Knowledge-based Contexts for Historical Named Entity Recognition & Linking

Emanuela Boros[1], Carlos-Emiliano González-Gallardo[1], Edward Giamphy[1,2], Ahmed Hamdi[1], José G. Moreno[1,3] and Antoine Doucet[1]

[1]University of La Rochelle, L3i, 17000 La Rochelle, France
[2]Preligens, 75009 Paris, France
[3]University of Toulouse, IRIT, 31000 Toulouse, France

Abstract

This paper summarizes the participation of the L3i laboratory of the University of La Rochelle in the *Identifying Historical People, Places, and other Entities* (HIPE) evaluation campaign of CLEF 2022 in both tasks: named entity recognition and classification (NERC), coarse- and fine-grained, and entity linking (EL) in historical newspapers and classical commentaries. For both tasks, we developed models based on our previous models, which ranked first at CLEF-`hipe-2020`. The NERC model is a Transformer-based architecture and the EL model is a BiLSTM-based architecture. For NERC, our main contribution is two-fold: (1) *data-wise* improvement – we propose a knowledge-based strategy to provide related context information to the NERC model; (2) *model-wise* improvement – we adapt the NERC model to the task of detecting coarse- and fine-grained entities in non-standard text via adapters and we include the knowledge-based contexts as context jokers. Our approaches ranked first on 84.6% of the leaderboards we participated in for NERC and 85.7% of them for EL.

Keywords

historical documents, fine-grained named entity recognition, named entity linking, knowledge bases, language models

1. Introduction

The identification of entities in historical documents, such as people and places, can be seen as a building block of historical knowledge that allows easier access and better information retrieval [1, 2, 3, 4]. Also, knowledge about historical events is gradually fading, especially among the younger generations. Thus, preserving the historical memory of the information that can be extracted from historical documents and bringing them to a larger audience, not limited to researchers and experts in the humanities [5, 6, 7], could lead to better and wider access to cultural heritage.

Although named entity recognition (NER) and linking (EL) systems have been developed to process modern data collections in general, NER and EL systems for processing historical

CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy
emanuela.boros@univ-lr.fr (E. Boros); carlos.gonzalez_gallardo@univ-lr.fr (C. González-Gallardo); edward.giamphy@preligens.com (E. Giamphy); ahmed.hamdi@univ-lr.fr (A. Hamdi); jose.moreno@irit.fr (J. G. Moreno); antoine.doucet@univ-lr.fr (A. Doucet)
0000-0001-6299-9452 (E. Boros); 0000-0002-0787-2990 (C. González-Gallardo); 0000-0002-2722-5168 (E. Giamphy); 0000-0002-8964-2135 (A. Hamdi); 0000-0002-8852-5797 (J. G. Moreno); 0000-0001-6160-3356 (A. Doucet)

documents are less common [8, 9]. Because these documents are not digitally born, they are scanned and processed by optical character recognition (OCR) tools to extract their textual content. However, the OCR process is not error free and misrecognizes some of the content. This can be due to the level of degradation of the document being scanned, the digitization process, and also the quality of the OCR tool. This causes digitization errors in the recognized text, such as misspelled locations or names.

In this context, the first CLEF-HIPE-2020 edition [10, 11, 12] proposed the tasks of named entity recognition and classification (NERC), both fine- and coarse-grained, and entity linking (EL) in historical newspapers written in English, French and German. The evaluation showed that neural-based systems with pre-trained language models or Transformer-based approaches [13, 14, 15] clearly prevailed in NERC [16], beating symbolic conditional random field (CRF) [17, 18], pattern-based approaches or BiLSTMs [19] by a large margin.

For its second edition, the HIPE evaluation campaign[1] took advantage of the availability of several NE annotated datasets produced by several European cultural heritage projects [20]. In this paper, we present our participation in the *Identifying Historical People, Places, and other Entities* (HIPE) evaluation campaign of CLEF 2022 in both tasks: NERC, fine-grained and coarse-grained, and EL in historical newspapers. For both tasks, we based our models on those that we proposed at CLEF-HIPE-2020 [13]. The NERC model was mainly based on the Transformer architecture [21] and the EL model was based on a BiLSTM architecture [22]. For NERC, our main contribution is two-fold: (1) we propose a knowledge-based system, where we build a multilingual knowledge base resting on Wikipedia and Wikidata to provide related context information to the NERC model (data-wise improvement); (2) we adapt the NERC model to the task of detecting coarse- and fine-grained entities in non-standard text by learning modular language- and task-specific representations via newly-proposed additional adapters [23, 24], small bottleneck layers inserted between the weights of two auxiliary Transformer layers (model-wise improvement) [25]. Furthermore, for taking advantage of the additional Wikipedia-based contexts, we include them in the model with mean-pooled representations that we refer to as *context jokers*. Official results of our participation show the effectiveness of our models over the CLEF-HIPE-2022 benchmark.

