be seen from Table 9 that except for Yadav et al. [188], we are the only survey giving particular attention to the works using neural networks. In contrast to Yadav et al., we are focusing on the most recent works on NER and NED, which lead us to the new tendency to use neural network models.

10 Conclusion

The role of this survey was to present and classify the most recent studies that have been done on EL. However, it has been highlighted from the beginning that EL represents *the task of recognising entities mentioned in the text and linking them to the corresponding entries in a knowledge repository* [17]. Hence, we mainly focused on the papers that have been proposed for both NER and NED (where we focused on 43 research papers and 55 resources including corpora and tools). Studies focusing on English and also the ones proposed for other languages were considered. This survey focuses on the most recent studies where 95% of presented papers were published after 2015.

After analysing the studied papers, we concluded that the majority of the works focused on one only, either NER or NED and only a few works were dedicated to presenting the whole pipeline leading to EL. For the works focusing on NER, the coarse-grained NER approach is mainly used compared to the fine-grained NER where the majority of the studies focused on four entity types only (PER, ORG, LOC and MISC). This is mainly due to the scarcity of datasets dedicated to fine-grained NER. The most used corpora are the CoNLL corpora (including CoNLL2003 for English and German and CoNLL2002 for Spanish) focusing on the four entities mentioned earlier. Also, the coarse-grained NER returns more accurate results than the fine-grained NER, mainly due to the way used for constructing the corpora. Indeed, almost all the corpora used for the coarse-grained NER were constructed manually, whereas almost all the corpora used for fine-grained NER were constructed automatically by relying on ontologies such as Wikipedia or YAGO. Due to the lack of diversity related to the constructed resources, only a few studies were multilingual were the majority of the papers focused on English.

Table 9 Comparison with other surveys

Survey	[17]	[63]	[124]	[188]	[136]	The presented survey
Year	2018	2018	2017	2019	2016	2023
Number of references	87	191	85	83	51	200
Number of described works	29	48	60	31	43	38
Number of work after 2015	30	13	0	23	0	36
Number of described tools and API	11	0	0	0	6	25
Number of described datasets	13	48	4	15	3	31
Number of approaches using neural networks	5	1	0	21	0	16
Focuses on multilingual approaches	No	Yes	Yes	Yes	Yes	Yes

Moreover, it was also concluded that the systems proposed for NER are much more promising than those presented for NED where disambiguating entities are much more complex than recognising them. However, the lack of reliable benchmarks is the most important issue related to a fair comparison among the different proposed studies. almost all the presented works compare themselves to up to three other works of their choice (presenting the most similar approaches). This solution is strong enough when it is applied to CoNLL corpora, as the majority of the studies used them for result comparison. However, when a corpus is used only by a few studies, it is difficult to conclude a strong and fair comparison. More benchmarks such as **Gerbil** should be constructed to provide this reliable comparison among frameworks.

Finally, we observed that the latest tendency regarding entity linking is to apply it to medical data where there are events that trigger a sudden increase in the number of publications, such as the COVID-19 pandemic. For instance, PubMed added new 952,919 citations only in 2020. In this context, new word embedding models (such as BioBERT [97], PubMedBERT [64] and SciBERT [18]) are pre-trained on large biomedical corpora through unsupervised tasks and then fine-tuned [65, 66] for specific tasks, including EL. These models achieved state-of-the-art performance in several EL benchmarks [150].

Funding The research reported in this paper was funded by the Innovate UK-funded Knowledge Transfer Partnership (ref. KTP011673) between Aston University and Folding Space.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Human participants and/or animals No human participants or animals were involved in the presented study.

References

1. Abhishek A, Anand A, Awekar A (2017) Fine-grained entity type classification by jointly learning representations and label embeddings. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. Association for Computational Linguistics, Valencia, Spain, pp 797–807. http://www.aclweb.org/anthology/E17-1075
2. Abhishek A, Taneja SB, Malik G, Anand A, Awekar A (2019) Fine-grained entity recognition with reduced false negatives and large type coverage. In: Proceedings of the 1st conference of the automated knowledge base construction. Automated Knowledge Base Construction, Amherst, USA. https://openreview.net/forum?id=HylHE-9p6m
3. Agerri R, Rigau G (2016) Robust multilingual named entity recognition with shallow semi-supervised features. Artif Intell 238:63–82
4. Aggarwal N, Buitelaar P (2014) Wikipedia-based distributional semantics for entity relatedness. In: 2014 AAAI Fall Symposium Series
5. Aguilar G, López-Monroy AP, González FA, Solorio T (2019) Modeling noisiness to recognize named entities using multitask neural networks on social media. arXiv:1906.04129

