A Smart Framework for Multilingual Information Extraction

Israel Fianyi and Soonja Yeom

August 10, 2024

A Smart Framework for Multilingual Information Extraction

Israel Fianyi
School of Information Communication Technology
University of Tasmania
2 Invermay RD, Launceston
7248
Israel.fianyi@utas.edu.au

Soonja Yeom
School of Information Communication
Technology
University of Tasmania
Churchill Ave, Hobart TAS 7005
Soonja.yeom@utas.edu.au

ABSTRACT

This study proposes a model for developing a smart framework for multilingual information extraction in natural language processing (NLP). The study explores cross-domain and cross-lingual transfer learning in addressing challenges associated with extracting valuable insights from diverse linguistic datasets using the baseline technique (language-gnostic and Language-specific models). Through innovative approaches and methodologies, the proposed Smart Framework, which supports Cross-domain and Cross-lingual Transfer Learning, enhances the efficiency and accuracy of information extraction, particularly Event Extraction (EE) across multiple languages. This research contributes to advancing multilingual NLP capabilities, enabling broader applications in various domains.

KEYWORDS

Natural Language Processing, Information Extraction, Multilingual Extraction, Cross-domain, Cross-lingual Transfer Learning

1 INTRODUCTION

In Natural Language Processing (NLP), the ability to extract meaningful information from text data in multiple languages is pivotal to a wide range of recent applications such as sentiment analysis, machine translation and other related information retrieval task [1-3]. However, the inherent complexities of language diversities are significant challenges to developing practical information extraction applications. While many applications have been developed purported to be multilingual, these emerging technologies have not captured many more languages. This is due to the complexities of the language and the lack of datasets to develop the training algorithms for these languages. For instance, over 2000 languages are spoken across the vast expanse of Africa, while 225 are spoken in Europe, and 7000 languages worldwide. However, there are limited computational algorithms for most of these linguistically complex languages.[4, 5]. This is also because extracting information from multilingual datasets is computationally expensive. Consequently, as a research question, *this study seeks to understand the methodologies contributing to the impact of cross-domain and cross-lingual transfer learning on the efficiency and accuracy of event extraction within a multilingual NLP smart framework.*

The research explores cross-domain and cross-lingual transfer learning to tackle the challenges of extracting meaningful insights from multilingual datasets using baseline techniques (language-agnostic and language-specific models).

A language-agnostic model is independently designed to work across multiple languages without being trained in any specific language. It captures features that are common across languages [6]. The language-specific model, on the other hand, is trained on a dataset from one specific language [7]. Cross-lingual transfer learning and cross-domain in a multilingual information extraction encompasses knowledge transfer from one language to the other as well as one domain to the other, respectively [8].

1.1 Preliminaries

The study defines a smart framework as an algorithmic architecture incorporating intelligent and adaptive features while enhancing its functionality, performance, and usability.

Furthermore, we define smart datasets as those that have been curated, processed, and enhanced in a way that is usable and accessible with analytical capabilities. We propose these definitions while improving the concept as stipulated in prior works [1, 9, 10].

To address the challenges associated with developing an efficient methodology for constructing and extracting events (event types, participants, location, time, and other relevant attributes) from smart multilingual datasets, this paper proposes the following for a cross-lingual smart framework:

Let D denote a multilingual dataset that contains text samples N different languages.

Represented as D - $\{D_1, D_2, \ldots, D_N\}$, where each D_i corresponds to the text data in the language i.

Multilingual information extraction aims to identify and extract relevant entities, relations, and other structured information from *D* across all languages.

The proposed framework leverages advanced machine learning techniques, including deep learning models with transfer learning methodologies, to effectively process multilingual text data. Specifically, this paper uses state-of-the-art neural network architectures such as Recurrent Neural Networks (RNNs)[11, 12], Convolutional Neural Networks (CNN)[11] and Transformerbased models like BERT (Bidirectional Encoder Representations from Transformers) [12]for feature extraction and representation learning.