The paper is organized as follows: Section 2 introduces the task and the datasets. Section 3 presents our knowledge-retrieval modules. Sections 4 and 5 respectively present our NERC and EL systems and their corresponding performance. Conclusions are drawn in Section 6, where future work is also presented.

2. Datasets

The CLEF-HIPE-2022 competition proposed corpora composed of historical newspapers and classical commentaries covering circa 200 years. The historical newspaper data is composed of five datasets in English, Finnish, French, German and Swedish which originate from various projects and national libraries in Europe, from which, we experimented with the `hipe-2020` dataset. `hipe-2020` includes newspaper articles from Swiss, Luxembourgish and American

[1] https://hipe-eval.github.io/HIPE-2022/

Table 1
Overview of the `hipe-2020` and `ajmc` datasets. LOC = location, ORG = organization, PERS = person, PROD = product, TIME = time, WORK = human work, OBJECT = physical object, and SCOPE = specific part of work.

		FR			DE			EN		
	Type	train	dev	test	train	dev	test	train	dev	test
hipe-2020	LOC	3,089	774	854	1,740	588	595	–	384	181
	ORG	836	159	130	358	164	130	–	118	76
	PERS	2,525	679	502	1,166	372	311	–	402	156
	PROD	200	49	61	112	49	62	–	33	19
	TIME	276	68	53	118	69	49	–	29	17
ajmc	PERS	577	123	139	620	162	128	618	130	96
	WORK	378	99	80	321	70	74	467	116	95
	LOC	15	0	9	31	10	2	39	3	3
	OBJECT	10	0	0	6	4	2	3	0	0
	DATE	2	0	3	2	0	0	12	5	3
	SCOPE	639	169	129	758	157	176	684	162	151

newspapers in French, German, and English (19C-20C) and it contains 19,848 linked entities [10, 11, 12].

We also experimented with the classical commentaries data from the Ajax Multi-Commentary (`ajmc`) project that is composed of digitized 19C commentaries published in French, German, and English [26], annotated with both universal named entities (person, location, organisation) and domain-specific named entities (bibliographic references to primary and secondary literature).

Table 1 presents the statistics regarding the number and type of entities in the aforementioned datasets divided according to the training, development, and test sets.

3. Knowledge-based Contexts

One of the main challenges of NER applied to historical newspapers and classical commentaries concerns the digitization process of these heritage materials. The OCR output contains errors which produce noisy text and complications, similar to those studied in [27]. Introducing external grammatically correct contexts into NERC systems have been shown to have a positive impact over the entities identification [28]. It consists on adding complementary and related sentences, paragraphs or documents from external resources like Wikipedia or knowledge graphs (KG) to enrich the surrounding of an entity, which helps NERC systems on detecting the correct label. KGs structure information in a connected form, by representing entities (e.g., people, places) as nodes, and relationships between entities (e.g., being part of, being located in) as edges. Thus, we propose two main techniques for generating additional contexts:

- *Wikipedia Knowledge Retrieval Module*: We create a local instance of ElasticSearch[2], which provides dense vector field indexing and a k-nearest neighbor (kNN) search API. Given a query vector, this API obtains the k closest vectors and returns those documents as search hits.
- *Knowledge Graph Embedding Retrieval Module*: We produce English contexts by extending the indexing scheme to a knowledge graph embedding model over the Wikidata5m[3] [29] dataset.

3.1. Wikipedia Knowledge Retrieval Module

We download the latest (02/04/2022) XML dumps[4] of the French and German Wikipedia and transform them into plain text using the Wikipedia2Vec [30] utility[5]. We focus on French and German since for English we create another type of retrieval module which also contains Wikipedia paragraphs. Similar to Wang et al. [28], we define a document, inside our instance of ElasticSearch, as a triplet composed of a sentence, a title, and a paragraph. We create a dense vector index over the sentence embedding field computed with a pre-trained multilingual Sentence-BERT model[6] [31, 32]. During context retrieval, for a given sentence from the datasets described in Section 2, we compute its dense vector representation with the same multilingual Sentence-BERT pre-trained model and take it as a query to retrieve the top-k semantically similar documents based on a k-nearest neighbors algorithm (k-NN) cosine similarity search over the sentence embedding field (Figure 1).