6. Aguilar G, Maharjan S, López-Monroy AP, Solorio T (2019) A multi-task approach for named entity recognition in social media data. arXiv:1906.04135
7. Akbik A, Bergmann T, Vollgraf R (2019) Pooled contextualized embeddings for named entity recognition. In: Proceedings of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp 724–728
8. Akbik A, Blythe D, Vollgraf R (2018) Contextual string embeddings for sequence labeling. In: Proceedings of the 27th International Conference on Computational Linguistics, pp 1638–1649
9. Al-Rfou R, Kulkarni V, Perozzi B, Skiena S (2015) Polyglot-ner: Massive multilingual named entity recognition. In: Proceedings of the 2015 SIAM international conference on data mining. SIAM, pp 586–594
10. Al-Rfou R, Perozzi B, Skiena S (2013) Polyglot: distributed word representations for multilingual nlp. arXiv:1307.1662
11. Alhelbawy A, Gaizauskas R (2014) Graph ranking for collective named entity disambiguation. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp 75–80
12. Alotaibi F, Lee M (2014) A hybrid approach to features representation for fine-grained arabic named entity recognition. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp 984–995
13. Andor D, Alberti C, Weiss D, Severyn A, Presta A, Ganchev K, Petrov S, Collins M (2016) Globally normalized transition-based neural networks. arXiv:1603.06042
14. Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z (2007) Dbpedia: a nucleus for a web of open data. In: The semantic web. Springer, pp 722–735
15. Baevski A, Edunov S, Liu Y, Zettlemoyer L, Auli M (2019) Cloze-driven pretraining of self-attention networks. arXiv:1903.07785
16. Baldridge J (2005) The opennlp project. http://opennlp.apache.org/index.html. Accessed 2 Feb 2012
17. Balog K (2018) Entity-oriented search. Springer, New York
18. Beltagy I, Lo K, Cohan A (2019) Scibert: a pretrained language model for scientific text. arXiv:1903.10676
19. Benajiba Y, Rosso P (2008) Arabic named entity recognition using conditional random fields. In: Proceedings of Workshop on HLT & NLP within the Arabic World, LREC, vol 8. Citeseer, pp 143–153
20. Benajiba Y, Rosso P, Benedíruiz JM (2007) Anersys: An arabic named entity recognition system based on maximum entropy. In: International conference on intelligent text processing and computational linguistics. Springer, pp 143–153
21. Benikova D, Biemann C, Kisselew M, Pado S (2014) Germeval 2014 named entity recognition shared task: companion paper
22. Bhatia S (2019) Entity linking in enterprise search: combining textual and structural information. In: Linking and mining heterogeneous and multi-view data. Springer, pp 183–199
23. Bird S, Klein E, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc."
24. Blum A, Mitchell T (1998) Combining labeled and unlabeled data with co-training. In: Proceedings of the eleventh annual conference on Computational learning theory, pp 92–100
25. Bodenreider O (2004) The unified medical language system (umls): integrating biomedical terminology. Nucleic Acids Res 32(suppl-1):D267–D270
26. Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J (2008) Freebase: a collaboratively created graph database for structuring human knowledge. In: Proceedings of the 2008 ACM SIGMOD international conference on management of data. AcM, pp 1247–1250
27. Brümmer M, Dojchinovski M, Hellmann S (2016) Dbpedia abstracts: A large-scale, open, multilingual nlp training corpus. In: Proceedings of the tenth international conference on Language Resources and Evaluation (LREC'16), pp 3339–3343
28. Cano AE, Rizzo G, Varga A, Rowe M, Stankovic M, Dadzie AS (2014) Making sense of microposts:(# microposts2014) named entity extraction & linking challenge. In: CEUR workshop proceedings, vol 1141, pp 54–60
29. Cao Y, Hou L, Li J, Liu Z (2018) Neural collective entity linking. arXiv:1811.08603
30. Castelli V, Raghavan H, Florian R, Han DJ, Luo X, Roukos S (2012) Distilling and exploring nuggets from a corpus. In: Proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval. ACM, pp 1006–1006
31. Ceccarelli D, Lucchese C, Orlando S, Perego R, Trani S (2013) Dexter: an open source framework for entity linking. In: Proceedings of the sixth international workshop on exploiting semantic annotations in information retrieval. ACM, pp 17–20