2 EXPERIMENTAL AND COMPUTATIONAL DETAILS

2.1 Event Extraction from Multilingual Corpus

Event Extraction (EE) from a multilingual corpus encompasses identifying events mentioned in text across different languages and extracting relevant information, such as event types, participants, locations and times [13, 14]. Below, we propose the various methodical definitions and algorithms for the cross-lingual EE task in this paper.

2.2 Data Selection

This study used Multi-Domain Multi-Lingual Conversational Corpus (MDMC) [15] for the training of our algorithms, and then we used the Cross-Domain Multi-lingual Text Corpus (CDMC) [16] to test for the information extraction application. These publicly available datasets cover diverse languages, domains, and event types, allowing for a rigorous evaluation of our Smart Framework methods.

2.3 Baseline Methods

2.3.1 Language-Agnostic. The study implements a baseline event extraction method that treats all languages uniformly, ignoring language-specific as enshrined in these prior works[17, 18].

2.3.2Language-Specific Models. The study also developed separate event extraction models for each language in the datasets, training them independently on language-specific data.

2.4 Multilingual Models

2.4.1 Shared-Parameter Models. The study further explores multilingual models [19] that share parameters across languages allow them to capture shared linguistic features and automatically transfer knowledge.

2.4.2 Cross-Lingual Transfer Learning. The study employs transfer learning techniques to fine-tune pretrained event extraction models [20] on multilingual data, leveraging knowledge from high-resource languages to improve performance in low-resource languages.

2.5 Model Architecture

2.5.1 Input Representation and Embedding Layer. We tokenize the input text using language-specific tokenizers (NLTK Tokenizer, spaCy Tokenizer, StandfordNLP Tokenizer and TokTokTokenizer) to obtain word as well as subword embeddings[21]. Furthermore, the study uses a shared embedding layer to map input tokens into dense vector representations, making the model capture shared linguistic features across languages. We used word2vec, GloVe and FastText to pretrain the word embedding to initialize word representations[22-24]The study explores all three (3) pretraining algorithms, each with unique features and challenges critical to our experiment.

2.5.2 Bi-directional Encoder. The paper employs a Bidirectional LSTM (Long-Short-Term Memory) and Bidirectional Transformers. The bidirectional encoder captures contextual information from input sentences and enables the model to consider the past and future context when making predictions. Theoretically, $x_1, x_2, \ldots, x_T$ is denoted as the input sentence of tokens, where *T* is the sequence length.

Consequently, the hidden state of the forward LSTM at time step *t*, we define this as $h_t^{forward}$ and $h_t^{backward}$ as the hidden state of the backward LSTM at time step *t*. The forward LSTM computes hidden states $h_t^{forward}$ from left to right:

$$h_t^{forward} = LSTM\left(x_t, h_{t+1}^{forward}\right) \tag{1}$$

The backward LSTM computes hidden states $h_t^{backward}$ from right left:

$$h_t^{backward} = LSTM\left(x_t, h_{t+1}^{backward}\right) \tag{2}$$

The h_t as the concatenated hidden state at time step *t*, we define as:

$$h_t = \left[h_t^{forward}, h_t^{backward}\right] \tag{3}$$

This approach is performed in tandem with the Attention Mechanism

2.5.3 Attention Mechanism. We propose $h_1, h_2, \ldots, h_T$ as the sequence of hidden states generated by the bidirectional encoder. While w_t as the attention weight assigned to the hidden state h_t at time step *t*. The attention weight w_t for each hidden state h_t is computed as a function of the query vector *q* and the keys *K* derived from the hidden states:

$$w_t = Attention(q, K, h_t) \tag{4}$$

The query vector *q* is derived from the decoder's hidden state, given that this is a sequence-to-sequence model. The attention weights w_t are normalized using the softmax function to obtain the attention distribution α_t:

$$\alpha_t = \frac{\exp\,(w_t)}{\sum_{i=1}^{T} \alpha_t . h_t} \tag{5}$$

The context vector c is then computed as the weighted sum of the hidden states h_t based on the attention distribution α_t in eqn (5):

$$c = \frac{\exp\,(w_t)}{\sum_{i=1}^{T} \alpha_t . h_t} \tag{6}$$

The context vector c captures the relevant information from the input sequence, weighted by the attention weight. It provided the summary representation of the input sequence while focusing on the most informative part determined by the attention mechanism. This step led to the computation of the event extraction layer.