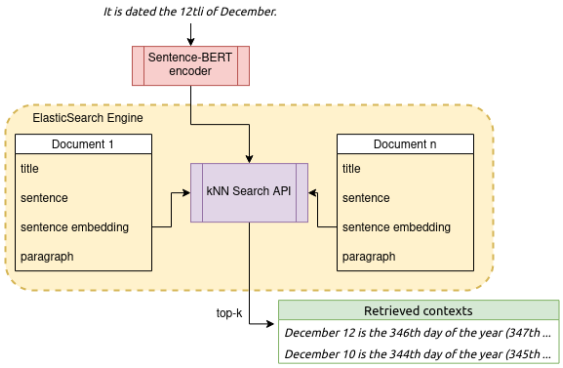

Figure 1: Context retrieval for the Wikipedia Knowledge Retrieval Module.

[2]We utilized ElasticSearch v8.1.
[3]https://deepgraphlearning.github.io/project/wikidata5m
[4]https://dumps.wikimedia.org/
[5]https://wikipedia2vec.github.io/wikipedia2vec/
[6]https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2

3.2. Knowledge Graph Embedding Retrieval Module

Wikidata5m is a large-scale KG with aligned entity descriptions. It integrates around five million Wikidata[7] entities, which are described in the first paragraph of the corresponding Wikipedia pages. We index the Wikidata5m dataset along with the dense vectors produced by the RotatE KG embedding model [33] pre-trained over the same dataset[8]. RotatE defines each relation between entities as a rotation from the source entity to the target entity in the dense vector space. In this case, we describe "an ElasticSearch document" as a triplet formed by an entity identifier, an entity description, and an entity embedding. We create a standard index on the entity identifier field and two dense vector indexes: the former on the entity embedding field, and the latter on the embeddings from the entity description field obtained with the same Sentence-BERT model as in the previous module. We propose two different methods for context retrieval (Figure 2) to evaluate the influence of the KG embedding on the semantic similarity:

- *KG Embedding Retrieval Module 1:* it takes into consideration the entity embedding index and follows the same principle utilized in the *Wikipedia Knowledge Retrieval Module*. For a given sentence, the top-*k* semantically similar documents are retrieved over the sentence embedding field.
- *KG Embedding Retrieval Module 2:* it retrieves the top-*1* semantically similar document. Then, a second search over the entity dense vector index is performed to retrieve the top-*k* similar documents based on the KG embeddings of the entities.

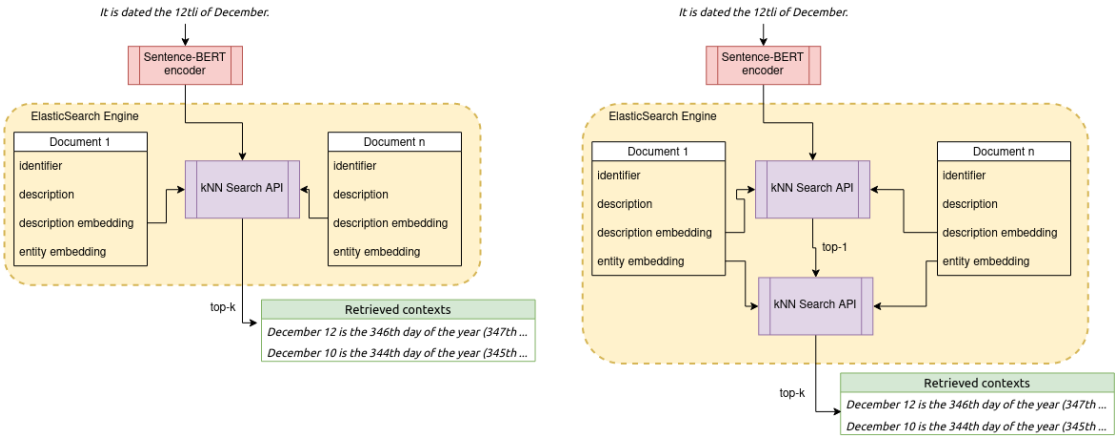

Figure 2: Context retrieval for KG Embedding Retrieval Module 1 (left) and KG Embedding Retrieval Module 2 (right).

4. Named Entity Recognition and Classification

In CLEF-HIPE-2022 [20], the named entity recognition and classification (NERC) task consists in the recognition and classification of entities, such as people and locations, within historical

[7] https://www.wikidata.org/
[8] https://graphvite.io/docs/latest/pretrained_model.html

multilingual newspapers and classical commentaries. According to the organizers [10], it is composed of two sub-tasks with different levels of difficulty:

- *Subtask 1.1 - NERC-Coarse:* the identification and categorization of entity mentions according to high-level entity types (e.g., Person, Location).
- *Subtask 1.2 - NERC-Fine:* the recognition and classification of entity mentions at different levels, finer-grained entity types and nested entities, up to one level of depth (nested entities).