32. Cheng X, Roth D (2013) Relational inference for wikification. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp 1787–1796
33. Chinchor N, Robinson P (1997) Muc-7 named entity task definition. In: Proceedings of the 7th conference on message understanding, vol 29, pp 1–21
34. Chisholm A, Hachey B (2015) Entity disambiguation with web links. Trans Assoc Comput Linguist 3:145–156
35. Chiu JP, Nichols E (2016) Named entity recognition with bidirectional lstm-cnns. Trans Assoc Comput Linguist 4:357–370
36. Clark K, Luong MT, Manning CD, Le QV (2018) Semi-supervised sequence modeling with cross-view training. arXiv:1809.08370
37. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P (2011) Natural language processing (almost) from scratch. J Mach Learn Res 12(Aug):2493–2537
38. Cornolti M, Ferragina P, Ciaramita M (2013) A framework for benchmarking entity-annotation systems. In: Proceedings of the 22nd international conference on World Wide Web. ACM, pp 249–260
39. Cucerzan S (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 (EMNLP-CoNLL), pp 708–716
40. Darwish K (2013) Named entity recognition using cross-lingual resources: Arabic as an example. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp 1558–1567
41. Del Corro L, Abujabal A, Gemulla R, Weikum G (2015) Finet: context-aware fine-grained named entity typing. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp 868–878
42. Delpeuch A (2019) Opentapioca: lightweight entity linking for wikidata. arXiv:1904.09131
43. Derczynski L, Nichols E, van Erp M, Limsopatham N (2017) Results of the wnut2017 shared task on novel and emerging entity recognition. In: Proceedings of the 3rd workshop on noisy user-generated text, pp 140–147
44. Dernoncourt F, Lee JY, Szolovits P (2017) Neuroner: an easy-to-use program for named-entity recognition based on neural networks. arXiv:1705.05487
45. Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805
46. Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesot, pp 4171–4186. https://doi.org/10.18653/v1/N19-1423
47. Ding N, Xu G, Chen Y, Wang X, Han X, Xie P, Zheng HT, Liu Z (2021) Few-nerd: a few-shot named entity recognition dataset. arXiv:2105.07464
48. Doğan RI, Lu Z (2012) An improved corpus of disease mentions in pubmed citations. In: Proceedings of the 2012 workshop on biomedical natural language processing. Association for Computational Linguistics, pp 91–99
49. Dojchinovski M, Kliegr T (2012) Recognizing, classifying and linking entities with wikipedia and dbpedia. In: Workshop on intelligent and knowledge oriented technologies (WIKT), pp 41–44
50. Eshel Y, Cohen N, Radinsky K, Markovitch S, Yamada I, Levy O (2017) Named entity disambiguation for noisy text. arXiv:1706.09147
51. Fabian M, Gjergji K, Gerhard W (2007) et al.: Yago: A core of semantic knowledge unifying wordnet and wikipedia. In: 16th International World Wide Web conference, WWW, pp 697–706
52. Fellbaum C (1998) Wordnet: Wiley online library. The Encyclopedia of Applied Linguistics
53. Ferragina P, Scaiella U (2010) Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In: Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, pp 1625–1628
54. Ferragina P, Scaiella U (2011) Fast and accurate annotation of short texts with wikipedia pages. IEEE Softw 29(1):70–75
55. Finkel JR, Grenager T, Manning C (2005) Incorporating non-local information into information extraction systems by Gibbs sampling. In: Proceedings of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics, pp 363–370
56. Florian R, Ittycheriah A, Jing H, Zhang T (2003) Named entity recognition through classifier combination. In: Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4. Association for Computational Linguistics, pp 168–171

57. Francis-Landau M, Durrett G, Klein D (2016) Capturing semantic similarity for entity linking with convolutional neural networks. arXiv:1604.00734
58. Freitas C, Carvalho P, Gonçalo Oliveira H, Mota C, Santos D (2010) Second harem: advancing the state of the art of named entity recognition in portuguese. In: quot; In: Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, Daniel Tapias (eds) Proceedings of the International Conference on Language Resources and Evaluation (LREC 2010)(Valletta 17-23 May de 2010) European Language Resources Association. European Language Resources Association
59. Ganea OE, Ganea M, Lucchi A, Eickhoff C, Hofmann T (2016) Probabilistic bag-of-hyperlinks model for entity linking. In: Proceedings of the 25th international conference on World Wide Web. International World Wide Web Conferences Steering Committee, pp 927–938
60. Ganea OE, Hofmann T (2017) Deep joint entity disambiguation with local neural attention. arXiv:1704.04920
61. Globerson A, Lazic N, Chakrabarti S, Subramanya A, Ringaard M, Pereira F (2016) Collective entity resolution with multi-focal attention. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp 621–631
62. Godin F, Vandersmissen B, De Neve W, Van de Walle R (2015) Multimedia lab@ acl wnut ner shared task: Named entity recognition for twitter microposts using distributed word representations. In: Proceedings of the workshop on noisy user-generated text, pp 146–153
63. Goyal A, Gupta V, Kumar M (2018) Recent named entity recognition and classification techniques: a systematic review. Comput Sci Rev 29:21–43
64. Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, Naumann T, Gao J, Poon H (2020) Domain-specific language model pretraining for biomedical natural language processing
65. Guellil I, Chenni N, Berrachedi Y, Abboud MN, Wu J, Wu H, Alex B (2022) Detecting adverse drug events from social media: A brief literature review. In: The sixth widening NLP workshop: proceedings of the workshop. ACL Anthology
66. Guellil I, Wu J, Wu H, Sun T, Alex B (2022) Edinburgh_ucl_health@ smm4h'22: From glove to flair for handling imbalanced healthcare corpora related to adverse drug events, change in medication and self-reporting vaccination. In: Proceedings of COLING. International conference on computational Linguistics, vol 2022. Europe PMC Funders, p 148
67. Guo Z, Barbosa D (2018) Robust named entity disambiguation with random walks. Semantic Web 9(4):459–479
68. Gupta N, Singh S, Roth D (2017) Entity linking via joint encoding of types, descriptions, and context. In: Proceedings of the 2017 conference on empirical methods in Natural Language Processing, pp 2681–2690
69. Gurevych I, Zamorani NC, Kim J (2012) Proceedings of the 3rd workshop on the people's web meets nlp: Collaboratively constructed semantic resources and their applications to nlp. In: Proceedings of the 3rd workshop on the People's Web Meets NLP: collaboratively constructed semantic resources and their applications to NLP
70. Hänig C, Thomas S, Bordag S (2014) Modular classifier ensemble architecture for named entity recognition on low resource systems
71. He K, Mao R, Huang Y, Gong T, Li C, Cambria E (2023) Template-free prompting for few-shot named entity recognition via semantic-enhanced contrastive learning. IEEE Trans Neural Netw Learn Syst
72. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, pp 1026–1034
73. He Z, Liu S, Li M, Zhou M, Zhang L, Wang H (2013) Learning entity representation for entity disambiguation. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp 30–34
74. Helwe C, Elbassuoni S (2019) Arabic named entity recognition via deep co-learning. Artif Intell Rev 52(1):197–215
75. Hoffart J, Seufert S, Nguyen DB, Theobald M, Weikum G (2012) Kore: keyphrase overlap relatedness for entity disambiguation. In: Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, pp 545–554
76. Hoffart J, Yosef MA, Bordino I, Fürstenau H, Pinkal M, Spaniol M, Taneva B, Thater S, Weikum G (2011) Robust disambiguation of named entities in text. In: Proceedings of the conference on empirical methods in Natural Language Processing. Association for Computational Linguistics, pp 782–792
77. Hoffmann R, Zhang C, Ling X, Zettlemoyer L, Weld DS (2011) Knowledge-based weak supervision for information extraction of overlapping relations. In: Proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, pp 541–550