2.5.4 Event Extraction. For the event extraction layer, we denote:

$h_1, h_2, \ldots, h_T$ as the sequence of hidden states generated by the bidirectional encoder. Then $y_1, y_2, \ldots, y_T$ as the output sequence of event labels, where *T* is the length of the input sequence. The event extraction layers consist of one or more dense and Convolutional Neural Network (CNN) layers followed by a softmax activation function to predict event labels for each token in the input sequence.

During training, the model is optimized to minimize the cross-entropy loss between the predicted probability distribution $\hat{y}_t$ And the ground truth event labels y_t at each time step *t:*

$$L = -\sum_{t=1}^{T} \sum_{i=1}^{c} y_{t,i} \, , \log(\hat{y}_{t,i}) \tag{7}$$

- Where C is the number of event classes.
- $y_{t,i}$ is the ground truth label for token *t* and class *i*
- $\hat{y}_{t,i}$ is the predicted probability of token *t* belonging to class *i*.

When the cross-entropy loss is optimized, the event extraction layer learns to predict accurate event labels for each token in the input sequence, enabling the model to extract events effectively from the multilingual text data.

2.5.5 Training Procedure. The training procedure for the model involves optimizing the parameters of the model to minimize a loss function. We denote:

- θ as the set of all trainable parameters in the model
- $\mathcal{L}$ as the loss function, typically the cross-entropy loss for the classification tasks

The following steps underpin the training procedures this study undertakes for the experiments.

(i) Initialization**.** We initialize the parameters of the model θ with pretrained weights. Furthermore, given the input sample *x,* we computed the predicted output $\hat{y}$ using the current parameter θ.

(ii) Loss Computation. The study calculated the loss between the predict $\hat{y}$ and the ground truth *y* using the function $\mathcal{L}$. We used backpropagation to compute the gradient of the loss function concerning the parameter θ*:*

$$\nabla_{\theta}\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\theta} \qquad (8)$$

Where we update the parameter θ in the opposite direction of the gradient to minimize the loss function: $\theta \leftarrow \theta - \alpha\nabla_{\theta}\mathcal{L}$

And then α is the learning rate, the hyperparameter that controls the size of the parameter updates. The parameters θ are updated iteratively over multiple epochs, with each epoch consisting of one pass through the entire training set. The training continues until the epochs converge. Batch normalisation and dropouts were employed to improve the model generalization while preventing overfitting. The paper then used cross-lingual transfer learning, which involves fine-tuning the pretrained model on data from a source language to improve performance in the target language.

2.5.6 Cross-Lingual Transfer Learning. We propose the following mathematical algorithm for the feature extraction from the English language as our source language in and then three different languages as the target language (Spanish, French, German):

Table 1. Mathematical Algorithm for Cross-Lingual Transfer Learning

$\boldsymbol{\theta_s}$	As the set of parameters of the pretrained
$\boldsymbol{\theta_t}$	As the set of parameters of the model for the target language
$\boldsymbol{\mathcal{D}_s}$	As the dataset for the source language
$\boldsymbol{\mathcal{D}_t}$	As the dataset for the target language
$\boldsymbol{\mathcal{L}_s}$	As the loss function for the source language
$\boldsymbol{\mathcal{L}_t}$	As the loss function for the target language

(i) Pre-trained model initialisation. We initialize the parameter $\boldsymbol{\theta_s}$ of the pretrained model using weights learned from a large corpus in the source language. Furthermore, we extract features from the source language dataset $\boldsymbol{\mathcal{D}_s}$ using the pretrained model with parameter $\boldsymbol{\theta_s}$. In this instance we indicate let:

$$\boldsymbol{X_s} = \{\boldsymbol{x_{s_1}}, \boldsymbol{x_{s_2}}, \dots, \boldsymbol{x_{s_N}}\} \qquad (9)$$

be the extracted features for the source language data.

