Named Entity Linking on Handwritten Document Images

Oliver Tüselmann[0000−0002−8892−3306] and Gernot A. Fink[0000−0002−7446−7813]

Department of Computer Science, TU Dortmund University,
44227 Dortmund, Germany
{oliver.tueselmann, gernot.fink}@cs.tu-dortmund.de

Abstract. Named Entity Linking (NEL) is an information extraction task that semantically enriches documents by recognizing mentions of entities in a text and matching them against an entry in a Knowledge Base (KB). This semantic information is fundamentally important for realizing a semantic search. Furthermore, it serves as a feature for subsequent tasks (i.e. Question Answering) as well as for improving the user experience. Current NEL approaches and datasets from the Document Image Analysis community are mainly focusing on machine-printed documents and do not consider handwriting. This is mainly due to the lack of annotated NEL handwriting datasets. To fill this gap, we manually annotated the well known IAM and George Washington datasets with NEL labels and created a synthetic handwritten version of the AIDA-CoNLL dataset. Furthermore, we present an evaluation protocol as well as a baseline approach.

Keywords: Named entity linking · Named entity disambiguation · Document image analysis · Information retrieval · Handwritten documents

1 Introduction

Over the last few years, we observe an increasing interest of the Document Image Analysis (DA) community towards the semantic analysis of documents. Thereby, the understanding of a document is becoming increasingly important. Several tasks from the field of Natural Language Processing (NLP), such as Named Entity Recognition (NER) [3,5,7,39] and Question Answering (QA) [22,23], have already been investigated in the context of document images. Beside these two, Named Entity Linking (NEL) constitutes another task of high interest in the DA community. NEL recognizes entities in unstructured texts and links them to their corresponding entries in a Knowledge Base (KB) (see Fig. 1). KBs (e.g. Wikidata [40], DBpedia[20]) contain information about entities and usually their relations. Entries in a KB are often so-called Named Entities (NEs). These are objects from the real world, such as persons, places, organizations or events. NEL is particularly useful for indexing document images and thus realizing a semantic search. Moreover, the extracted information can contribute to subsequent tasks and can improve the user experience [35].

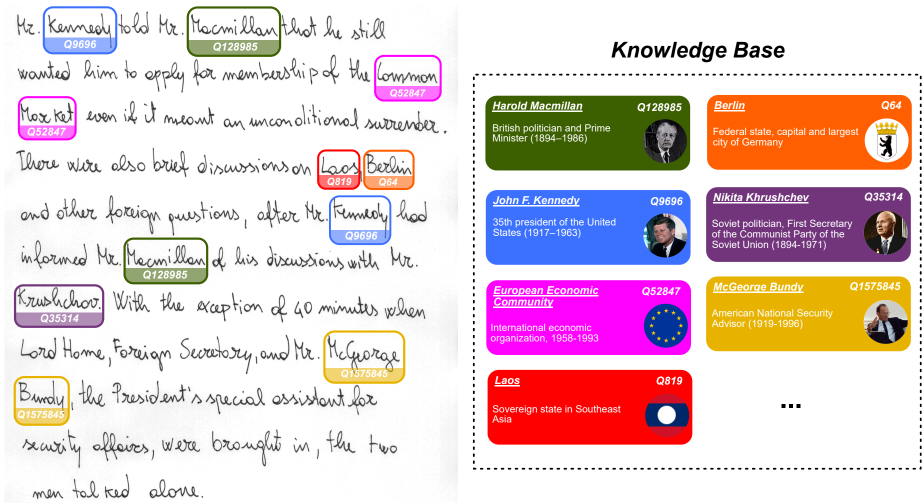


Fig. 1. An example for linking Named Entities against Wikidata on a handwritten document image from the IAM database.

The information extraction task of NEL is an ongoing research topic in the NLP community. Multiple approaches and datasets already exist for this task [35]. In DA, there is also an active community working on the recognition and mapping of entities in historical newspapers [7]. Approaches and datasets mainly focus on machine-printed documents and do not consider handwriting separately. In [22], Mathew et al. recently showed that Question Answering approaches trained on machine-printed data perform rather limited on handwritten document images and they suggested that specialized models are needed. This is mainly due to the fact that handwriting generally has more variability compared to machine-printed text, leading to more errors during recognition. Furthermore, there is a difference between machine and handwritten recognition errors, which would make a simple adaptation less than optimal [22].

To the best of our knowledge, there is only one work dealing with the task of NEL on handwritten document images [13]. This work focuses on a historical, non-public dataset. The lack of publicly available datasets hinders the investigation, development and comparison of NEL approaches on handwritten documents. To tackle this problem, we present three new datasets for this task. The first dataset is synthetically generated based on AIDA-CoNLL [14], the standard benchmark in the textual domain. In contrast, the other two datasets feature real handwriting, that we manually labeled with NEL annotations. We make the datasets publicly available for the community and present an evaluation protocol as well as an initial baseline approach [1].