4.1. NERC Architecture

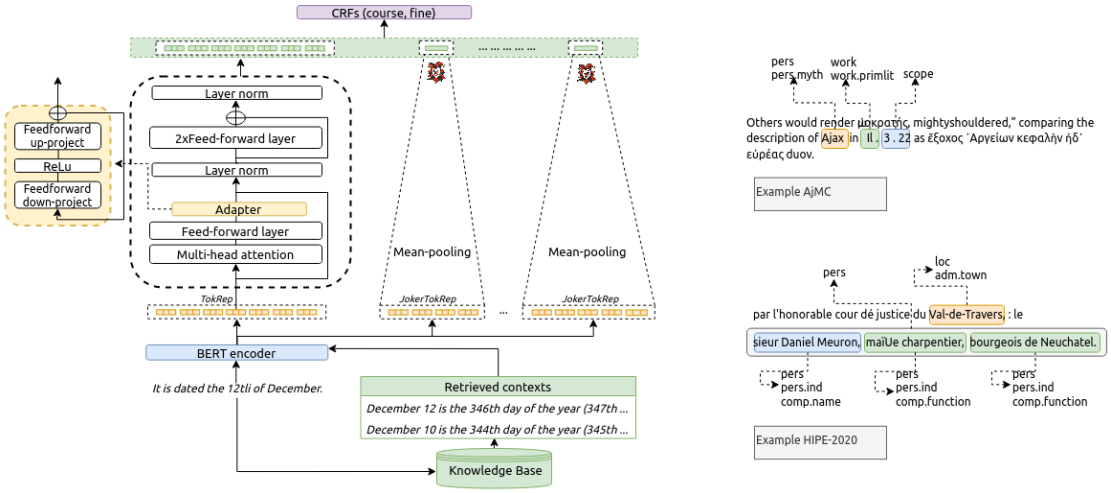

Figure 3: The NERC model architecture with additional contexts and examples of sentences from both datasets.

Our proposed architecture is presented in Figure 3. In the right, we detail our model that consists in a *Base Model* with new adapter layers, and the encoding of the additional contexts (*context jokers*). As an overview, after the contexts are generated for an initial sentence, we encode the tokens of the sentence with the *Base Model*, while the additional contexts are only passed through the BERT pre-trained model encoder. These representations are afterward concatenated, followed by the prediction CRF-based layers. In the left, we present two example sentences from `hipe-2020` and `ajmc`, for demonstrating the different levels of the entity types.

Base Model Our base model is based on the architecture proposed for CLEF-HIPE-2020 [13, 25] that consists in a hierarchical, multitask learning approach, with a fine-tuned encoder based on BERT. The previous model included an encoder with two Transformer [21] layers on top of the BERT pre-trained model encoder. This year, we add adapter modules to these layers [34]. The adapters are added to each Transformer layer after the projection following multi-headed attention. The adapter consists of a bottleneck which contains few parameters relative to the attention and feed-forward layers in the original model. This acts as a task adapter [24] for fine-grained NER. The attention modules in the Transformer layers adapt not

only to the task, but also to the noisy input which proved to increase performance of NER in such special conditions [25]. Finally, the multitask prediction layer consists of separate CRF layers[9].

Context Jokers  In order to include the additional contexts generated as explained in Section 3, we introduce the *context jokers*. Each additional context is passed through the BERT pre-trained model encoder[10] which is afterward mean-pooled along the sequence axis[11]. We call this representation the *context joker*. The *context jokers* are afterward concatenated with the sequential representation of the initial tokens of the sentence, as seen in Figure 3 and they are discarded at the moment of prediction. We call them *jokers* because we see them as wild cards unobtrusively inserted in the representation of the current sentence for improving the recognition of the fine-grained entities. However, we also consider that these jokers can affect the results in a way not immediately apparent and could also be harmful to the performance of a NERC system.

4.2. Experiments and Internal Results

CLEF-HIPE-2022 consists in assessing both tasks, NERC and EL in terms of precision (P), recall (R), and F-measure (F1) at macro and micro levels [35, 10][12]. Two evaluation scenarios are considered: strict (exact boundary matching) and fuzzy boundary matching. For our internal NERC results, we report only the `strict` matching (*NERC-Coarse* and *NERC-Fine*). Our experimental setup consists in a baseline model and three settings with different levels of knowledge-based contexts:

- `no-context`: *Base Model* with `bert-base-multilingual-cased`[13];
- `v0-language-specific`: context jokers are generated with *Wikipedia Knowledge Retrieval Module*;
- `v1-en-wk5m`: context jokers are generated with *KG Embedding Retrieval Module 1*;
- `v2-en-wk5m`: contexts jokers are generated with *KG Embedding Retrieval Module 2*.