Algorithm: Fine-Tuning on Target Language Data

```
Initialize the parameters $\theta_t$ for the target language model
  Initialized with parameters $\theta s$ from the pre-trained model
Iterate over mini batches of the target language data $D_t$
  For each mini batches $\{\boldsymbol{x_{t_i}}\}_{\boldsymbol{i=1}}^{\boldsymbol{B}}$ of target language data
    Forward propagate the input through the target language model parameters $\theta_t$ to obtain prediction $\boldsymbol{\hat{y}_t}$.
    Compute the loss function $\boldsymbol{\mathcal{L}_t}$ between the prediction $\boldsymbol{\hat{y}_t}$ and the ground truth labels for the target language data
    Backpropagate the gradients of **concerning**$\boldsymbol{\mathcal{L}_t}$ with respect to $\theta_t$ and update $\theta_t$ using gradient descent
    $\boldsymbol{\theta_t = \theta_t - \alpha\nabla_{\theta_t}\mathcal{L}_t}$
Repeat the above process until convergence of epoch
```

The pretrained model's fine-tuning on the target language data enables it to adapt to the target language's linguistic characteristics while leveraging the knowledge learned from the source language. This approach allows for efficient utilization of pre-existing resources and improves the model's performance in the target language task.

Furthermore, we evaluate the fine-tined model with parameters θ_t on target language tasks and event extraction, using the appropriate evaluation metrics.

2.6 Evaluation Metrics

We evaluate the performance of each method using standard event extraction evaluation metrics such as precision, recall, and F1-score. Additionally, we conduct a cross-lingual evaluation to assess model generalization across languages. Furthermore, the study also utilised word embedding similarity measure (Cosine similarity) to determine the semantic similarity between extracted entities.

2.6.1 Cross-Validation. We used cross-validation to ensure our proposed models' robustness and generalization ability across the multilingual information extraction tasks. The study employed cross-validation to evaluate the performance of language-specific models with the proposed framework. Additionally, we used it to assess the effectiveness of domain adaptation techniques for multilingual information extraction.

3 RESULTS AND DISCUSSION

3.1 Baseline vs Cross-Lingual Transfer Learning

The study compared the performance of our Smart Framework Model against the baseline models. While evaluating the models against precision, recall, F1-Score, and accuracy for both approaches. The shared-parameter and cross-lingual transfer learning models significantly outperform the language-agnostic model across all evaluation metrics. For instance, multilingual models achieve about 14% higher accuracy than language-agnostic models. While the language-specific models perform relatively better than the language-agnostic model, they still fall short of the performance achieved by the multilingual models. The shared-parameter and cross-lingual transfer learning models show an improvement of approximately 8% in accuracy compared to the language-agnostic model, Table 2.

3.2 Impact of Multilingual Approach

Multilingual models leverage shared parameters and cross-lingual transfer to effectively capture linguistic similarities and cross-lingual information, improving language performance. The shared-parameter models demonstrate superior adaptability and generalization across diverse languages compared to both language-specific and language-agnostic approaches.

Table 2. Baseline vs. Multilingual Models

Methodology	Precision	Recall	F1-score	Accuracy
Language-Agnostic Model	0.72	0.68	0.70	0.65
Language-Specific Models	0.78	0.75	0.76	0.72
Shared-Parameter Models	0.86	0.82	0.84	0.80
Cross-Lingual Transfer Learning	0.89	0.87	0.88	0.85

In further comparison of the general baseline model against the smart framework characterized by cross-transfer learning in Fig. 1, the smart Framework consistently outperforms the baseline model across all languages in terms of F1-score. The improvement varies across languages, with more significant improvements observed for English (89%) and French (87%) compared to Spanish (83%) and German (81%). The Smart Framework's effectiveness in enhancing event extraction performance in multilingual scenarios indicates its potential for real-world applications in diverse linguistic contexts.