[1] https://patrec.cs.tu-dortmund.de/cms/en/home/Resources/index.html

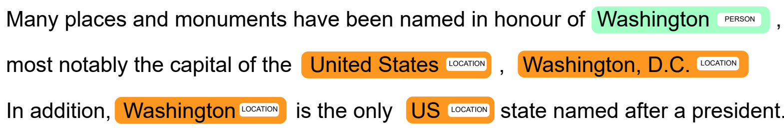


Fig. 2. A visualization of the mention detection step. The task does not include the prediction of Named Entity tags.

2 Named Entity Linking

The general goal of NEL is to recognize entities in an unstructured text and assign them to a unique identity from a given KB. To achieve a mapping between mentions in a text and entities in a KB, state-of-the-art approaches usually divide the task into a mention detection and an entity disambiguation phase [15]. In the following, we describe the two main steps of NEL approaches in more detail.

In the literature the term Named Entity Linking is often used as a synonym for Named Entity Disambiguation (NED). It is important to distinguish both tasks, as NED skips the mention detection step. Therefore, only gold standard mentions are considered and have to be disambiguated.

Mention Detection The first stage of a NEL system detects mentions of entities from a given KB in an unstructured text (see Fig. 2). For this task Named Entity Recognition (NER) models are often used [35]. These models do not only recognize NEs but also predict if an identified NE is, e.g., a person, an organization or an event. Even if some NEL systems use these class predictions for their further steps, only the identification of mentions is important in this phase. Traditional NER methods are mainly implemented based on handcrafted rules, orthographic features or dictionaries [42]. Statistical-based methods such as Hidden Markov Models and Conditional Random Fields (CRFs) achieved further progress in this field [41]. In recent years, deep neural network approaches greatly improved the recognition accuracies. Especially combinations of recurrent neural networks, CRFs and pre-trained word embeddings have been successful [18]. For a detailed overview of NER, see [42].

Entity Disambiguation In most approaches, the disambiguation process is further separated into the candidate generation and entity ranking stages [10,15,19]. The candidate generation step creates a list of possible entities that are associated to an identified mention from the previous step (see Fig. 3). An intuitive realization is to find entities in the KB which textually match the mention [2,35,36]. To counteract the variability of mentions, heuristics like the Levenshtein distance or normalization are used [35]. Since entities are often associated with different names, state-of-the-art methods usually rely on dictionaries with additional aliases for entities [35]. Such dictionaries are often generated from the disambiguation and redirect pages of Wikipedia. There are several specialized resources available online containing aliases and synonyms of entities [14]. Most

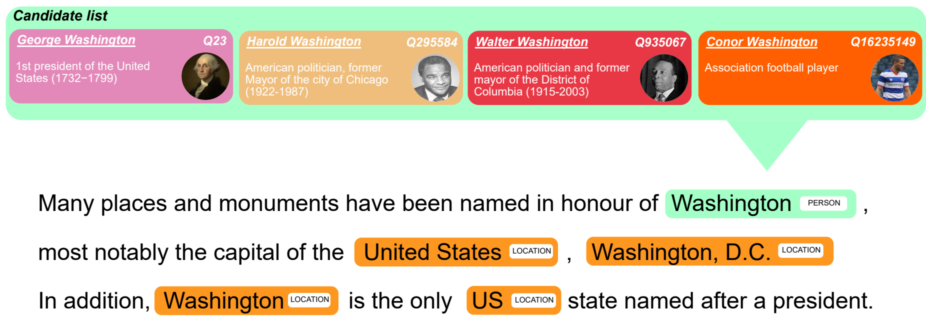


Fig. 3. An example showing the generation of candidates for the first mention of *Washington* in the sample.

systems rely on precalculated prior probabilities of correspondence between mention and entities $p(e|m)$ [10,15,17,29,44]. This probability is commonly computed based on Wikipedia hyperlinks, where the URL provides the entity e and the corresponding text of the link provides the mention m [38].

Entity Ranking is the final step of the traditional NEL pipeline and is usually interpreted as a retrieval problem. The system assigns a score to each entity from the candidate generation step, indicating how well the entity matches the given mention (see Fig. 4). For the disambiguation process it is crucial to capture the semantic information from the context in which the mention appears [35]. Therefore, state-of-the-art approaches usually produce a vector representation for the given mention in its context and also for each candidate entity [2,35]. Finally, the similarity between the mention and entity representation is computed. Traditional NEL approaches typically use handcrafted features to calculate similarities between a mention and its candidate entities [14,31,36]. In recent years, approaches based on neural networks outperformed traditional ones [35]. The general idea is to represent the mention, context and entities as vector representations and to use a neural architecture to compute similarity scores between the given entities and the mention. There are different strategies in the literature to encode entities. A widely used strategy is to map entities into a continuous vector space, such that entity representations are embedded into the same semantic space as words [19,44]. Another strategy uses relations between entities in a KB and graph embedding methods [34]. Lately, neural encoders are used to convert textual descriptions of entities into embeddings and tackle the task with a self-attention model based on a pre-trained BERT model [43].

While it is common to separate the entity recognition and disambiguation step, a few systems provide a joint model [4,17]. It is possible that some mentions do not have their corresponding entity in the given KB. Therefore, NEL approaches should also be able to recognize if a mention does not exist in the KB. This prediction could be interpreted and realized as a classification problem with rejection [35]. For a more detailed overview of textual NEL, see [2,35,36].

