Gazetteer Enhanced Named Entity Recognition for Code-Mixed Web Queries

Besnik Fetahu, Anjie Fang, Oleg Rokhlenko, Shervin Malmasi
Amazon.com, Inc.
Seattle, WA, USA
{besnikf,njfn,olegro,malmasi}@amazon.com

ABSTRACT

Named entity recognition (NER) for Web queries is very challenging. Queries often do not consist of well-formed sentences, and contain very little context, with highly ambiguous queried entities. Code-mixed queries, with entities in a different language than the rest of the query, pose a particular challenge in domains like e-commerce (e.g. queries containing *movie* or *product names*).

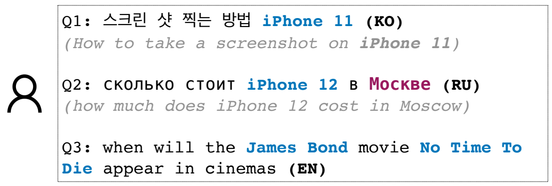
This work tackles NER for code-mixed queries, where entities and non-entity query terms co-exist simultaneously in different languages. Our contributions are twofold. First, to address the lack of code-mixed NER data we create EMBER, a large-scale dataset in six languages with four different scripts. Based on Bing query data, we include numerous language combinations that showcase real-world search scenarios. Secondly, we propose a novel gated architecture that enhances existing multi-lingual Transformers with a Mixture-of-Experts model to dynamically infuse multi-lingual gazetteers, allowing it to simultaneously differentiate and handle entities and non-entity query terms in multiple languages.

Experimental evaluation on code-mixed queries in several languages shows that our approach efficiently utilizes gazetteers to recognize entities in code-mixed queries with an F1=68%, an absolute improvement of +31% over a non-gazetteer baseline.

ACM Reference Format:
Besnik Fetahu, Anjie Fang, Oleg Rokhlenko, Shervin Malmasi. 2021. Gazetteer Enhanced Named Entity Recognition for Code-Mixed Web Queries. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3404835.3463102

1 INTRODUCTION

NER is crucial for query understanding. In Web search, named entities (NE) can be used for filtering or re-ranking of top−k candidate documents [8]. In other scenarios, like voice search, voice assistants can better understand the query intent and thus re-route to an appropriate answering mechanism. Depending on the queried concepts, a recurring aspect is the use of code-mixed queries, with query terms in multiple languages [2, 9]. Figure 1 shows example monolingual and code-mixed Web queries in different languages.



Q1: 스크린 샷 찍는 방법 **iPhone 11** **(KO)**
*(How to take a screenshot on **iPhone 11**)*

Q2: сколько стоит **iPhone 12** в **Москве** **(RU)**
(how much does iPhone 12 cost in Moscow)

Q3: when will the **James Bond** movie **No Time To Die** appear in cinemas **(EN)**

Figure 1: Example code-mixed and monolingual queries from our dataset in different languages and scripts.

Advances in pre-trained multi-lingual transformers [3] have made it easier to parse such code-mixed queries. Yet, as shown in our experiments, performing NER on such queries is still challenging as the models lack the required entity knowledge across all languages.

Furthermore, adapting NER models to emerging entities (EE) in Web queries is challenging. EEs are prevalent in Web queries [18] and have variable surface forms [12], making the application of pre-trained NER models difficult.

In this work, we tackle the problem of NER for code-mixed Web queries, where NEs can be in a language or multiple languages that are different from the non-NE query terms. Our novel approach enhances pre-trained multi-lingual Transformers, here XLM-RoBERTa (XLMR) [3], with external entity knowledge. We inject language-specific gazetteer information, using an LSTM-based contextualized embedding to represent potential NEs matches in queries. The textual and gazetteer information are combined dynamically through a Mixture-of-Experts model (MoE) [21], which conditionally determines how to combine both representations of each token for the NER task. Using the dynamically-computed representation from different modalities, our approach can flexibly encode text and external knowledge from multi-lingual resources like Wikidata.