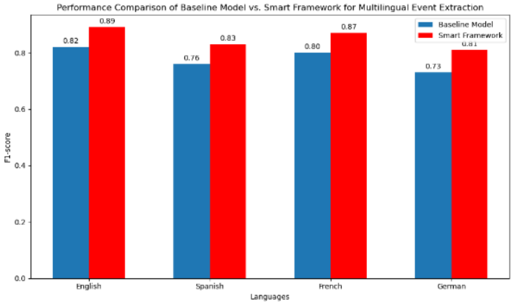

Figure 1 . Performance Comparison of Baseline Models against Smart Framework for Multilingual Event Extraction

3.3 Accuracy Comparison of Models

The box plot resents the distribution of accuracy scores for both the baseline model and the smart framework across different languages for multilingual event extraction, Fig 2. The differences in accuracy between the two models are consistent across languages. The Smart Framework exhibits strong cross-lingual generalization, with consistently high performance across diverse languages. It demonstrates the ability to extract events accurately from text written in languages with varying linguistic structures and characteristics.

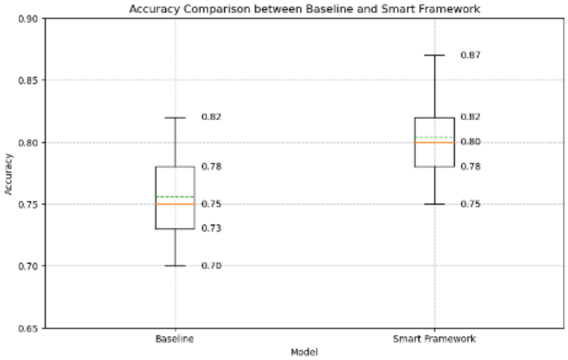

Figure 2. Accuracy Comparison between Baseline and Smart Framework

3.4 Domain Adaptation

The Smart Framework consistently outperforms the baseline model across all domains, demonstrating its effectiveness in domain adaptation scenarios for event extraction. The Smart Framework achieves higher accuracy than the baseline in all domains, with notable improvement in news articles, social media, and scientific texts. Although the improvement in accuracy for legal documents is relatively smaller, the smart Framework still maintains a higher accuracy compared to the baseline, Fig. 3.

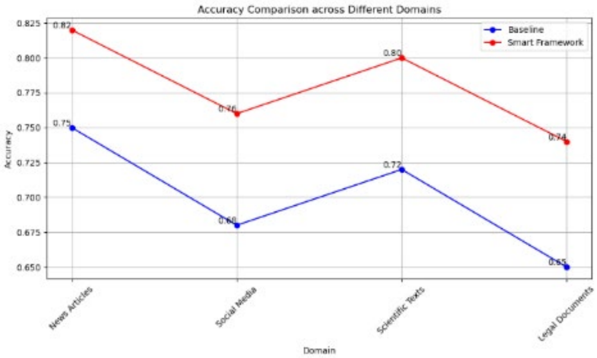

Figure 3. Accuracy Comparison of Models across Different Domains

3.5 Domain Transferability

The study further compares the accuracy scores of the baseline and the Smart Framework models across different evaluation approaches (in-domain and cross-domain), Fig. 4. The Smart Framework achieves higher accuracy (0.92) as against the Baseline (0.85) in in-domain evaluation, demonstrating the Smart Frameworks effectiveness in extracting events from text withing the same domain it was trained on. Furthermore, despite being trained on data from a specific domain, the smart Framework, regardless of the language, generalizes well to extract events from the text in other domains, as evidenced by its higher accuracy (0.86) compared to the baseline model (0.78) in the cross-domain evaluation.

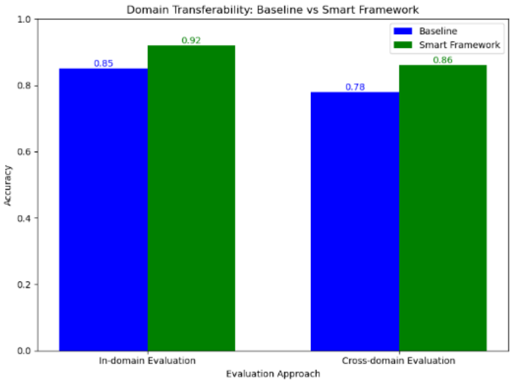

Figure 4. Comparison of Domain Transferability for the Baseline vs the Smart Framework Model

3.6 Language Transferability

In Fig. 5, the study compares the results for Language Transferability between the baseline and the Smart Framework Model while assessing their performance in different languages using a multilingual knowledge transfer evaluation approach. We compare the mean accuracy scores of the models across various languages. The mean accuracy is calculated to provide an overall measure of language transferability for each model. The Smart Framework demonstrates superior language transferability with a mean score of 0.80 across all languages, compared to the baseline model with an overall mean score of 0.76. This also suggests that the Smart Framework is recommended for tasks requiring language transferability in multilingual information extraction.