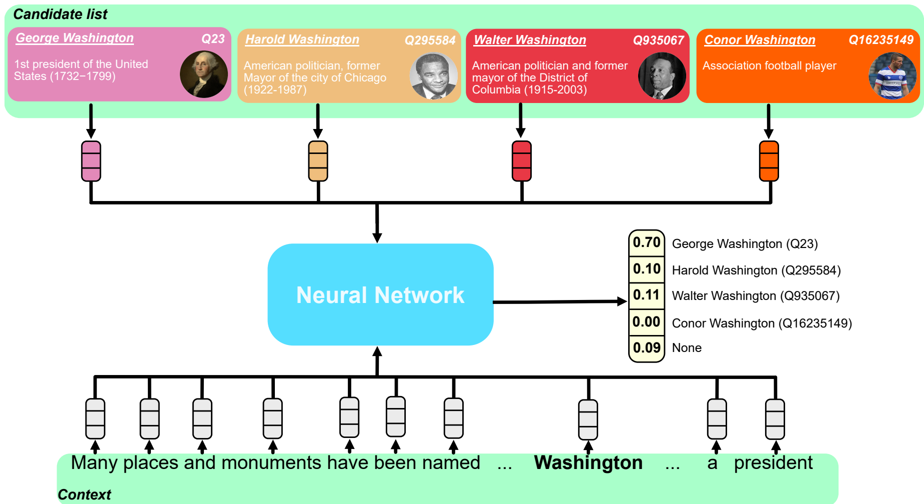


Fig. 4. An overview of the entity ranking step for the first mention of *Washington* given the candidate list generated in figure 3.

Document Image Domain In recent years, several works on the recognition and mapping of entities in historical document images were proposed [8,27]. Traditional approaches use rule based systems [12] or just concentrate on specific types of entities [37]. There are also multiple works that use available NEL approaches from the literature to analyze historical data. For example, Ruiz et al. [32] apply NEL as a sub task for generating an overview of the manually transcribed Bentham dataset. Hereby, approaches working on historic data often have to deal with the problem that most publicly available KBs do not contain most of the entities mentioned in the historic texts [25].

A fundamental work of Pontes et al. [28] evaluates the effects of synthetically generated Optical Character Recognition (OCR) errors in machine-printed documents on the NED process. They show that the performance of NED systems can decrease by 20% when about 15% of the words are not transcribed properly. Lately, Ehrmann et al. provided a large annotated NER and Linking dataset on historical newspapers [7] and organized the CLEF 2020 HIPE Shared Task [8]. Furthermore, Pontes et al. [27] proposed recently a multilingual end-to-end NEL approach based on the model of Kolitsas et al [17].

In the handwriting domain, Hendriks et al. do a case study in [13] on a handwritten, historic, dutch document corpus. They use a fuzzy matching strategy to detect mentions. Afterwards, they select candidates by using fuzzy string matching and additional context information, e.g., dates. Finally, an active learning approach is used for linking entities to entries in their KB. It is important to note that this work just concentrates on person names and uses a small KB.

3 Dataset

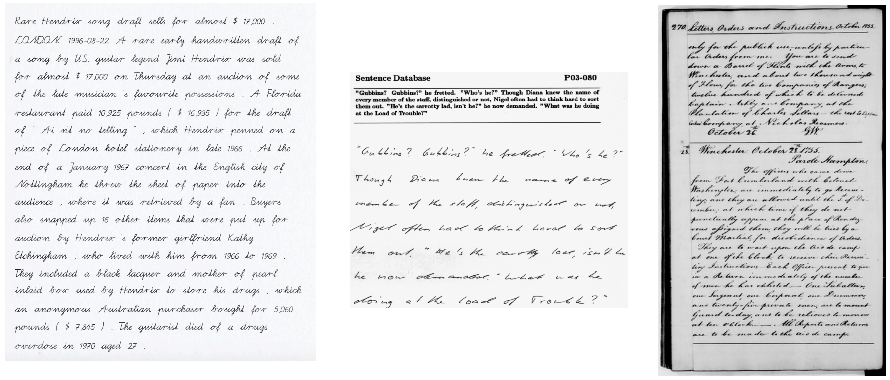
There are currently no publicly available datasets dealing with NEL on handwritten document images. There is just the historical dutch dataset proposed by [13], which is currently not publicly available and is highly specialized for the specific use-case. To evaluate and develop NEL approaches on handwritten data, we provide three new annotated datasets for this task. In the following, we describe each of the datasets in more detail.