78. Houlsby N, Ciaramita M (2014) A scalable gibbs sampler for probabilistic entity linking. In: European conference on information retrieval. Springer, pp 335–346
79. Huang Y, He K, Wang Y, Zhang X, Gong T, Mao R, Li C (2022) Copner: Contrastive learning with prompt guiding for few-shot named entity recognition. In: Proceedings of the 29th international conference on computational linguistics, pp 2515–2527
80. Huang Z, Xu W, Yu K (2015) Bidirectional lstm-crf models for sequence tagging. arXiv:1508.01991
81. Ji H, Grishman R, Dang HT, Griffitt K, Ellis J (2010) Overview of the tac 2010 knowledge base population track. In: Third text analysis conference (TAC 2010), vol 3, p 3
82. Ji H, Nothman J, Hachey B, Florian R (2015) Overview of tac-kbp2015 tri-lingual entity discovery and linking. In: TAC
83. Jiang R, Banchs RE, Li H (2016) Evaluating and combining name entity recognition systems. In: Proceedings of the sixth named entity workshop, pp 21–27
84. Jiang Y, Hu C, Xiao T, Zhang C, Zhu J (2019) Improved differentiable architecture search for language modeling and named entity recognition. In: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), pp 3576–3581
85. Jin H, Dong T, Hou L, Li J, Chen H, Dai Z, Yincen Q (2022) How can cross-lingual knowledge contribute better to fine-grained entity typing? In: Findings of the Association for Computational Linguistics: ACL 2022, pp 3071–3081
86. Jin H, Hou L, Li J, Dong T (2018) Attributed and predictive entity embedding for fine-grained entity typing in knowledge bases. In: Proceedings of the 27th international conference on computational linguistics, pp 282–292
87. Jin H, Hou L, Li J, Dong T (2019) Fine-grained entity typing via hierarchical multi graph convolutional networks. In: Proceedings of the 2019 conference on empirical methods in Natural Language Processing and the 9th International joint conference on Natural Language Processing (EMNLP-IJCNLP), pp 4970–4979
88. Khashabi D, Sammons M, Zhou B, Redman T, Christodoulopoulos C, Srikumar V, Rizzolo N, Ratinov L, Luo G, Do Q, et al (2018) Cogcompnlp: your swiss army knife for nlp. In: Proceedings of the eleventh international conference on Language Resources and Evaluation (LREC 2018)
89. Kim JD, Ohta T, Tateisi Y, Tsujii J (2003) Genia corpus-a semantically annotated corpus for bio-textmining. Bioinformatics 19(suppl-1):i180–i182
90. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv:1412.6980
91. Kiselev VY, Andrews TS, Hemberg M (2019) Challenges in unsupervised clustering of single-cell rna-seq data. Nat Rev Genet 20(5):273–282
92. Kolitsas N, Ganea OE, Hofmann T (2018) End-to-end neural entity linking. arXiv:1808.07699
93. Kulkarni S, Singh A, Ramakrishnan G, Chakrabarti S (2009) Collective annotation of wikipedia entities in web text. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 457–466
94. Kuru O, Can OA, Yuret D (2016) Charner: character-level named entity recognition. In: Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical Papers, pp 911–921
95. Lal A, Tomer A, Chowdary CR (2017) Sane: system for fine grained named entity typing on textual data. In: Proceedings of the 26th international conference on World Wide Web companion. International World Wide Web Conferences Steering Committee, pp 227–230
96. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C (2016) Neural architectures for named entity recognition. arXiv:1603.01360
97. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J (2020) Biobert: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 36(4):1234–1240
98. Lee JY, Dernoncourt F, Szolovits P (2017) Transfer learning for named-entity recognition with neural networks. arXiv:1705.06273
99. Lee K, He L, Lewis M, Zettlemoyer L (2017) End-to-end neural coreference resolution. arXiv:1707.07045
100. Leitner E, Rehm G, Moreno-Schneider J (2019) Fine-grained named entity recognition in legal documents. In: International conference on semantic systems. Springer, pp 272–287
101. Lewis DD, Gale WA (1994) A sequential algorithm for training text classifiers. In: SIGIR'94. Springer, pp 3–12
102. Lewis DD, Yang Y, Rose TG, Li F (2004) Rcv1: a new benchmark collection for text categorization research. J Mach Learn Res 5(Apr):361–397