Evaluation on real Web queries [5] on several settings of code-mixed queries, show that our approach is able to capture NEs in several languages, irrespective of the original query language. Furthermore, for EEs, where the training data does not contain a specific target set of NEs, the dynamic integration of text and gazetteer representations through MoE allows us to adapt quickly to EEs. Comparing our approach to a multi-lingual NER (e.g. the XLMR baseline), we achieve an NER performance improvement of more than 30% absolute points improvement in terms of F1.

In this work, we make the following contributions:

- We introduce mLOWNER, a multi-lingual NER training dataset for short texts in 6 languages (4 scripts); and EMBER, a code-mixed NER dataset covering the same languages and scripts.

SIGIR '21, July 11–15, 2021, Virtual Event, Canada
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8037-9/21/07…$15.00
https://doi.org/10.1145/3404835.3463102

- A multi-lingual NER approach for code-mixed queries that flexibly encodes text and external multi-lingual gazetteers;
- An evaluation study on code-mixed queries, showing the potential and limitations of our approach.

2 RELATED WORK

NER is a fundamental task in natural language processing (NLP), and has seen extensive use in information retrieval [1]. It is often applied to extract NEs and their types, with the purpose of query understanding, which in turn can be used in different verticals (e.g. travel search) [4], or improve ranking [8]. Furthermore, NEs can be used to re-adjust term weights in a query for query expansion [13].

The rapid growth of multilingual resources on the web has helped develop cross-lingual IR approaches [20]. In particular, queries can often contain terms in more than one language (i.e. code-mixed queries). This is especially the case, in e-commerce search or searching for creative works (e.g. movies, songs, etc.), where NEs can be in a different language than the rest of the query terms [7]. While NER models aid IR systems, their application for code-mixed Web queries remains challenging [2, 10]. Moreover, queries are too short to provide rich contextual information [8, 15], and in turn more difficult. In this paper, we investigate how to effectively identify NEs from code-mixed Web queries.

Bhargava et al. in a recent work [2] propose an NER hybrid approach for code-mixed queries, consisting of a gazetteer and tree based identifier. The gazetteer based identifier is rule based, and as such it cannot deal with unseen entities in the gazetteer. Gupta et al. [9] leverage linguistic features to train a conditional random field (CRF) model, where the output is further processed using multi-lingual gazetteer lists. Gazetteers have been proven to be effective for NER [16]. Yet, a major issue is on how to encode gazetteer knowledge and jointly combine it with textual representation or other modalities. Furthermore, using hand-crafted features extracted from gazetteers raises issues on how to replace gazetteers during test time without any pre-training. This is one of the major bottlenecks of existing work [2, 16], and a major diverging point of our work. We propose approaches that efficiently incorporate gazetteers on pre-trained NER models for code-mixed queries, without having to retrain models for a new application scenario.

Finally, an important aspect is the ability to handle multi-linguality. Priyadharshini et al. [19] propose the use of pre-trained multilingual word embeddings to encode sentences for NER. We follow such guidelines, however, we use the more recent transformer based cross-lingual model XLMR [3]. XLMR is suitable for our task to encode Web queries in different languages, as it has been pre-trained using a multilingual corpus of more than 100 languages. We differ from [19] as we are able to simultaneously handle code-mixed queries, and through the efficient integration of gazetteers we can recognize entities with optimal performance.

3 CODE-MIXED QUERY NER APPROACH

To perform NER for code-mixed queries, we adapt our existing gazetteer NER model architecture [17] in order to effectively handle: (i) NEs and non-query terms with varying language and script (in more extreme cases NEs are in several languages), and (ii) noisy query terms due to automated speech recognition (ASR) in voice search scenarios. Our approach has three components that address these aspects: (i) multi-lingual text encoder, (ii) external gazetteer encoder with entities from openly available corpora, and (iii) dynamic integration of textual and gazetteer query representations.

Given user query $\mathbf{q} = \{q_1, \ldots, q_n\}$ consisting of N query tokens, we compute the query representation as follows.