3.1 Synthetic HW-AIDA-CoNLL

To train and evaluate NEL models on handwritten document images, we created a synthetic dataset called HW-AIDA-CoNLL. It consists of synthetically generated handwritten pages using the text of AIDA-CoNLL, the standard benchmark for textual NEL. The benchmark provides a training (AIDA-train [14]), validation (AIDA-A [14]) and six test datasets. The special aspect of this benchmark is that there is not only a so-called in-domain test dataset (AIDA-B [14]), which is evaluated on the same type of data as used during training and validation. But there are five out-of-domain datasets (MSNBC [6], CWEB [9], ACE2004 [31], AQUAINT [24] and WIKI [31]), on which the system is not fine-tuned and thus provides an indication of its adaptivity on other domains. Even though the five out-of-domain datasets are very useful for evaluating NED approaches, several NEL publications consider only a subset of them [4,15,17]. We follow this trend and restrict our synthetically generated out-of-domain datasets to MSNBC, CWEB and WIKI.

The AIDA-CoNLL dataset [14] was manually annotated by Hoffart et al. and is based on the CoNLL 2003 dataset [33]. The data is divided into AIDA-train for training, AIDA-A for validation, and AIDA-B for testing. It is one of the biggest manually annotated NEL datasets available, containing 1393 news articles and 27817 linkable mentions. The MSNBC dataset [6] contains 20 news articles from 10 different topics and provides 656 linkable mentions. The datasets was cleaned and updated by Guo et al. [11]. CWEB [9] contains 320 documents and 11154 linkable mentions. Therefore, web pages from the ClueWeb Corpora in English language were annotated. Finally, WIKI [31] is composed of 320 documents and 6821 linkable mentions that correspond to existing hyperlinks in Wikipedia articles. The NEL annotations of CWEB and WIKI are generated automatically, while the others were checked or created manually and are therefore more reliable.

For generating the synthetic handwritten document images of these datasets, we used nearly the same approach as proposed in [22]. We synthetically generated each text document as an image using a handwritten font. The font is randomly sampled from over 300 publicly available True Type fonts that resemble handwriting. Each word of the document is rendered onto a transparent background. The font size is set randomly from a range of 28–52 pts and the intensity of the text stroke is varied between 0 and 50. In the next step, all word images are pasted onto a background image in the same order as in the original passage



(a) HW-AIDA-CoNLL (b) IAM-DB (c) George Washington

Fig. 5. Examples of document images for the synthetically generated HW-AIDA-CoNLL as well as the IAM-DB and George Washington datasets.

using alpha blending. We randomly sample the background image from a small set of manuscripts like textures. While pasting the words onto the background image, we break the lines whenever a word does not fit on the actual line. To provide variability, we randomly set the page width to a minimum of 800 and a maximum of 1600 pixel. Furthermore, we randomly padded the borders on all four sides of the document. To provide usability for approaches that work on full pages, we additionally divided long documents into multiple pages. Sentence segmentation is given for the documents and there is no overlap between the fonts in the training and test set.

3.2 IAM-DB

The IAM Database [21] is a major benchmark for several DA tasks. The documents contain modern English sentences and were written by a total of 657 different people. The pages contain text from the genres listed in Fig. 6. The wide range of genres makes this an excellent choice of dataset for NEL. The database consists of 1539 scanned text pages containing a total of 13353 text lines and 115320 words.

Since NEL is a semantic task, we use an optimized partitioning of the IAM-DB data into training, validation and testing as proposed by Tüselmann et al [39]. For the manual NEL annotations of the IAM-DB, we use Wikidata [40] as our KB. It is one of the largest public databases with over 95 million entries and it is widely used in the literature. The advantage of Wikidata is that its IDs do not change over time as Wikipedia page numbers or titles sometimes do. Given a Wikidata ID, it is possible to extract the page number or title of their corresponding Wikipedia page if needed for an approach.

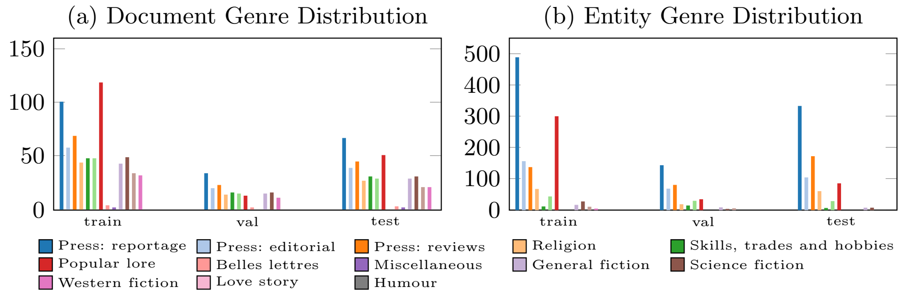


Fig. 6. The amount of document pages (a) and entities (b) per genre in training, validation and test set for the IAM database [39].

Our NEL dataset contains 3650 links between word images and Wikidata IDs. There are a total of 2405 entity-mention pairs in the dataset. These are divided into 1238 pairs in the training set, 785 pairs in the test set and 382 pairs in the validation set. Fig. 6 shows that most linkable entities are located in the news articles and only a few in books and novels. This poses a major challenge to the NEL model, as fictitious entities should not be mapped to mismatched entities from the KB.