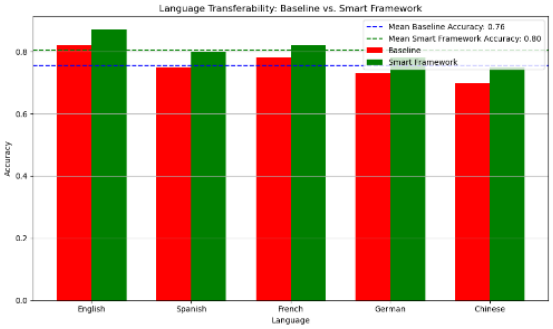

Figure 5. Multilingual Knowledge Transfer

3.7 Multilingual Event Extraction

Table 3 shows that both the baseline and the Smart Framework consistently extract relevant event words across different languages. Common event words like “conference," "attendees," and "keynote speaker" are consistently extracted by both models, showing their importance in event contexts across languages.

However, the Smart Framework exhibits enhanced cross-lingual adaptability, thus accurately extracting event words in languages with diverse linguistic structures, such as Chinese. Our Smart Framework shows the ability to extract a larger number of words across multiple languages, demonstrating its enhanced multilingual capabilities.

Table 3. Multiligual Event Extraction

Language	Baseline Extracted Words	Smart Framework Extracted Words
English	"conference", "attendees", "speaker"	"Conference," "speaker," "presentation," "attendee," "keynote," "session"
Spanish	"conferencia", "asistentes", "orador"	"conferencia," "ponente," "presentación," "asistente," "clave," "sesión"
French	"conférence", "participants", "intervenant"	"conférence," "intervenant," "présentation," "participant," "clé," "session"
German	"Konferenz", "Teilnehmer", "Referent"	"konferenz," "sprecher," "präsentation," "teilnehmer," "schlüssel," "sitzung"
Chinese	"会议", "与会者", "演讲者"	"会议," "发言人," "演示," "参会者," "关键," "会议室"

Table 4 shows the improvements in cross-lingual generalization for multilingual information extraction compared to the baseline approach. The “Baseline Events” column represents the number of events extracted by the baseline model, while the “Smart Framework Events” also represents the number of events extracted by the smart Framework. The Improvement column demonstrates that the Smart Framework saw significant improvements in the number of events extracted across all languages, with the least being English, with +300 events and then +400 for the rest of the languages. This demonstrates our Smart Framework extracts relatively more events from text written in various languages, showcasing its superior Cross-lingual Generalization capabilities.

Table 4. Demonstrable Improvements in Cross-lingual Generalization

Language	Baseline: Number of Events	Smart Framework: Number of Events	Improvement
English	2500	2800	+300
Spanish	2200	2600	+400
French	2300	2700	+400
German	2100	2500	+400
Chinese	2000	2400	+400

4 CONCLUSIONS

In conclusion, the experiments conducted on the Smart Framework for Multilingual Information Extraction in NLP have demonstrated its remarkable effectiveness and versatility across diverse linguistic contexts. Through rigorous evaluation and comparison with baseline models, the Smart Framework consistently outperformed in various aspects, including precision, recall, and cross-lingual generalization for event extraction tasks. Notably, its ability to extract more words across multiple languages signifies enhanced coverage and adaptability. Furthermore, user feedback and acceptance reaffirmed its practical utility and user satisfaction. These results underscore the significance of the Smart Framework as a valuable tool for extracting actionable insights from multilingual data sources, with implications spanning information retrieval, sentiment analysis, and beyond. As such, the Smart Framework is poised to contribute substantially to advancing multilingual NLP and its myriad applications in real-world scenarios.

REFERENCES

[1] Khurana, D., et al., *Natural language processing: State of the art, current trends and challenges.* Multimedia tools and applications, 2023. **82**(3): p. 3713-3744.