103. Li W, Li H, Ge J, Zhang L, Li L, Wu B (2023) Cdaner: Contrastive learning with cross-domain attention for few-shot named entity recognition. In: 2023 international joint conference on neural networks (IJCNN). IEEE, pp 1–8
104. Liao L, He X, Zhang H, Chua TS (2018) Attributed social network embedding. IEEE Trans Knowl Data Eng 30(12):2257–2270
105. Limsopatham N, Collier NH (2016) Bidirectional lstm for named entity recognition in twitter messages
106. Ling X, Singh S, Weld DS (2015) Design challenges for entity linking. Trans Assoc Comput Linguist 3:315–328
107. Ling X, Weld DS (2012) Fine-grained entity recognition. In: Twenty-sixth AAAI conference on artificial intelligence
108. Liu C, Zhao F, Kang Y, Zhang J, Zhou X, Sun C, Wu F, Kuang K (2023) Rexuie: a recursive method with explicit schema instructor for universal information extraction. arXiv:2304.14770
109. Luong MT, Pham H, Manning CD (2015) Effective approaches to attention-based neural machine translation. arXiv:1508.04025
110. Ma J, Ballesteros M, Doss S, Anubhai R, Mallya S, Al-Onaizan Y, Roth D (2022) Label semantics for few shot named entity recognition. arXiv:2203.08985
111. Ma T, Jiang H, Wu Q, Zhao T, Lin CY (2022) Decomposed meta-learning for few-shot named entity recognition. arXiv:2204.05751
112. Ma X, Hovy E (2016) End-to-end sequence labeling via bi-directional lstm-cnns-crf. arXiv:1603.01354
113. Mendes PN, Jakob M, García-Silva A, Bizer C (2011) Dbpedia spotlight: shedding light on the web of documents. In: Proceedings of the 7th international conference on semantic systems. ACM, pp 1–8
114. Milne D, Witten IH (2008) Learning to link with wikipedia. In: Proceedings of the 17th ACM conference on Information and knowledge management. ACM, pp 509–518
115. Mohit B, Schneider N, Bhowmick R, Oflazer K, Smith NA (2012) Recall-oriented learning of named entities in arabic wikipedia. In: Proceedings of the 13th conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, pp 162–173
116. Moro A, Cecconi F, Navigli R (2014) Multilingual word sense disambiguation and entity linking for everybody. In: International Semantic Web Conference (Posters & Demos), pp 25–28
117. Moro A, Raganato A, Navigli R (2014) Entity linking meets word sense disambiguation: a unified approach. Trans Assoc Comput Linguist 2:231–244
118. Moussallem D, Ferreira TC, Zampieri M, Cavalcanti MC, Xexéo G, Neves M, Ngomo ACN (2018) Rdf2pt: Generating Brazilian Portuguese texts from rdf data. arXiv:1802.08150
119. Moussallem D, Usbeck R, Röder M, Ngomo ACN (2018) Entity linking in 40 languages using mag. In: European Semantic Web conference. Springer, pp 176–181
120. Moussallem D, Usbeck R, Röeder M, Ngomo ACN (2017) Mag: A multilingual, knowledge-base agnostic and deterministic entity linking approach. In: Proceedings of the knowledge capture conference. ACM, p 9
121. Mulang IO, Singh K, Vyas A, Shekarpour S, Sakor A, Vidal ME, Auer S, Lehmann J (2019) Context-aware entity linking with attentive neural networks on wikidata knowledge graph. arXiv:1912.06214
122. Murphy KP, Weiss Y, Jordan MI (1999) Loopy belief propagation for approximate inference: An empirical study. In: Proceedings of the fifteenth conference on uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc, pp 467–475
123. Murty S, Verga P, Vilnis L, Radovanovic I, McCallum A (2018) Hierarchical losses and new resources for fine-grained entity typing and linking. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp 97–109
124. Nadeau D, Sekine S (2007) A survey of named entity recognition and classification. Lingvisticae Investigationes 30(1):3–26
125. Navigli R, Jurgens D, Vannella D (2013) Semeval-2013 task 12: Multilingual word sense disambiguation. In: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pp 222–231
126. Navigli R, Litkowski KC, Hargraves O (2007) Semeval-2007 task 07: Coarse-grained english all-words task. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), pp 30–35
127. Navigli R, Ponzetto SP (2012) Babelnet: the automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artif Intell 193:217–250
128. Neelakantan A, Chang MW (2015) Inferring missing entity type instances for knowledge base completion: New dataset and methods. arXiv:1504.06658
129. Neelakantan A, Collins M (2015) Learning dictionaries for named entity recognition using minimal supervision. arXiv:1504.06650