3.3 George Washington

The George Washington (GW) dataset [30] is a widely used benchmark in DA. It includes 20 pages of correspondences between George Washington and his associates produced in 1755. The documents were written by a single person in historical English. The manual NEL annotations were created in the same way as for the IAM database. The NEL dataset contains 218 links between word images and Wikidata IDs. There are a total of 145 entity-mention pairs in the dataset. These are divided into 105 pairs in the training set, 31 pairs in the test set and 9 pairs in the validation set.

4 Baseline Approach

In this section, we present our baseline approach for NED and Linking. We propose a two-stage system that works on segmented word images. The model transforms a document image into a textual representation and determines the linking of NEs using a textual NEL model (see Fig. 7). The approach is based on a state-of-the-art Handwriting Text Recognizer (HTR) [16] and NEL model [15]. The first step of our approach is to feed all word images into the recognizer in their order of occurrence on the document pages. The recognized text is then processed sequentially by the NEL model. Thereby, it extracts relevant mentions from the text by using a state-of-the-art NER model [1]. Given the relevant mentions, candidates are generated and finally disambiguated. In the following, we will provide further information about each component of this approach.

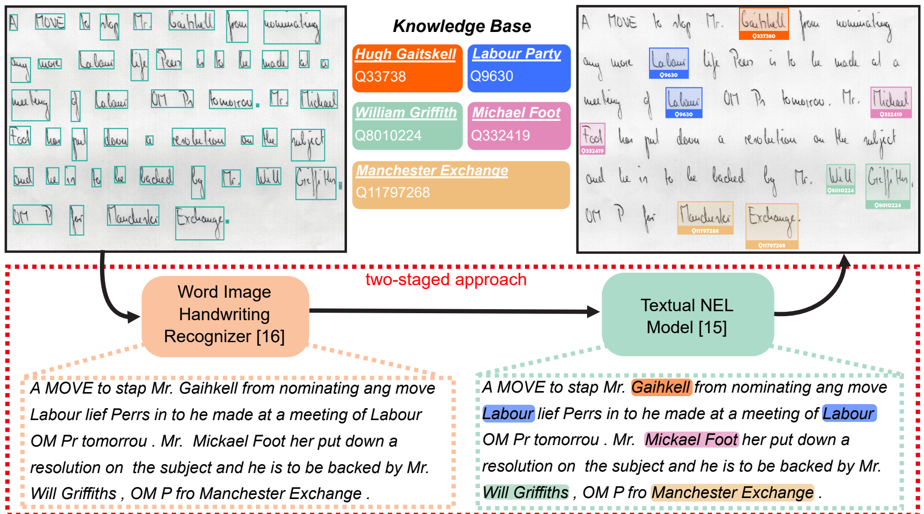


Fig. 7. An overview of our two-staged baseline approach for linking entities on handwritten document images against Wikidata.

4.1 Handwriting Text Recognizer

We use the attention-based sequence-to-sequence HTR model proposed by Kang et al. [16] for transcribing word images. It works on character-level and does not require any information about the language, except for an alphabet. The approach has the advantage that it does not require any dataset-specific preprocessing steps and provides satisfying results for the majority of datasets. Apart from adjusting the size of the input images, the maximum word length and the alphabet, we use the parameters proposed in [16]. As the approach does not use any linguistic resources during recognition, it has the advantage of not penalizing out-of-vocabulary words. However, the use of a lexicon often helps with minor recognition errors. Therefore, we additionally apply a fixed vocabulary consisting of its training, validation and test words on the outputs of the HTR model. In the following, *HTR+D* denotes a model using a dictionary. Table 1 shows the Character Error Rate (CER) and Word Error Rate (WER) for the recognition models on the datasets introduced in Sec. 3. We report similar error rates for the IAM and GW datasets as published in the literature [16]. Improvements are possible with further optimizations and dataset specific adaptations.

4.2 Named Entity Linking

For NEL, we use the Radboud Entity Linker (REL), a state-of-the-art NEL approach proposed by van Hulst et al. [15]. Their model follows the standard entity linking architecture, consisting of mention detection, candidate selection and entity disambiguation.

Table 1. Handwriting recognition (HTR) rates measured in Character Error Rate (CER) and Word Error Rate (WER) for the used datasets.

		AIDA-B	MSNBC	WIKI	CWEB	IAM	GW
HTR	*CER*	6.9	5.7	7.4	8.7	7.1	5.2
	WER	18.3	16.9	18.2	18.7	20.4	14.5
HTR+D	*CER*	5.7	4.4	6.3	8.0	6.4	4.1
	WER	9.2	7.6	9.8	12.7	10.8	6.1

Mention Detection For mention detection, the FLAIR framework [1] is used. It is a NER model that roughly follows the architecture proposed by Lample et al. [18]. The approach transforms the input words into a vector representation using a pre-trained word embedding model (e.g. BERT [26]). These representations are encoded using a Bidirectional Long Short-Term Memory (BLSTM). Finally, a CRF uses the encoding from the BLSTM and predicts NE tags for each input word.