Multi-Lingual Textual Representation. We represent the query tokens using the multi-lingual transformer XLMR [3], which has several advantages as it contains prior knowledge about query terms in different languages. For $\mathbf{q}$ we compute the query text representation indicated as $\mathbf{h_q} \in \mathbb{R}^{N \times L}$.

However, as shown later in our experimental evaluation, XLMR alone is insufficient, as these models are trained on *monolingual* instances, and thus cannot efficiently handle code-mixed queries, with terms transitioning from one language to another.

Gazetteer Encoder. Gazetteers allow us to inject explicit information about NEs from a target language or set of languages. Gazetteers are easy to obtain from open resources like Wikipedia or Wikidata. A gazetteer $\mathcal{G}$ consists of entities and their type, e.g. ⟨"`No Time to Die`", `CreativeWork`⟩.

A token or span of tokens from $\mathbf{q}$ are matched to the longest matching entry in $\mathcal{G}$, yielding a sparse encoding $\mathbf{g_q} \in \mathbb{N}^{N \times k}$, where $\mathbf{g}_q = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, and $\mathbf{x}_i \in (0, 1)_k$ is a binary vector of length k, with k being the number of entity types in $\mathcal{G}$ encoded in BIO format. We compute a dense representation for $\mathbf{g}_q$ by projecting it through a dense layer θ and additionally encoding it using a bi-directional LSTM [11], $\mathbf{G}_q = \left[\overrightarrow{\texttt{LSTM}}(\theta[\mathbf{g}_q]); \overleftarrow{\texttt{LSTM}}(\theta[\mathbf{g}_q])\right] \in \mathbb{R}^{N \times L}$.

Joint Textual and Gazetteer Representation. Through the mixture of experts approach (MoE) [21] we jointly encode the textual (syntactic) and contextual gazetteer representations, namely $\mathbf{h}_q$ and $\mathbf{G}_q$. At token level, MoE dynamically adjusts the weight of the two representations to correctly predicting the token NER class. Since we only have two representations, we use a Sigmoid function to split the importance considering both representations:

$$\mathbf{w}_{moe} = \sigma\left(\Lambda[\mathbf{h}_q; \mathbf{G}_q]^T\right) \tag{1}$$

$$\mathbf{h} = \mathbf{w}_{moe} \cdot \mathbf{h}_q + (1 - \mathbf{w}_{moe}) \cdot \mathbf{G}_q \tag{2}$$

where, $\Lambda \in \mathbb{R}^{2L}$ are trainable parameters. From $\mathbf{h}$ using a CRF layer [14] we compute the token NER tags. The model is trained in two stages. In the first stage, the gazetteer and the MoE modules are trained, ensuring that the model is not biased towards XLMR's pre-trained knowledge. While, in the second stage all components are jointly trained to optimize for NER performance.

4 DATA AND EXPERIMENTAL SETUP

We now describe our data, languages, and NER evaluation scenarios. The data and code are publicly available at the paper link.[1]

4.1 Evaluation Scenarios

We experiment with 6 languages (English, Dutch, Russian, Turkish, Farsi, and Korean) under the following scenarios:

- **Monolingual**: monolingual query and NE tokens.
- **Code-Mixed**: NE tokens in a different language than the rest of the query terms.

[1] https://registry.opendata.aws/code-mixed-ner

- **Multi-Lingual Code-Mixed**: queries with NEs in several languages to avoid bias where all NEs are in the same language.
- **Noisy Code-Mixed**: noisy non-NE query terms of the same language as NE are added to ensure that the language/script switch and entity boundaries differ in order to test NER robustness.

4.2 Datasets

We now describe our data contributions: (1) mLOWNER: a low-context multi-lingual NER training dataset; (2) EMBER: a multi-lingual NER dataset based on Bing queries; and (3) multi-lingual gazetteers extracted from WikiData.