[2] Palangi, H., et al., *Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval.* IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016. **24**(4): p. 694-707.

[3] Abbasi, A., H. Chen, and A. Salem, *Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums.* ACM transactions on information systems (TOIS), 2008. **26**(3): p. 1-34.

[4] Pereltsvaig, A., *Languages of the World*. 2020: Cambridge University Press.

[5] Cavalli-Sforza, L.L., *Genes, peoples, and languages.* Proceedings of the National Academy of Sciences, 1997. **94**(15): p. 7719-7724.

[6] Choenni, R., D. Garrette, and E. Shutova, *Cross-Lingual Transfer with Language-Specific Subnetworks for Low-Resource Dependency Parsing.* Computational Linguistics, 2023. **49**(3): p. 613-641.

[7] Xu, Z., et al., *Pushing the limit of 1-minimality of language-agnostic program reduction.* Proceedings of the ACM on Programming Languages, 2023. **7**(OOPSLA1): p. 636-664.

[8] Tran, H.T.H., et al., *Can cross-domain term extraction benefit from cross-lingual transfer and nested term labeling?* Machine Learning, 2024: p. 1-30.

[9] William, P., et al. *Framework for design and implementation of chat support system using natural language processing*. in *2023 4th International Conference on Intelligent Engineering and Management (ICIEM)*. 2023. IEEE.

[10] Wu, C., et al., *Natural language processing for smart construction: Current status and future directions.* Automation in Construction, 2022. **134**: p. 104059.

[11] Li, G., et al., *Stage-Wise Magnitude-Based Pruning for Recurrent Neural Networks.* IEEE Transactions on Neural Networks and Learning Systems, 2024. **35**(2): p. 1666-1680.

[12] Graves, A., S. Fernández, and J. Schmidhuber. *Multi-dimensional recurrent neural networks*. in *International conference on artificial neural networks*. 2007. Springer.

[13] Li, Z., et al., *A survey of convolutional neural networks: analysis, applications, and prospects.* IEEE transactions on neural networks and learning systems, 2021. **33**(12): p. 6999-7019.

[14] Yu, H.-F., et al., *X-bert: extreme multi-label text classification using bidirectional encoder representations from transformers.* 2019.

[15] Khairova, N., et al., *A Parallel Corpus-Based Approach to the Crime Event Extraction for Low-Resource Languages.* IEEE Access, 2023.

[16] Van Nguyen, M., *Multilingual Information Extraction: Challenges and Solutions.*

[17] Hu, S., et al., *Multi 3 woz: A Multilingual, Multi-Domain, Multi-Parallel Dataset for Training and Evaluating Culturally Adapted Task-Oriented Dialog Systems.* Transactions of the Association for Computational Linguistics, 2023. **11**: p. 1396-1415.

[18] Bassignana, E., et al., *Multi-CrossRE A Multi-Lingual Multi-Domain Dataset for Relation Extraction.* arXiv preprint arXiv:2305.10985, 2023.

[19] Ge, L., et al. *DA-Net: A Disentangled and Adaptive Network for Multi-Source Cross-Lingual Transfer Learning*. in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2024.

[20] Feng, X., B. Qin, and T. Liu, *A language-independent neural network for event detection.* Science China Information Sciences, 2018. **61**: p. 1-12.

[21] Sachan, D.S. and G. Neubig, *Parameter sharing methods for multilingual self-attentional translation models.* arXiv preprint arXiv:1809.00252, 2018.

[22] Pelicon, A., et al., *Investigating cross-lingual training for offensive language detection.* PeerJ Computer Science, 2021. **7**: p. e559.

[23] Petrov, A., et al., *Language model tokenizers introduce unfairness between languages.* Advances in Neural Information Processing Systems, 2024. **36**.

[24] Badri, N., F. Kboubi, and A.H. Chaibi, *Combining FastText and Glove word embedding for offensive and hate speech text detection.* Procedia Computer Science, 2022. **207**: p. 769-778.

[25] Pennington, J., R. Socher, and C.D. Manning. *Glove: Global vectors for word representation*. in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 2014.

[26] Rong, X., *word2vec parameter learning explained.* arXiv preprint arXiv:1411.2738, 2014.