Candidate Selection The selection of candidates follows the idea proposed by Le and Titov [19], where for each mention up to seven candidate entities are selected. Four candidates are selected based on mention-entity prior statistics, collected from Wikipedia, CrossWikis [38] and the YAGO dictionary [14]. Furthermore, three candidates are chosen based on their similarity to the 50 word context (c) surrounding the mention. REL uses word and entity embeddings provided by Yamada et al. [44] to estimate the similarity between an entity and the context of a mention. The similarity is obtained for 30 entities with the highest mention-entity prior and is calculated for each of these entities e by $\mathbf{e}^T \sum_{w \in c} \mathbf{w}$, whereby $\mathbf{w}$ and $\mathbf{e}$ are word and entity embeddings.

Entity Disambiguation REL follows the Ment-norm approach proposed by Le and Titov [19] for mapping mentions to their entries in a given KB. The linking is done by a local score, composed of mention-entity prior and context similarity, as well as the coherence with other linking decisions in the document. The coherence between all entity decisions is done by assuming K latent relations between mentions. To optimize the local and global conditions, max-product loopy belief propagation is used [10]. The final score for an entity of a mention is obtained by a two-layered neural network that combines mention-entity prior information and the max-marginal probability for the entity and its document. Finally, the posterior probabilities of the linked entities are computed by applying a logistic function over the final scores of the neural network.

5 Experiments

We evaluate NED and NEL on the handwritten datasets introduced in Sec. 3 using our two-stage baseline approach. Before we present and discuss the evaluation results, we describe the evaluation protocol in Sec. 5.1.

Table 2. Named Entity Disambiguation (NED) and Linking (NEL) performances measured in InKB micro F1 score based on strong annotation matching. Training was done on AIDA-train.

		AIDA-B	MSNBC	WIKI	CWEB	IAM	GW
NED	*Annotation*	87.9	91.9	75.7	76.7	69.4	22.6
	HTR+D	74.8	80.1	63.1	60.1	54.9	19.4
	HTR	52.0	59.0	47.9	45.8	31.4	12.9
NEL	*Annotation*	79.1	75.4	40.1	47.8	49.8	15.2
	HTR+D	68.0	64.2	33.2	34.9	38.1	12.5
	HTR	56.7	55.7	30.6	33.8	27.5	7.1

5.1 Evaluation Protocol

For NEL, the InKB micro F1 score based on strong annotation matching (see equation 3) is often used in the literature to compare approaches. InKB only focuses on those cases in which either the mention has a valid entity in the KB or the NEL approach predicts an entity of the KB. Strong annotation matching implies that a prediction is only correct if it recognizes the mention and links it to the proper entity. Therefore, it is not enough to recognize a part of the mention and match it correctly. For computing the F1 score, it is important to distinguish between micro and macro averaging. The macro F1 computes the F1 score independently for each class and then takes the average, whereas the micro F1 aggregates the contributions of all classes. The evaluation of NED provides a special case as the mentions are provided as input and do not have to be detected. Therefore, the number of predicted mentions is equal to the number of mentions in gold standard, which leads to F1 being equal to precision, recall and accuracy [36].

$$Precision = \frac{\text{\# of correctly detected and disambiguated mentions}}{\text{\# of predicted mentions by model}} \tag{1}$$

$$Recall = \frac{\text{\# of correctly detected and disambiguated mentions}}{\text{\# of mentions in gold standard}} \tag{2}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3}$$

5.2 Results

In this section, we present and discuss the NED and NEL performances on our created datasets using our baseline approach. To evaluate the impact of the HTR errors, we need the scores of the NED and NEL models on perfect recognition results. In the following, we denote approaches that work on the textual annotations of the word images instead of the HTR results as *Annotation*.

The results of our baseline approach show that NEL is a considerably harder task compared to NED. This is shown in table 2, where the F1 scores for NEL are significantly lower than for NED on all datasets, even without using handwriting recognition. The NED and NEL models perform well on all textual datasets and the scores are similar to the published results obtained in [15]. The scores show that the IAM dataset is comparatively challenging for both tasks. This is probably due to the small context and the large variety of topics [35].

The recognition errors have a strong impact on the performance of the models. In comparison to handwriting recognition only, the scores improved significantly by adding a dictionary. This is mainly due to the lower WER on the datasets. However, even with small recognition error rates as given for the MSNBC dataset, a strong decrease of the F1 score occurs. Interestingly, REL uses the same model and approach for NED as described in [28]. However, we cannot obtain similar results as observed in their OCR evaluation, even though similar error rates were considered there.

For several reasons, NED and NEL perform particularly poorly on the GW dataset. It is historical data in which only 38% of people and place mentions are indexed in Wikidata. Furthermore, we work on textual annotations that are all in lowercase characters, which makes mention detection and linking considerably harder. For the detection we already used a NER model from FLAIR, which was trained on lowercase letters. The dataset also contains difficult conditions for linking, since there are several abbreviations like *gw* for *George Washington* and often only the last name and military rank of the person is given. In addition, a lot of contextual knowledge is needed as the data commonly involves people and places in the area of Virginia between 1700 and 1800. Finally, the training dataset is very small, which makes fine-tuning difficult.