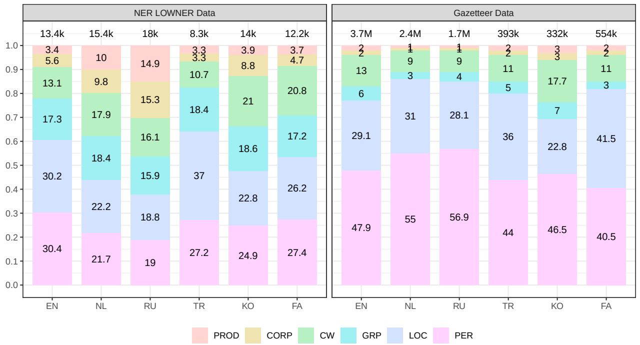


Figure 2: Distribution (in %) of NER training data and gazetteer data per class for the different evaluation languages.

mLOWNER: is an extension of the English LOWNER dataset [17] to the languages in Figure 2, which we use for training the multi-lingual NER models. It consists of low-context sentences from Wikipedia annotated using the NER class taxonomy from [6] (Person, Location, Group, CreativeWork, Corporation, Product). Figure 2 shows detailed the dataset statistics in 6 languages.

Multi-Lingual Gazetteers: Multi-lingual entity names are extracted from Wikidata, specifically from entity types matching our NER taxonomy. We construct gazetteers for all the language pairs in Table 1. Figure 2 shows gazetteer statistics for all languages.

EMBER: To evaluate NER models on code-mixed queries, we first process 10 million distinct Bing search queries (via the ORCAS dataset [5]) to extract templates, and then generate different code-mixed queries for evaluation. We generate 471k queries per language, with an average query length of 4 ± 1.75 tokens per query. **Query Templatization:** For each query in the ORCAS dataset we run an existing NER model[2] to detect NEs, which we map to their entity type (e.g. CW, ORG etc.). This allows us to collapse multiple queries into a single template, if the resulting query and the entity type are the same. For instance, for the queries *"what is **westworld**"*, *"what is **the african union**"*, we construct the query templates *"what is* ⟨CW|ORG⟩*"*. We obtain 97K unique English templates, which are then translated into the languages in Figure 2.

Evaluation Datasets: We generate several datasets for the different evaluation settings: (i) *mono-lingual*, the query and NE terms are in the same language, (ii) *code-mixed*, the non-NE query terms and the NE are in different languages, (iii) *noisy code-mixed*, similar to (ii), with additionally injected noise consisting of random bi-grams before and after the NE slots, and (iv) *multi-lingual code-mixed*, similar to (ii), but with NEs being in one or more languages.

We extract multi-lingual entity labels from Wikidata and use them to replace the entity placeholders in the query templates in the different languages. Table 1 shows the different dataset configurations along with the query language and the NE language(s). We use the dataset IDs to refer to the various evaluation configurations.

ID	*dataset type*	*lang. pairs*
Mono	*monolingual*	$\langle q : l_i \rangle,\ l_i \in \{\text{EN, NL, RU, TR, KO, FA}\}$
CoMix	*code-mixed*	$\langle q : l_i;\ e : l_j \rangle,\ l_i = \{\text{EN}\},\ l_j = \{\text{NL, RU, TR, KO, FA}\}$ $\langle q : l_i;\ e : l_j \rangle,\ l_i = \{\text{NL, RU, TR, KO, FA}\},\ l_j = \{\text{EN}\}$
NoisyCoMix	*noisy code-mixed*	$\langle q : \text{RU};\ e : \text{EN} \rangle, \langle q : \text{EN};\ e : \text{RU} \rangle$
MulCoMix	*multi-lingual code-mixed*	$\langle q : l_i;\ \langle e_1 : EN,\ e_2 : l_i \rangle \rangle,\ l_i = \{\text{NL, RU, TR, KO, FA}\}$

Table 1: Evaluation scenarios: EMBER Mono, CoMix, NoisyCoMix with an average of 471k queries per dataset, whereas MulCoMix has an average of 90k queries.