130. Ngomo ACN, Röder M, Moussallem D, Usbeck R, Speck R (2018) Bengal: An automatic benchmark generator for entity recognition and linking. In: Proceedings of the 11th international conference on Natural Language Generation, pp 339–349
131. Nickel M, Tresp V, Kriegel HP (2011) A three-way model for collective learning on multi-relational data. In: ICML, vol 11, pp 809–816
132. Nothman J, Ringland N, Radford W, Murphy T, Curran JR (2013) Learning multilingual named entity recognition from wikipedia. Artif Intell 194:151–175
133. Nousi P, Tzelepi M, Passalis N, Tefas A (2022) Chapter 7 - lightweight deep learning. In: A. Iosifidis, A. Tefas (eds.) Deep Learning for Robot Perception and Cognition. Academic Press, pp. 131–164. https://doi.org/10.1016/B978-0-32-385787-1.00012-9
134. Oramas S, Ferraro A, Correya AA, Serra X (2017) Mel: a music entity linking system. In: Hu X, Cunningham SJ, Turnbull D, Duan Z (eds) ISMIR 2017. 18th International Society for Music Information Retrieval Conference; 2017 Oct 23-27; Suzhou, China [Canada]: ISMIR; 2017
135. Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: bringing order to the web. Tech. rep, Stanford InfoLab
136. Patil N, Patil AS, Pawar B (2016) Survey of named entity recognition systems with respect to indian and foreign languages. Int J Comput Appl 134(16)
137. Peng H, Roth D (2016) Two discourse driven language models for semantics. arXiv:1606.05679
138. Pennington J, Socher R, Manning CD (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp 1532–1543
139. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. arXiv:1802.05365
140. Phang J, Févry T, Bowman SR (2018) Sentence encoders on stilts: Supplementary training on intermediate labeled-data tasks. arXiv:1811.01088
141. Piccinno F, Ferragina P (2014) From tagme to wat: a new entity annotator. In: Proceedings of the first international workshop on Entity recognition & disambiguation. ACM, pp 55–62
142. Pradhan S, Loper E, Dligach D, Palmer M (2007) Semeval-2007 task-17: English lexical sample, srl and all words. In: Proceedings of the fourth international workshop on semantic evaluations (SemEval-2007), pp 87–92
143. Pradhan S, Moschitti A, Xue N, Uryupina O, Zhang Y (2012) Conll-2012 shared task: Modeling multilingual unrestricted coreference in ontonotes. In: Joint conference on EMNLP and CoNLL-Shared Task. Association for Computational Linguistics, pp 1–40
144. Raiman JR, Raiman OM (2018) Deeptype: multilingual entity linking by neural type system evolution. In: Thirty-second AAAI conference on artificial intelligence
145. Ramuhalli P, Udpa L, Udpa SS (2005) Finite-element neural networks for solving differential equations. IEEE Trans Neural Netw 16(6):1381–1392
146. Ratinov L, Roth D, Downey D, Anderson M (2011) Local and global algorithms for disambiguation to wikipedia. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, pp 1375–1384
147. Reimers N, Eckle-Kohler J, Schnober C, Kim J, Gurevych I (2014) Germeval-2014: nested named entity recognition with neural networks
148. Röder M, Usbeck R, Hellmann S, Gerber D, Both A (2014) N^3-a collection of datasets for named entity recognition and disambiguation in the nlp interchange format. In: LREC, pp 3529–3533
149. Rosales-Méndez H, Hogan A, Poblete B (2018) Voxel: a benchmark dataset for multilingual entity linking. In: International semantic Web conference. Springer, pp 170–186
150. Ruas P, Couto FM (2022) Nilinker: attention-based approach to nil entity linking. J Biomed Inform 132:104137
151. Sakor A, Mulang IO, Singh K, Shekarpour S, Vidal ME, Lehmann J, Auer S (2019) Old is gold: linguistic driven approach for entity and relation linking of short text. In: Proceedings of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp 2336–2346
152. Sang EF, De Meulder F (2003) Introduction to the conll-2003 shared task: Language-independent named entity recognition. arXiv:preprintcs/0306050
153. Sang TK (2002) Erik. f. 2002. introduction to the conll-2002 shared task: Language-independent named entity recognition. In: Proceedings of Conference on Natural Language Learning
154. Sasaki F, Dojchinovski M, Nehring J (2016) Chainable and extendable knowledge integration web services. In: International Semantic Web Conference. Springer, pp 89–101