6 Conclusion

In this work, we introduce the task of Named Entity Linking on handwritten document images. We also propose and publish the first Entity Linking datasets for this task (HW-AIDA-CoNLL, IAM and George Washington) along with a suitable evaluation protocol. The datasets offer two tasks, Named Entity Disambiguation and Named Entity Linking. We investigate a two-stage approach for both tasks on word-segmented handwritten document images in this work. Even though our experiments show that the two tasks are particularly difficult for handwritten document images, we are already achieving promising results with our baseline approach.

References

1. Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., Vollgraf, R.: FLAIR: An easy-to-use framework for state-of-the-art NLP. In: Annual Conf. of the North American Chapter of the Association for Computational Linguistics. pp. 54–59. Minneapolis, MN, USA (2019)
2. Al-Moslmi, T., Ocaña, M.G., Opdahl, A.L., Veres, C.: Named entity extraction for knowledge graphs: A literature overview. IEEE Access **8**, 32862–32881 (2020)
3. Boroş, E., Pontes, E.L., Cabrera-Diego, L.A., Hamdi, A., Moreno, J.G., Sidère, N., Doucet, A.: Robust named entity recognition and linking on historical multilingual documents. In: Working Notes of Conf. and Labs of the Evaluation Forum. Thessaloniki, Greece (2020)
4. Cao, N.D., Izacard, G., Riedel, S., Petroni, F.: Autoregressive entity retrieval. In: Int. Conf. on Learning Representations. Vienna, Austria (2021)
5. Carbonell, M., Fornés, A., Villegas, M., Lladós, J.: A neural model for text localization, transcription and named entity recognition in full pages. Pattern Recognition, Letters **136**, 219–227 (2020)
6. Cucerzan, S.: Large-scale named entity disambiguation based on Wikipedia data. In: Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. pp. 708–716. Prague, Czech Republic (2007)
7. Ehrmann, M., Romanello, M., Bircher, S., Clematide, S.: Introducing the CLEF 2020 HIPE shared task: Named entity recognition and linking on historical newspapers. In: European Conf. on Information Retrieval. pp. 524–532. Lisbon, Portugal (2020)
8. Ehrmann, M., Romanello, M., Flückiger, A., Clematide, S.: Overview of CLEF HIPE 2020: Named entity recognition and linking on historical newspapers. In: Int. Conf. of the CLEF Association. pp. 288–310. Thessaloniki, Greece (2020)
9. Gabrilovich, E., Ringgaard, M., Subramanya, A.: FACC1: Freebase annotation of ClueWeb corpora, Version 1 (Release date 2013-06-26, Format version 1, Correction level 0) (2013)
10. Ganea, O., Hofmann, T.: Deep joint entity disambiguation with local neural attention. In: Proc. Conf. on Empirical Methods in Natural Language Processing. pp. 2619–2629. Copenhagen, Denmark (2017)
11. Guo, Z., Barbosa, D.: Robust entity linking via random walks. In: Proc. ACM Int. Conf. on Information and Knowledge Management. pp. 499–508. Shanghai, China (2014)
12. Heino, E., Tamper, M., Mäkelä, E., Leskinen, P., Ikkala, E., Tuominen, J., Koho, M., Hyvönen, E.: Named entity linking in a complex domain: Case second world war history. In: Proc. Int. Conf. on Language, Data, and Knowledge. pp. 120–133. Galway, Ireland (2017)
13. Hendriks, B., Groth, P., van Erp, M.: Recognizing and linking entities in old dutch text: A case study on VOC notary records. In: Proc. Int. Conf. Collect and Connect: Archives and Collections in a Digital Age. pp. 25–36. Leiden, Netherlands (2020)
14. Hoffart, J., Yosef, M.A., Bordino, I., Fürstenau, H., Pinkal, M., Spaniol, M., Taneva, B., Thater, S., Weikum, G.: Robust disambiguation of named entities in text. In: Proc. Conf. on Empirical Methods in Natural Language Processing. pp. 782–792. Edinburgh, UK (2011)
15. van Hulst, J.M., Hasibi, F., Dercksen, K., Balog, K., de Vries, A.P.: REL: An entity linker standing on the shoulders of giants. In: Proc. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval. pp. 2197–2200. Xi'an, China (2020)