French Our preliminary results for French, `hipe-2020` and `ajmc` datasets, are shown in Table 2. They reveal that generating contexts with *KG Embedding Retrieval Module 1 & 2* brings considerable improvements for HIPE even if our *Base Model* provides the higher precision for NERC-Coarse and the *Wikipedia Knowledge Retrieval Module* the higher recall for both granularities. Adding any type of context to `ajmc` seems to slightly affect the precision while

[9] There is a CRF layer for each level of the entity types (NE-COARSE-LIT, NE-COARSE-METO, NE-FINE-LIT, NE-FINE-METO, NE-FINE-COMP, NE-NESTED), thus six layers. If a dataset does not have fine-grained entities (e.g., English in `hipe-2020`, we maintain the same numbers of layers, and the model will learn to predict no entity.

[10] We do not utilize in this case the additional Transformer layers with adapters, since these were specifically proposed for noisy text and they do not bring any increase in performance as observed by Boroş et al. [25].

[11] The maximum length of each context corresponds to the one handled by the language model. Thus, for example, for a BERT-base model, the maximum is 512.

[12] We utilized the HIPE-scorer https://github.com/hipe-eval/HIPE-scorer.

[13] https://huggingface.co/bert-base-multilingual-cased

Table 2
NERC results on French (Internal).

	hipe-2020			ajmc		
	P	R	F1	P	R	F1
	no-context					
NERC-Coarse	**0.765**	0.755	0.76	**0.833**	0.792	0.812
NERC-Fine	0.651	0.665	0.658	**0.691**	**0.697**	**0.694**
	v0-language-specific					
NERC-Coarse	0.758	**0.768**	0.763	0.83	0.800	0.815
NERC-Fine	0.632	**0.694**	0.662	0.69	**0.697**	0.693
	v1-en-wk5m					
NERC-Coarse	0.762	0.767	**0.765**	0.83	0.803	0.816
NERC-Fine	0.643	0.69	0.666	0.625	0.633	0.629
	v2-en-wk5m					
NERC-Coarse	0.756	0.758	0.757	0.828	**0.814**	**0.821**
NERC-Fine	**0.655**	0.692	**0.673**	0.69	**0.697**	0.693

Table 3
NERC results on German (Internal).

	hipe-2020			ajmc		
	P	R	F1	P	R	F1
	no-context					
NERC-Coarse	0.754	0.73	0.742	0.910	0.877	0.893
NERC-Fine	0.598	0.657	0.626	0.895	0.872	0.883
	v0-language-specific					
NERC-Coarse	**0.761**	0.756	0.759	0.933	0.877	0.904
NERC-Fine	0.644	0.684	0.664	**0.912**	0.869	0.890
	v1-en-wk5m					
NERC-Coarse	0.759	0.767	0.763	0.930	0.898	0.913
NERC-Fine	**0.677**	0.684	**0.681**	0.909	**0.887**	**0.898**
	v2-en-wk5m					
NERC-Coarse	0.76	**0.774**	**0.767**	**0.935**	**0.906**	**0.920**
NERC-Fine	0.654	**0.701**	0.676	0.906	**0.887**	0.897

the contexts produced by the *KG Embedding Retrieval Module 2* has a positive impact for the NERC-Coarse recall.

German As for German, our preliminary results presented in Table 3 show the larger improvements when applying contexts for both ajmc and hipe-2020, specially with *KG Embedding*

Table 4
NERC results on English (Internal).

	hipe-2020			ajmc		
	P	R	F1	P	R	F1
	no-context					
NERC-Coarse	**0.604**	0.563	**0.583**	0.789	0.859	0.823
NERC-Fine	–	–	–	0.740	0.833	0.784
	v1-en-wk5m					
NERC-Coarse	0.565	**0.601**	**0.583**	0.828	**0.871**	0.849
NERC-Fine	–	–	–	0.755	**0.839**	0.795
	v2-en-wk5m					
NERC-Coarse	0.565	**0.601**	**0.583**	**0.86**	0.868	**0.864**
NERC-Fine	–	–	–	**0.782**	0.836	**0.808**

Retrieval Module 1 & 2. We assume that this is due the considerably smaller training dataset than for the other languages.

English Our preliminary results for English, shown in Table 4, indicate that generating contexts with *KG Embedding Retrieval Module 1 & 2* brings considerable improvements on ajmc for both granularities. Adding contexts to hipe-2020 has a double effect. They negatively impact precision while improving recall. This is due to the lack of English training documents and the fact that the contexts were generated using the French and German hipe-2020 training datasets[14].

4.3. CLEF-HIPE-2022 Results

The official CLEF-HIPE-2022 competition was restricted to two submissions. We, thus, selected our baseline (no-context) and our best context generator models (v2-en-wk5m). In order to improve the performance of our models, we stacked, for each language, a language-specific language model. For English, we add bert-base-cased[15], while for French and German, we add the open-source French and German Europeana BERT models pretrained on the open source Europeana digitized newspapers provided by The European Library and published by the MDZ Digital Library team (dbmdz)[16].