4.3 Baseline

As a baseline we consider the XLMR model without gazetteer knowledge. We fine-tune on mLOWNER, which represents the ablation of our approach without the gazetteer knowledge.

5 RESULTS & DISCUSSION

5.1 Monolingual Query NER Evaluation

We first evaluate the NER approaches on the monolingual scenario, where query terms and NEs are in the same language. The results in Table 2 are important as they highlight the difficulty of segmenting short textual snippets like Web queries, and can better contextualize our results from the code-mixed experiments that follow.

lang.	**P**		**R**		**F1**	
	B	Ours	B	Ours	B	Ours
EN	0.409	**0.659** (▲ 25.0)	0.424	**0.579** (▲ 15.5)	0.417	**0.616** (▲ 19.9)
NL	0.367	**0.739** (▲ 37.2)	0.457	**0.749** (▲ 29.2)	0.407	**0.744** (▲ 33.7)
RU	0.377	**0.755** (▲ 37.8)	0.357	**0.702** (▲ 34.5)	0.367	**0.728** (▲ 36.1)
TR	0.397	**0.619** (▲ 22.2)	0.439	**0.695** (▲ 25.6)	0.417	**0.655** (▲ 23.8)
KO	0.367	**0.656** (▲ 28.9)	0.408	**0.786** (▲ 37.8)	0.387	**0.715** (▲ 32.8)
FA	0.363	**0.679** (▲ 31.6)	0.402	**0.621** (▲ 21.9)	0.382	**0.649** (▲ 26.7)

Table 2: NER results on the Mono datasets. In brackets we show the difference in absolute points our approach against the baseline (B).

In all cases our approach significantly outperforms the baseline, with an average absolute F1 improvement of ▲28.8%. This increase is attributed to the external gazetteer knowledge, and the flexibility of our model to combine it together with the textual representation using the MoE module. Gazetteers offer several advantages as they can be replaced during inference time without any fine-tuning, which we show in the code-mixed query evaluation scenarios.

Emerging Entities. Web queries often contain emerging entities, unseen during training. Our data represents this scenario well: across different languages of Mono, the training and test set of entities have an overlap of less than 2%. Such information about new NEs can be added to our model's gazetteer knowledge without

[2] We used spaCy https://spacy.io/api

any re-training. Our strong performance confirms our assumption that gazetteers can be used to increase the flexibility of NER models to accurately spot emerging NEs.

5.2 Code-Mixed Query NER Evaluation

Code-mixed queries represent challenging cases for NER models. Even multi-lingual models like XLMR fail to utilize their pre-trained knowledge in parsing code-mixed queries, possibly because during pre-training the inputs are monolingual. Furthermore, code-mixed queries can introduce ambiguity as the same surface forms may point to different entities in different languages.

Table 3 shows the NER F1 results for the competing approaches for the `CoMix` datasets. Similar to the monolingual case, our approach obtains high improvements across all language pairs, with an average F1 increase of ▲29%. We notice a higher improvement when the query language is in EN, while the NEs are in another language, with an F1 score improvement of ▲34.8%. Whereas, for the case when the NEs are in EN, the improvement is ▲23.2%. The latter is a more realistic scenario, with English NEs in non-English queries (e.g. product search, movie titles, etc.).

Although not directly comparable, an interesting observation to make from Table 3 (lang. pairs with English NEs) and Table 2, is that our approach does not degrade in its NER performance. We obtain on average 2% better performance for `CoMix-EN` than `Mono-EN`. Similar is the performance for the baseline, however, with an average decrease of 1.7%. This confirms that for pre-trained models we can simply swap the external gazetteer knowledge, such that it reflects the particular use case, and our approach is able to perform NER on code-mixed queries with a small NER performance drop.