16. Kang, L., Toledo, J.I., Riba, P., Villegas, M., Fornés, A., Rusiñol, M.: Convolve, attend and spell: An attention-based sequence-to-sequence model for handwritten word recognition. In: Proc. German Conf. on Pattern Recognition. pp. 459–472. Stuttgart, Germany (2018)
17. Kolitsas, N., Ganea, O., Hofmann, T.: End-to-end neural entity linking. In: Proc. Conf. on Computational Natural Language Learning. pp. 519–529. Brussels, Belgium (2018)
18. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. In: Annual Conf. of the North American Chapter of the Association for Computational Linguistics. pp. 260–270. San Diego, CA, USA (2016)
19. Le, P., Titov, I.: Improving entity linking by modeling latent relations between mentions. In: Annual Meeting of the Association for Computational Linguistics. pp. 1595–1604. Melbourne, Australia (2018)
20. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. Semantic Web **6**(2), 167–195 (2015)
21. Marti, U., Bunke, H.: The IAM-database: An English sentence database for offline handwriting recognition. Int. Journal on Document Analysis and Recognition **5**(1), 39–46 (2002)
22. Mathew, M., Gómez, L., Karatzas, D., Jawahar, C.V.: Asking questions on handwritten document collections. Int. Journal on Document Analysis and Recognition **24**, 235–249 (2021)
23. Mathew, M., Karatzas, D., Jawahar, C.V.: DocVQA: A dataset for VQA on document images. In: IEEE Workshop on Applications of Computer Vision. pp. 2199–2208. Waikoloa, HI, USA (2021)
24. Milne, D., Witten, I.H.: Learning to link with Wikipedia. In: Proc. ACM Int. Conf. on Information and Knowledge Management. pp. 509–518. Napa Valley, CA, USA (2008)
25. Munnelly, G., Pandit, H.J., Lawless, S.: Exploring linked data for the automatic enrichment of historical archives. In: Extended Semantic Web Conf. pp. 423–433. Crete, Greece (2018)
26. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. In: Annual Conf. of the North American Chapter of the Association for Computational Linguistics. pp. 2227–2237. New Orleans, LA, USA (2018)
27. Pontes, E.L., Cabrera-Diego, L.A., Moreno, J.G., Boroş, E., Hamdi, A., Sidère, N., Coustaty, M., Doucet, A.: Entity linking for historical documents: Challenges and solutions. In: Int. Conf. on Asia-Pacific Digital Libraries. pp. 215–231. Kyoto, Japan (2020)
28. Pontes, E.L., Hamdi, A., Sidere, N., Doucet, A.: Impact of OCR quality on named entity linking. In: Int. Conf. on Asia-Pacific Digital Libraries. pp. 102–115. Kuala Lumpur, Malaysia (2019)
29. Raiman, J., Raiman, O.: DeepType: Multilingual entity linking by neural type system evolution. In: Proc. AAAI Conf. on Artificial Intelligence. pp. 5406–5413. New Orleans, LA, USA (2018)
30. Rath, T.M., Manmatha, R.: Word spotting for historical documents. Int. Journal on Document Analysis and Recognition **9**(2-4), 139–152 (2007)

31. Ratinov, L., Roth, D., Downey, D., Anderson, M.: Local and global algorithms for disambiguation to Wikipedia. In: Annual Meeting of the Association for Computational Linguistics. pp. 1375–1384. Portland, Oregon (2011)
32. Ruiz, P., Poibeau, T.: Mapping the Bentham corpus: Concept-based navigation. Journal of Data Mining and Digital Humanities (2019)
33. Sang, E.F.T.K., Meulder, F.D.: Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In: Proc. Conf. on Computational Natural Language Learning. pp. 142–147. Edmonton, Canada (2003)
34. Sevgili, Ö., Panchenko, A., Biemann, C.: Improving neural entity disambiguation with graph embeddings. In: Annual Meeting of the Association for Computational Linguistics. pp. 315–322. Florence, Italy (2019)
35. Sevgili, Ö., Shelmanov, A., Arkhipov, M.Y., Panchenko, A., Biemann, C.: Neural entity linking: A survey of models based on deep learning. CoRR **abs/2006.00575** (2020)
36. Shen, W., Wang, J., Han, J.: Entity linking with a knowledge base: Issues, techniques, and solutions. IEEE Transactions on Knowledge and Data Engineering **27**(2), 443–460 (2015)
37. Smith, D.A., Crane, G.R.: Disambiguating geographic names in a historical digital library. In: European Conf. on Digital Libraries. pp. 127–136. Darmstadt, Germany (2001)
38. Spitkovsky, V.I., Chang, A.X.: A cross-lingual dictionary for English Wikipedia concepts. In: Proc. Int. Conf. on Language Resources and Evaluation. pp. 3168–3175. Istanbul, Turkey (2012)
39. Tüselmann, O., Wolf, F., Fink, G.A.: Are end-to-end systems really necessary for NER on handwritten document images? In: Proc. Int. Conf. on Document Analysis and Recognition. pp. 808–822. Lausanne, Switzerland (2021)
40. Vrandečić, D., Krötzsch, M.: Wikidata: A free collaborative knowledgebase. Communications of the ACM **57**(10), 78–85 (2014)
41. Wen, Y., Fan, C., Chen, G., Chen, X., Chen, M.: A survey on named entity recognition. In: Int. Conf. on Communications, Signal Processing, and Systems. pp. 1803–1810. Urumqi, China (2019)
42. Yadav, V., Bethard, S.: A survey on recent advances in named entity recognition from deep learning models. In: Proc. Int. Conf. on Computational Linguistics. pp. 2145–2158. Santa Fe, NM, USA (2018)
43. Yamada, I., Shindo, H.: Pre-training of deep contextualized embeddings of words and entities for named entity disambiguation. CoRR **abs/1909.00426** (2019)
44. Yamada, I., Shindo, H., Takeda, H., Takefuji, Y.: Joint learning of the embedding of words and entities for named entity disambiguation. In: Proc. Conf. on Computational Natural Language Learning. pp. 250–259. Berlin, Germany (2016)