French, German, English Our official results for French, German and English are shown in Tables 5, 6, and 7 respectively. Adding contexts with the *KG Embedding Retrieval Module 2* reveals

[14]These training sets were combined and used for training the model. Since the English hipe-2020 has only NERC-Coarse entities, we discarded the NERC-Fine and the nested entities from the the French and German hipe-2020, before training.

[15]We utilized the English BERT model https://huggingface.co/bert-base-cased.

[16]We utilized the bert-base-french-europeana-cased and bert-base-german-europeana-cased from https://huggingface.co/dbmdz/.

Table 5
NERC results on French (CLEF-HIPE-2022).

	hipe-2020			ajmc		
	P	R	F1	P	R	F1
	no-context					
NERC-Coarse	**0.786**	**0.831**	**0.808**	0.78	0.817	0.798
NERC-Fine	0.679	0.767	0.720	0.623	0.669	0.645
	v2-en-wk5m					
NERC-Coarse	0.782	0.827	0.804	**0.81**	**0.842**	**0.826**
NERC-Fine	**0.697**	**0.779**	**0.736**	**0.646**	**0.694**	**0.669**

Table 6
NERC results on German (CLEF-HIPE-2022).

	hipe-2020			ajmc		
	P	R	F1	P	R	F1
	no-context					
NERC-Coarse	0.757	**0.792**	0.774	0.913	0.903	0.908
NERC-Fine	**0.658**	**0.724**	**0.689**	0.860	**0.901**	0.880
	v2-en-wk5m					
NERC-Coarse	**0.78**	0.787	**0.784**	**0.946**	**0.921**	**0.934**
NERC-Fine	0.657	0.71	0.682	**0.915**	0.898	**0.906**

Table 7
NERC results on English (CLEF-HIPE-2022).

	hipe-2020			ajmc		
	P	R	F1	P	R	F1
	no-context					
NERC-Coarse	0.604	**0.619**	0.612	**0.831**	0.851	0.841
NERC-Fine	–	–	–	0.745	0.822	0.781
	v2-en-wk5m					
NERC-Coarse	**0.624**	0.617	**0.620**	0.824	**0.876**	**0.850**
NERC-Fine	–	–	–	**0.754**	**0.848**	**0.798**

a general improvement for all languages for ajmc. The additional contexts for hipe-2020 behave differently. For French, our baseline model performed better for coarse granularity with exact boundary matching. For German, contexts improved performance for coarse granularity while slightly negatively affecting fine granularity. Finally, for English, the *KG Embedding Retrieval Module 2* boosted the performance for the coarse-grained entities.

5. Entity Linking

In CLEF-HIPE-2022, the EL task consists in the disambiguation of named entities using two settings:

- **EL-only:** The ground-truth regarding the entity mentions is provided, hence the entity disambiguation runs uses the gold entity mentions of NERC and the only variable is the EL system;
- **End-to-end EL:** No prior knowledge of the named entities is given, therefore EL has to be performed over the named entities predicted with the NERC models (`no-context` and `v2-en-wk5m`).

Table 8
EL results (CLEF-HIPE-2022) for the `hipe-2020` dataset.

Language	Setting	P	R	F1	P	R	F1
		relaxed			strict		
French	EL-only	0.620	0.620	0.620	0.602	0.602	0.602
	`no-context`	**0.563**	**0.594**	**0.578**	**0.546**	**0.576**	**0.560**
	`v2-en-wk5m`	0.560	0.592	0.576	0.543	0.574	0.558
German	EL-only	0.497	0.497	0.497	0.481	0.481	0.481
	`no-context`	0.453	**0.473**	0.463	0.438	**0.458**	0.447
	`v2-en-wk5m`	**0.462**	0.466	**0.464**	**0.446**	0.451	**0.449**
English	EL-only	0.546	0.546	0.546	0.546	0.546	0.546
	`no-context`	**0.471**	0.465	0.468	**0.471**	0.465	0.468
	`v2-en-wk5m`	0.463	**0.474**	**0.469**	0.463	**0.474**	**0.469**

Our EL model is based on the same neural approach that we proposed for CLEF-HIPE-2020 [13]. It is combined with a filtering process to analyze the historical mentions and to disambiguate them using the Wikidata KB [36]. Combining information from Wikipedia, Wikidata, and DBpedia allows a thorough analysis of the characteristics of the entities and, as in CLEF-HIPE-2020, it helped our method to correctly disambiguate mentions in historical documents. Table 8 presents our EL scores for CLEF-HIPE-2022 in terms of P, R, and F1 for the `hipe-2020` dataset. It can be observed that adding contexts to German and English has a negative impact on the recall which is consistent with our NERC results (cf. Table 6 and Table 7). Results also show that applying additional contexts to French does not increase performances. The extended results and ranking of CLEF-HIPE-2022 are available at the official website of the evaluation campaign[17].