lang. pair	**P**		**R**		**F1**	
	B	Ours	B	Ours	B	Ours
⟨EN, NL⟩	0.386	**0.733** (▲ 34.7)	0.399	**0.627** (▲ 22.8)	0.392	**0.676** (▲ 28.4)
⟨EN, RU⟩	0.356	**0.780** (▲ 42.4)	0.378	**0.666** (▲ 28.8)	0.367	**0.719** (▲ 35.2)
⟨EN, TR⟩	0.366	**0.763** (▲ 39.7)	0.387	**0.639** (▲ 25.2)	0.376	**0.695** (▲ 31.9)
⟨EN, KO⟩	0.390	**0.843** (▲ 45.3)	0.363	**0.688** (▲ 32.5)	0.376	**0.758** (▲ 38.2)
⟨EN, FA⟩	0.347	**0.827** (▲ 48.0)	0.358	**0.695** (▲ 33.7)	0.352	**0.755** (▲ 40.3)
⟨NL, EN⟩	0.401	**0.717** (▲ 31.6)	0.484	**0.726** (▲ 24.2)	0.439	**0.721** (▲ 28.2)
⟨RU, EN⟩	0.406	**0.733** (▲ 32.7)	0.380	**0.656** (▲ 27.6)	0.392	**0.692** (▲ 30.0)
⟨TR, EN⟩	0.398	**0.533** (▲ 13.5)	0.434	**0.630** (▲ 19.6)	0.415	**0.578** (▲ 16.3)
⟨KO, EN⟩	0.389	**0.562** (▲ 17.3)	0.434	**0.615** (▲ 18.1)	0.410	**0.588** (▲ 17.8)
⟨FA, EN⟩	0.365	**0.646** (▲ 28.1)	0.366	**0.561** (▲ 19.5)	0.366	**0.601** (▲ 23.5)

Table 3: NER F1 scores on the `CoMix` datasets for the competing approaches (B=baseline).

5.3 Noisy Code-Mixed Query NER Evaluation

To assess if our model is simply exploiting language/script switch boundaries to predict NEs, we add noisy non-NE *bigrams* and *trigrams* of the same language as the NE, either before or after it.

The `NoisyCoMix` scenario is challenging due to false positives, and ensures that boundaries across languages/scrips are not only between NEs and non-NE terms, but rather irregular and mixed.

Table 4 shows the results for one language pair. The gap between the baseline and our approach remains the same, with an absolute F1 improvement of ▲35% and ▲27.7% for the ⟨EN, RU⟩ and ⟨RU, EN⟩ pairs, respectively. Similar patterns hold for other languages, not shown here for brevity.

These results are particularly encouraging, given that the noise has been injected for all NEs in our query set, thus, contributing to a major challenge. This may be considered similar to scenarios where the user query is given through voice, which often results in speech-to-text transcription errors.

lang. pair	*noise*	**P**		**R**		**F1**	
		B	Ours	B	Ours	B	Ours
⟨EN, RU⟩	bigram	0.277	**0.711** (▲ 43.4)	0.290	**0.571** (▲ 28.1)	0.283	**0.633** (▲ 35.0)
	trigram	0.278	**0.712** (▲ 43.4)	0.292	**0.571** (▲ 27.9)	0.285	**0.634** (▲ 34.9)
⟨RU, EN⟩	bigram	0.225	**0.535** (▲ 31.0)	0.211	**0.457** (▲ 24.6)	0.218	**0.493** (▲ 27.5)
	trigram	0.234	**0.547** (▲ 31.3)	0.223	**0.471** (▲ 24.8)	0.228	**0.506** (▲ 27.8)

Table 4: NER F1 scores on the `NoisyCoMix` datasets (B=Baseline).

5.4 Multi-Lingual Code-Mixed Queries

For queries containing more than one NE, in extreme cases we may see a combination of NEs in different languages, namely, NEs in the query language, and NEs in another language, such as English. Table 5 shows the NER results for the `MultiCoMix` scenario. Our approach achieves high improvements over the baseline. We simply combine the gazetteers from the different languages and use them during inference to spot the NEs in all languages.