155. Seyler D, Dembelova T, Del Corro L, Hoffart J, Weikum G (2018) A study of the importance of external knowledge in the named entity recognition task. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp 241–246
156. Shao Y, Hardmeier C, Nivre J (2016) Multilingual named entity recognition using hybrid neural networks. In: The Sixth Swedish Language Technology Conference (SLTC)
157. Shen W, Wang J, Han J (2014) Entity linking with a knowledge base: Issues, techniques, and solutions. IEEE Trans Knowl Data Eng 27(2):443–460
158. Shen Y, Yun H, Lipton ZC, Kronrod Y, Anandkumar A (2017) Deep active learning for named entity recognition. arXiv:1707.05928
159. Shi X, Knight K, Ji H (2014) How to speak a language without knowing it. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp 278–282
160. Shimaoka S, Stenetorp P, Inui K, Riedel S (2016) An attentive neural architecture for fine-grained entity type classification. arXiv:1604.05525
161. Shimaoka S, Stenetorp P, Inui K, Riedel S (2016) Neural architectures for fine-grained entity type classification. arXiv:1606.01341
162. Sienčnik SK (2015) Adapting word2vec to named entity recognition. In: Proceedings of the 20th nordic conference of computational linguistics, nodalida 2015, May 11–13, 2015, Vilnius, Lithuania, 109. Linköping University Electronic Press, pp 239–243
163. Sil A, Florian R (2017) One for all: towards language independent named entity linking. arXiv:1712.01797
164. Snyder B, Palmer M (2004) The English all-words task. In: Proceedings of SENSEVAL-3, the third international workshop on the evaluation of systems for the semantic analysis of text
165. Soto AR, Gallardo JJ, Diz AB (2017) Adapting Simplenlg to Spanish. In: Proceedings of the 10th international conference on natural language generation, pp 144–148
166. Spina D, Meij E, Oghina A, Bui MT, Breuss M, de Rijke M, et al (2012) A corpus for entity profiling in microblog posts. In: LREC workshop on language engineering for online reputation management
167. Spitkovsky VI, Chang AX (2012) A cross-lingual dictionary for English wikipedia concepts
168. Steinmetz N, Sack H (2013) Semantic multimedia information retrieval based on contextual descriptions. In: Extended Semantic Web Conference. Springer, pp 382–396
169. Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S, Tsujii J (2012) Brat: a web-based tool for nlp-assisted text annotation. In: Proceedings of the demonstrations at the 13th conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, pp 102–107
170. Stolcke A (2002) Srilm-an extensible language modeling toolkit. In: Seventh international conference on spoken language processing
171. Strassel S, Tracey J (2016) Lorelei language packs: data, tools, and resources for technology development in low resource languages. In: Proceedings of the tenth international conference on language resources and evaluation (LREC'16), pp 3273–3280
172. Strauss B, Toma B, Ritter A, De Marneffe MC, Xu W (2016) Results of the wnut16 named entity recognition shared task. In: Proceedings of the 2nd workshop on noisy user-generated text (WNUT), pp 138–144
173. Stubbs A, Kotfila C, Uzuner Ö (2015) Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/uthealth shared task track 1. J Biomed Inform 58:S11–S19
174. Sun Y, Lin L, Tang D, Yang N, Ji Z, Wang X (2015) Modeling mention, context and entity with neural networks for entity disambiguation. In: Twenty-fourth international joint conference on artificial intelligence
175. Tang J, Qu M, Mei Q (2015) Pte: Predictive text embedding through large-scale heterogeneous text networks. In: Proceedings of the 21th ACM SIGKDD International conference on knowledge discovery and data mining. ACM, pp 1165–1174
176. Taylor WL (1953) "cloze procedure": A new tool for measuring readability. Journal Q 30(4):415–433
177. Tong H, Faloutsos C, Pan JY (2006) Fast random walk with restart and its applications. In: Sixth international conference on data mining (ICDM'06). IEEE, pp 613–622
178. Topaz M, Murga L, Gaddis KM, McDonald MV, Bar-Bachar O, Goldberg Y, Bowles KH (2019) Mining fall-related information in clinical notes: comparison of rule-based and novel word embedding-based machine learning approaches. J Biomed Inform 90:103103
179. Toutanova K, Manning CD (2000) Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In: Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13. Association for Computational Linguistics, pp 63–70

180. Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G (2016) Complex embeddings for simple link prediction. In: International conference on machine learning, pp 2071–2080
181. Usbeck R, Ngomo ACN, Röder M, Gerber D, Coelho SA, Auer S, Both A (2014) Agdistis-graph-based disambiguation of named entities using linked data. In: International semantic web conference. Springer, pp 457–471
182. Usbeck R, Röder M, Ngonga Ngomo AC, Baron C, Both A, Brümmer M, Ceccarelli D, Cornolti M, Cherix D, Eickmann B (2015) et al.: Gerbil: general entity annotator benchmarking framework. In: Proceedings of the 24th international conference on World Wide Web. International World Wide Web Conferences Steering Committee, pp 1133–1143
183. Van Erp M, Rizzo G, Troncy R (2013) Learning with the web: Spotting named entities on the intersection of nerd and machine learning. In: # MSM, pp 27–30
184. Waitelonis J, Jürges H, Sack H (2016) Don't compare apples to oranges: extending gerbil for a fine grained nel evaluation. In: Proceedings of the 12th international conference on semantic systems, pp 65–72. ACM
185. Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. J. Big data 3(1):1–40
186. Xu B, Zhang Y, Liang J, Xiao Y, Hwang Sw, Wang W (2016) Cross-lingual type inference. In: International conference on database systems for advanced applications. Springer, pp 447–462
187. Xu M, Jiang H, Watcharawittayakul S (2017) A local detection approach for named entity recognition and mention detection. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol 1, pp 1237–1247
188. Yadav V, Bethard S (2019) A survey on recent advances in named entity recognition from deep learning models. arXiv:1910.11470
189. Yaghoobzadeh Y, Adel H, Schuetze H (2018) Corpus-level fine-grained entity typing. J Artif Intell Res 61:835–862
190. Yaghoobzadeh Y, Schütze H (2016) Corpus-level fine-grained entity typing using contextual information. arXiv:1606.07901
191. Yaghoobzadeh Y, Schütze H (2017) Multi-level representations for fine-grained typing of knowledge base entities. arXiv:1701.02025
192. Yaghoobzadeh Y, Schütze H (2018) Multi-multi-view learning: multilingual and multi-representation entity typing. arXiv:1810.10499
193. Yamada I, Shindo H, Takeda H, Takefuji Y (2017) Learning distributed representations of texts and entities from knowledge base. Trans Assoc Comput Linguist 5:397–411
194. Yang Y, Chang MW (2016) S-mart: Novel tree-based structured learning algorithms applied to tweet entity linking. arXiv:1609.08075
195. Yang Y, Irsoy O, Rahman KS (2018) Collective entity disambiguation with structured gradient tree boosting. arXiv:1802.10229
196. Yang Z, Cohen WW, Salakhutdinov R (2016) Revisiting semi-supervised learning with graph embeddings. arXiv:1603.08861
197. Yu X, Mayhew S, Sammons M, Roth D (2018) On the strength of character language models for multilingual named entity recognition. arXiv:1809.05157
198. Zhang S, Jiang H, Xu M, Hou J, Dai L (2015) The fixed-size ordinally-forgetting encoding method for neural network language models. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pp 495–500
199. Zhang Y, He S, Liu K, Zhao J (2016) A joint model for question answering over multiple knowledge bases. In: Thirtieth AAAI conference on artificial intelligence
200. Zwicklbauer S, Seifert C, Granitzer M (2016) Robust and collective entity disambiguation through semantic embeddings. In: Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval, pp 425–434