[17] https://hipe-eval.github.io/HIPE-2022/results

6. Conclusions

For the participation of our team (L3i) in CLEF-HIPE-2022, we proposed two neural-based methods for the tasks of NERC and EL. We conclude, for NERC, that our *joker*-based approach generally performed well, due to the additional KG-based contexts and model improvements in regards to the treatment of such contexts. For EL, the model we proposed for CLEF-HIPE-2020 confirmed its good performance, with and without context. Finally, we consider that external knowledge has brought clear improvements to both our approaches and future work on this subject could furthermore prove the utility and importance of high-quality symbolic knowledge.

Acknowledgments

This work has been supported by the ANNA (2019-1R40226) and TERMITRAD (2020-2019-8510010) projects funded by the Nouvelle-Aquitaine Region, France. We would like to also thank Nicolas Sidère and Jean-Loup Guillaume for the insightful discussions.

References

[1] F. Boschetti, A. Cimino, F. Dell'Orletta, G. Lebani, L. Passaro, P. Picchi, G. Venturi, S. Montemagni, A. Lenci, Computational analysis of historical documents: An application to italian war bulletins in world war i and ii, in: Workshop on Language resources and technologies for processing and linking historical documents and archives (LRT4HDA 2014), ELRA, 2014, pp. 70–75.

[2] M. Rovera, F. Nanni, S. P. Ponzetto, Providing advanced access to historical war memoirs through the identification of events, participants and roles (2019).

[3] A. Cybulska, P. Vossen, Historical event extraction from text, in: Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, 2011, pp. 39–43.

[4] E. Boschee, P. Natarajan, R. Weischedel, Automatic extraction of events from open source text for predictive forecasting, in: Handbook of Computational Approaches to Counterterrorism, Springer, 2013, pp. 51–67.

[5] S. Oberbichler, E. Boroş, A. Doucet, J. Marjanen, E. Pfanzelter, J. Rautiainen, H. Toivonen, M. Tolonen, Integrated interdisciplinary workflows for research on historical newspapers: Perspectives from humanities scholars, computer scientists, and librarians, Journal of the Association for Information Science and Technology 73 (2022) 225–239.

[6] S. Hechl, P.-C. Langlais, J. Marjanen, S. Oberbichler, E. Pfanzelter, Digital interfaces of historical newspapers: opportunities, restrictions and recommendations, Journal of Data Mining & Digital Humanities (2021).

[7] S. Oberbichler, E. Pfanzelter, Topic-specific corpus building: A step towards a representative newspaper corpus on the topic of return migration using text mining methods, Journal of Digital History 1 (2021) 74–98.

[8] M. Ehrmann, A. Hamdi, E. Linhares Pontes, M. Romanello, A. Douvet, A Survey of Named

Entity Recognition and Classification in Historical Documents, ACM Computing Surveys (2022 (to appear)). URL: https://arxiv.org/abs/2109.11406.
[9] E. Linhares Pontes, L. A. Cabrera-Diego, J. G. Moreno, E. Boros, A. Hamdi, N. Sidère, M. Coustaty, A. Doucet, Entity linking for historical documents: challenges and solutions, in: International Conference on Asian Digital Libraries, Springer, 2020, pp. 215–231.
[10] M. Ehrmann, M. Romanello, S. Bircher, S. Clematide, Introducing the CLEF 2020 HIPE shared task: Named entity recognition and linking on historical newspapers, in: J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, F. Martins (Eds.), Advances in information retrieval, Springer International Publishing, Cham, 2020, pp. 524–532.
[11] M. Ehrmann, M. Romanello, A. Flückiger, S. Clematide, Overview of clef hipe 2020: Named entity recognition and linking on historical newspapers, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2020, pp. 288–310.
[12] M. Ehrmann, M. Romanello, A. Flückiger, S. Clematide, Extended overview of clef hipe 2020: named entity processing on historical newspapers, in: CEUR Workshop Proceedings, 2696, CEUR-WS, 2020.
[13] E. Boros, E. L. Pontes, L. A. Cabrera-Diego, A. Hamdi, J. G. Moreno, N. Sidère, A. Doucet, Robust named entity recognition and linking on historical multilingual documents, in: Conference and Labs of the Evaluation Forum (CLEF 2020), volume 2696, CEUR-WS Working Notes, 2020, pp. 1–17.
[14] K. Labusch, C. Neudecker, Named entity disambiguation and linking historic newspaper ocr with bert., in: CLEF (Working Notes), 2020.
[15] S. Schweter, L. März, Triple e-effective ensembling of embeddings and language models for ner of historical german., in: CLEF (Working Notes), 2020.
[16] V. Provatorova, S. Vakulenko, E. Kanoulas, K. Dercksen, J. M. van Hulst, Named entity recognition and linking on historical newspapers: Uva. ilps & rel at clef hipe 2020 (2020).
[17] T. Kristanti, L. Romary, Delft and entity-fishing: Tools for clef hipe 2020 shared task, in: CLEF 2020-Conference and Labs of the Evaluation Forum, volume 2696, CEUR, 2020.
[18] C. B. El Vaigh, G. Le Noé-Bienvenu, G. Gravier, P. Sébillot, Irisa system for entity detection and linking at clef hipe 2020, in: CEUR Workshop Proceedings, 2020.
[19] P. J. O. Suárez, Y. Dupont, G. Lejeune, T. Tian, Sinner@ clef-hipe2020: Sinful adaptation of sota models for named entity recognition in french and german, in: CLEF (Working Notes), 2020.
[20] M. Ehrmann, M. Romanello, A. Doucet, S. Clematide, Introducing the hipe 2022 shared task: Named entity recognition and linking in multilingual historical documents, in: European Conference on Information Retrieval, Springer, 2022, pp. 347–354.
[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).
[22] N. Kolitsas, O.-E. Ganea, T. Hofmann, End-to-end neural entity linking, in: Proceedings of the 22nd Conference on Computational Natural Language Learning, 2018, pp. 519–529.
[23] S.-A. Rebuffi, H. Bilen, A. Vedaldi, Learning multiple visual domains with residual adapters, Advances in neural information processing systems 30 (2017).
[24] J. Pfeiffer, I. Vulić, I. Gurevych, S. Ruder, MAD-X: An Adapter-Based Framework for Multi-