On average we obtain an absolute F1 improvement of ▲21% over the XLMR baseline. This shows that our approach can simultaneously recognize NEs in different languages with high accuracy. For many of the language pairs the results are better when compared to the `CoMix` scenario, however, this is mainly due to the fact that the proportion of evaluation queries is lower, since we need to limit to queries that contain more than one NE. The potential improvement over the results in Table 3 can be attributed to the ability of the models in recognizing NEs for English.

lang. pairs	**P**		**R**		**F1**	
	B	Ours	B	Ours	B	Ours
⟨NL, ⟨NL, EN⟩⟩	0.478	**0.815** (▲ 33.7)	0.473	**0.700** (▲ 22.7)	0.475	**0.753** (▲ 27.8)
⟨RU, ⟨RU, EN⟩⟩	0.509	**0.793** (▲ 28.4)	0.318	**0.562** (▲ 24.4)	0.391	**0.658** (▲ 26.7)
⟨TR, ⟨TR, EN⟩⟩	0.524	**0.618** (▲ 9.4)	0.420	**0.555** (▲ 13.5)	0.466	**0.584** (▲ 11.8)
⟨KO, ⟨KO, EN⟩⟩	0.490	**0.654** (▲ 16.4)	0.358	**0.565** (▲ 20.7)	0.414	**0.606** (▲ 19.2)
⟨FA, ⟨FA, EN⟩⟩	0.498	**0.742** (▲ 24.4)	0.353	**0.489** (▲ 13.6)	0.413	**0.589** (▲ 17.6)

Table 5: NER F1 scores on the `MulCoMix` datasets (B=Baseline).

6 CONCLUSION

We presented a model, and new public datasets for code-mixed NER in search queries. We introduced two large datasets, mLOWNER and EMBER, covering 6 languages (4 scripts) that are suitable for training multi-lingual NER models. Additionally, we presented a Mixutre-of-Experts model to effectively integrate external entity knowledge into multilingual Transformer architectures.

Extensive experiments in several scenarios and covering numerous languages were performed, with our models achieving significant improvements over existing multi-lingual baselines, with more than ▲30% F1 NER improvement. We demonstrated that our models can adapt to emerging entities, simultaneously recognize entities in more than one language, and account for noisy terms in a query to perform the NER for code-mixed queries without any fine-tuning by simply swapping the external gazetteer knowledge. In the future we plan to extend this work to more languages, and evaluate Entity Linking and document retrieval for code-mixed queries.

ACKNOWLEDGMENTS

We would like to extend our gratitude to Eugene Agichtein, Marcus Collins, Giuseppe Castellucci, and Jason Ingyu Choi for their valuable input on this paper. We also thank the anonymous reviewers for their constructive remarks.

REFERENCES

[1] Krisztian Balog, Pavel Serdyukov, and Arjen P. de Vries. 2010. Overview of the TREC 2010 Entity Track. In *Proceedings of The Nineteenth Text REtrieval Conference, TREC 2010, Gaithersburg, Maryland, USA, November 16-19, 2010 (NIST Special Publication)*, Ellen M. Voorhees and Lori P. Buckland (Eds.), Vol. 500-294. National Institute of Standards and Technology (NIST). http://trec.nist.gov/pubs/trec19/papers/ENTITY.OVERVIEW.pdf

[2] Rupal Bhargava, Bapiraju Vamsi, and Yashvardhan Sharma. 2016. Named entity recognition for code mixing in indian languages using hybrid approach. *Facilities* 23, 10 (2016).

[3] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116* (2019).

[4] Brooke Cowan, Sven Zethelius, Brittany Luk, Teodora Baras, Prachi Ukarde, and Daodao Zhang. 2015. Named entity recognition in travel-related search queries. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. 3935–3941.