Imane Guellil is a research fellow interested in Natural Language Processing (NLP), including sentiment analysis, automatic translation, named entity recognition, disambiguation and linking. She works on different kinds of data, including news, social media and medical data. She is a pioneer working on the Algerian dialect and sentiment analysis of Arabizi and Arabic dialects. Her current position at the University of Edinburgh as a research fellow in Clinical NLP allows her to apply her NLP expertise to automatically detect adverse events from free text. Her previous experience at Aston University as a Data Science Researcher (Knowledge Transfer Partnership associate) has enabled her to embed her academic NLP/data science expertise into the industrial products of a UK company managing large document collections. Before working for Aston University, Dr. Guellil occupied an assistant professor position for four years at the Higher School of Applied Science in Algeria, where she supervised different groups of students on different NLP projects. She had the opportunity to collaborate with researchers from different universities worldwide. During these collaborations, Dr. Guellil produced around 25 papers (where 8 were published in Journals, 1 as a book chapter, 1 representing her thesis and the rest were published in conferences). Dr. Guellil has also acted as a reviewer for different journals, including TALLIP (ACM) and SNAM (Springer).

Antonio Garcia-Dominguez is a Lecturer in Software Engineering at the University of York. Before this, he was a Lecturer in Computer Science at Aston University (United Kingdom). His main research interests are software testing and model-driven engineering: in both of these fields, the increase in system sizes has required the adoption of AI-based approaches and non-relational database technologies to scale up. This combination motivated a partnership with Folding Space on the use of these approaches to extract value from large unstructured repositories. In addition to over 10 papers in peer-reviewed journals and over 40 papers in conferences and workshops, Antonio is a core contributor to several open-source projects. Some of these projects include the Eclipse Epsilon model management languages and tools, the GAmera mutation analysis tool, or the Eclipse Hawk model indexing framework.

Peter R. Lewis holds a Canada Research Chair in Trustworthy Artificial Intelligence, at Ontario Tech University. He is an associate professor in the Faculty of Business and Information Technology, where he leads the Trustworthy AI Lab. Peter's research advances both foundational and applied aspects of trustworthy, reflective, and socially intelligent systems. He is interested in where AI meets society, and how to help that relationship work well. His research is concerned with how to conceive and build AI systems that meet this challenge. His work draws on extensive experience applying AI commercially, as well as an academic background in nature-inspired, socially-sensitive, and self-aware intelligent systems. He co-edited the foundational book, Self-Aware Computing Systems, published by Springer, and is an Associate Editor of IEEE Technology & Society Magazine. He has published over 75 papers in academic journals and conference proceedings and led teams that have worked with dozens of companies in the areas of artificial intelligence, data science, and software development. Formerly, he was Director of Think Beyond Data, and co-founder of Beautiful Canoe. He was a Lecturer then Senior Lecturer at Aston University, UK, and a Postdoctoral Research Fellow at the University of Birmingham, UK. He obtained his PhD at the University of Birmingham, and is a member of the Aston Lab for Intelligent Collectives Engineering (ALICE).

Shakeel Hussain is a Technical Manager at Folding Space. Shakeel manages a team of software developers and helps the Folding Space develop innovative products and solutions to a wide range of problems experienced by both the private and public sectors. Shakeel originally studied Physics at Imperial College, London before developing an interest in computer science.

Geoffrey Smith As a founder of Folding Space in 1999, Geoffrey Smith leads the management team and is responsible for the business strategy and technology vision of the company. He has some 25+ years IT industry experience, being involved in technical, marketing, sales and product management at every level, including many board appointments. His particular speciality is new technology R&D and product development. Geoffrey originally trained as a Clinical Psychologist but left medicine for IT. He has undertaken post-graduate and doctoral research at various Universities including Cambridge, London, Harvard and MIT and holds four separate post-graduate degrees.