Task Cross-Lingual Transfer, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 7654–7673. URL: https://aclanthology.org/2020.emnlp-main.617. doi:10.18653/v1/2020.emnlp-main.617.

[25] E. Boroş, A. Hamdi, E. L. Pontes, L.-A. Cabrera-Diego, J. G. Moreno, N. Sidere, A. Doucet, Alleviating digitization errors in named entity recognition for historical documents, in: Proceedings of the 24th conference on computational natural language learning, 2020, pp. 431–441.

[26] M. Romanello, S. Najem-Meyer, B. Robertson, Optical character recognition of 19th century classical commentaries: the current state of affairs, in: The 6th International Workshop on Historical Document Imaging and Processing, 2021, pp. 1–6.

[27] S. Mayhew, T. Tsygankova, D. Roth, ner and pos when nothing is capitalized, 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), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 6256–6261. URL: https://aclanthology.org/D19-1650. doi:10.18653/v1/D19-1650.

[28] X. Wang, Y. Shen, J. Cai, T. Wang, X. Wang, P. Xie, F. Huang, W. Lu, Y. Zhuang, K. Tu, et al., Damo-nlp at semeval-2022 task 11: A knowledge-based system for multilingual named entity recognition, arXiv preprint arXiv:2203.00545 (2022).

[29] X. Wang, T. Gao, Z. Zhu, Z. Liu, J. Li, J. Tang, Kepler: A unified model for knowledge embedding and pre-trained language representation, arXiv preprint arXiv:1911.06136 (2019).

[30] I. Yamada, A. Asai, J. Sakuma, H. Shindo, H. Takeda, Y. Takefuji, Y. Matsumoto, Wikipedia2Vec: An efficient toolkit for learning and visualizing the embeddings of words and entities from Wikipedia, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, 2020, pp. 23–30.

[31] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using Siamese BERT-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), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3982–3992. URL: https://aclanthology.org/D19-1410. doi:10.18653/v1/D19-1410.

[32] N. Reimers, I. Gurevych, Making monolingual sentence embeddings multilingual using knowledge distillation, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 4512–4525.

[33] Z. Sun, Z.-H. Deng, J.-Y. Nie, J. Tang, Rotate: Knowledge graph embedding by relational rotation in complex space, arXiv preprint arXiv:1902.10197 (2019).

[34] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, S. Gelly, Parameter-efficient transfer learning for nlp, in: International Conference on Machine Learning, PMLR, 2019, pp. 2790–2799.

[35] J. Makhoul, F. Kubala, R. Schwartz, R. Weischedel, et al., Performance measures for information extraction, in: Proceedings of DARPA broadcast news workshop, Herndon, VA, 1999, pp. 249–252.

[36] E. Linhares Pontes, L. A. Cabrera-Diego, J. G. Moreno, E. Boros, A. Hamdi, A. Doucet, N. Sidere, M. Coustaty, Melhissa: a multilingual entity linking architecture for historical press articles, International Journal on Digital Libraries 23 (2022) 133–160.