[5] Nick Craswell, Daniel Campos, Bhaskar Mitra, Emine Yilmaz, and Bodo Billerbeck. 2020. ORCAS: 18 Million Clicked Query-Document Pairs for Analyzing Search. *CoRR* abs/2006.05324 (2020). arXiv:2006.05324 https://arxiv.org/abs/2006.05324

[6] Leon Derczynski, Eric Nichols, Marieke van Erp, and Nut Limsopatham. 2017. Results of the WNUT2017 Shared Task on Novel and Emerging Entity Recognition. In *Proceedings of the 3rd Workshop on Noisy User-generated Text, NUT@EMNLP 2017, Copenhagen, Denmark, September 7, 2017*, Leon Derczynski, Wei Xu, Alan Ritter, and Tim Baldwin (Eds.). Association for Computational Linguistics, 140–147. https://doi.org/10.18653/v1/w17-4418

[7] Hengyi Fu. 2017. Query Reformulation Patterns of Mixed Language Queries in Different Search Intents. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. 249–252.

[8] Jiafeng Guo, Gu Xu, Xueqi Cheng, and Hang Li. 2009. Named entity recognition in query. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. 267–274.

[9] Deepak Gupta, Shubham Tripathi, Asif Ekbal, and Pushpak Bhattacharyya. 2016. A hybrid approach for entity extraction in code-mixed social media data. *MONEY* 25 (2016), 66.

[10] Parth Gupta, Kalika Bali, Rafael E Banchs, Monojit Choudhury, and Paolo Rosso. 2014. Query expansion for mixed-script information retrieval. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. 677–686.

[11] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[12] Johannes Hoffart, Yasemin Altun, and Gerhard Weikum. 2014. Discovering emerging entities with ambiguous names. In *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014*, Chin-Wan Chung, Andrei Z. Broder, Kyuseok Shim, and Torsten Suel (Eds.). ACM, 385–396. https://doi.org/10.1145/2566486.2568003

[13] Saravanakumar Kandasamy and Aswani Kumar Cherukuri. 2020. Query expansion using named entity disambiguation for a question-answering system. *Concurrency and Computation: Practice and Experience* 32, 4 (2020), e5119.

[14] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williams College, Williamstown, MA, USA, June 28 - July 1, 2001*, Carla E. Brodley and Andrea Pohoreckyj Danyluk (Eds.). Morgan Kaufmann, 282–289.

[15] Zhen Liao, Xinying Song, Yelong Shen, Saekoo Lee, Jianfeng Gao, and Ciya Liao. 2017. Deep context modeling for web query entity disambiguation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1757–1765.

[16] Tianyu Liu, Jin-Ge Yao, and Chin-Yew Lin. 2019. Towards Improving Neural Named Entity Recognition with Gazetteers. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, Anna Korhonen, David R. Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, 5301–5307. https://doi.org/10.18653/v1/p19-1524

[17] Tao Meng, Anjie Fang, Oleg Rokhlenko, and Shervin Malmasi. 2021. GEMNET: Effective Gated Gazetteer Representations for Recognizing Complex Entities in Low-context Input. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

[18] Tu Ngoc Nguyen, Nattiya Kanhabua, and Wolfgang Nejdl. 2018. Multiple Models for Recommending Temporal Aspects of Entities. In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings (Lecture Notes in Computer Science)*, Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam (Eds.), Vol. 10843. Springer, 462–480. https://doi.org/10.1007/978-3-319-93417-4_30

[19] Ruba Priyadharshini, Bharathi Raja Chakravarthi, Mani Vegupatti, and John P McCrae. 2020. Named entity recognition for code-mixed Indian corpus using meta embedding. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 68–72.

[20] Royal Sequiera, Monojit Choudhury, Parth Gupta, Paolo Rosso, Shubham Kumar, Somnath Banerjee, Sudip Kumar Naskar, Sivaji Bandyopadhyay, Gokul Chittaranjan, Amitava Das, et al. 2015. Overview of FIRE-2015 Shared Task on Mixed Script Information Retrieval.. In *FIRE workshops*, Vol. 1587. 19–25.

[21] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc V. Le, Geoffrey E. Hinton, and Jeff Dean. 2017. Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net. https://openreview.net/forum?id=B1ckMDqlg