Integrating Arabic Open Information Extraction Model by Machine Learning to Extract Relation

Sally M.A. Elmorsy
Lecturer at the Higher Institute of Management
Mahala Elkoubra

Mohamed Bahh Eldin Abdellatief
10Th grade school

ABSTRACT
In this paper, The Arabic text presents various difficulties, as text mining in Arabic is challenging due to its inflectional solid and derivational nature. The applications of Arabic information extraction are still in their early stages and require an improved approach to take advantage of its potential. Information extraction requires a general method that can work on the text regardless of its domain (e.g., biomedical, sport, economics, etc.) and capture all its information. Machine learning techniques have been applied to make information extraction systems more portable. Machine learning aims to develop algorithms to assist or replace domain experts in knowledge engineering situations. By using learning algorithms that rely on open information extraction to automate information retrieval processes such as document classification, the modeling can reduce the workload of information workers and minimize the inconsistencies introduced by human error, focusing on the research domain and problem.

Keywords
Arabic Open Information, dependency parsing, Decision Tree

1. INTRODUCTION
Information extraction involves applying natural language processing to extract essential details from text documents automatically[1]. A significant disadvantage of current approaches is their intrinsic dependence on the application domain and the target language. Several machine-learning techniques have been applied to facilitate information extraction systems' portability [2]. The research domain uses natural language processing to extract meaningful information from text documents automatically. The main challenge in this area[1] is the dependence of current information extraction systems on the application domain and target language, which limits their portability and applicability. To address this problem, previous research has explored the use of machine learning techniques for information extraction. However, these approaches are still limited by their intrinsic dependence on the application domain and target language[3].

This paper introduces a relation extraction model using the Decision Tree machine learning technique[4]. The proposed model uses the labeled data outputted from the Open Information Extraction model, categorizing every word in a related part as one of the following: target (Verb, Subject, Object, Adjective, I, E, or not any of them as a supplement). This label is considered the target from the decision tree model while the predictors have been extracted from the (CONLL-U) formate features[5]. The features used were POSTAG (Part Of Speache Tag), XPOSTAG (language-specific part-of-speech tag), and DEPREL (Dependency Relation). The feature has been chosen as the most related feature with the words' position in sentences and its relation label type[6]. This method finds the extraction of relationships to be an activity of classification among the most used supervised methods[7]. The conducted decision tree model could facilitate using the proposed relation extraction model in large-scale text, increasing on the web.The model[8] uses the CONLL-U format feature to integrate with the result from the rule-based model proposed in the previous section. The text of the trained model reveals high performance and accuracy in the text from different fields[9][10]. The proposed approach in this paper aims to overcome these limitations by using a decision tree machine-learning technique for relation extraction. The model takes the output from the Open Information Extraction (OIE) model[5], which categorizes each word in a related part, and uses this output as the target for the decision tree. The predictors used in the model are extracted from the CONLL-U format features, which include POSTAG (Part Of Speech Tag), XPOSTAG (language-specific part-of-speech tag), and DEPREL (Dependency Relation). The decision tree is a popular approach for representing classifiers[11]. It is considered one of the most popular data-mining techniques for knowledge discovery[12]. The decision tree systematically derives useful rules and relationships from the information stored in a broad data source[13]. These extracted rules are typically used for classification/prediction purposes. Compared to other text mining methods, it is commonly used in diverse fields as it is versatile for data sizes or distributions [14]. In this chapter, it exploited the various features in the CONLLU format and combined them with the result obtained from the rule-based model to get a relation extraction easily and effectively. A decision tree technique has been used in training using the produced dataset from the finished OIE model to improve the proposed model[4]. The most related three CONLLU[11] features with the relation have been used as predictors, which are (POSTAG, XPOSTAG, DEPREL) and the word relation type as a target (Verb, Subject, Object, Adjective, or not any of them as a supplement). In this stage machine learning package in Python has been used, data for training have been collected in an Excel sheet, and an encoder has been used to convert the data to vectors[10]. After that, the model was trained in a part of the dataset and tested in the rest. The final results of the model reveal high performance and accuracy. The decision tree is trained on a dataset collected in an Excel sheet and then converted to vectors using an encoder[15]. The model was tested on a portion of the dataset, and the results showed high performance and accuracy in large Arabic data[6]. The proposed approach provides a more effective and efficient way to extract relationships from text.

2. LITERATURE REVIEW
Some research studies have been directed towards Machine Learning (ML) approaches, including unsupervised, semi-supervised, and supervised learning techniques, to automate the relationship extraction task completely. Unsupervised approaches use large volumes of unlabeled text and are based almost exclusively on clustering strategies and comparisons between features or meaning terms[16]. Hasegawa et al., 2004 focused on clustering NE pairs based on the similarities of

meaning terms that function between NEs. These writers have not considered the meaning of terms before and after the NEs. However, these two contexts can provide valuable information for exploring semantic relationships between NEs. In addition, the writers assume relations whose contexts are the same [17].

Zhang et al., Calculated the similarities of two parsing trees to be clustered using a hierarchical clustering model. Each cluster obtained is classified, and any bad clusters whose NE pair number is below the predetermined threshold are discarded. Any experiments have been geared towards semi-supervised learning techniques or bootstrapping strategies to resolve challenges with an unsupervised approach [18]. The last approach to ML techniques is the supervised form, which focuses on a completely labeled corpus[6]. This method finds the extraction of relationships to be an activity of classification. Among the most used supervised methods are Support Vector Machines (SVM)[19], Conditional Random Fields (CRF), Decision Trees, and Maximum Entropy (MaxEnt). A recent attempt to extract ties between the Arab NEs has been made[20], Which used a classifier dependent on MaxEnt. Based solely on morphological and part-of-speech (POS) details, this method achieves satisfactory results when applied to the ACE4 corpus.

Celli [21] combined two supervised strategies, i.e., simple decision tree and PART decision list algorithms, to extract three semantic relationships (role, social, and location) between NEs. These authors relied only on the context POS before and between the two entities without considering the context after the NEs. Kramdi et al., 2009 adopted the learning pattern algorithm LP2 that Ciravegna and Wilks, 2003 proposed to generate annotation rules. They obtained an F-score of 50%[22]. The resulting patterns produced by such a method often suffer from low precision. Another study that adopted the learning rules method was performed by Boujelben et al., 2013 [23]. The research seeks to use the association rule algorithm Apriori. This mining rule model aims at finding all the rules from a database that satisfy minimum support and minimum confidence values. The main advantage of supervised relation NE systems is that they can be applied to other domains. Additionally, their update is conducted with minimal time and effort in cases where a sufficient database is available [24].

A decision tree model is a popular approach for representing classifiers. It is considered one of the most popular data-mining techniques for knowledge discovery. The decision tree systematically derives from the information stored in a comprehensive data helpful source rules and relationships. These extracted rules are typically used for classification/prediction purposes. Compared to other text mining methods, it is commonly used in diverse fields as it is versatile for data sizes or distributions [14].

3. INTEGRATING THE DECISION TREE MODEL WITH A RULE-BASED MODEL

In order to obtain more effective performance and results in the relation extraction test, the decision tree model has been used as an additional step[7][14], as shown in figure1. The proposed model uses the label data, which is output from the previous model, which categorizes every word in a related part as one of the following: target (Verb, Subject, Object, Adjective, I, E or not any of them as a supplement). This label is considered the target from the decision tree model, while the predictors have been extracted from the COLLU formate features.

3.1 The Methodology and Experimental

The features used were (POSTAG, XPOSTAG, DEPREL). The selected feature has been chosen as the most related feature with the position of the words in sentences and its relation label type.

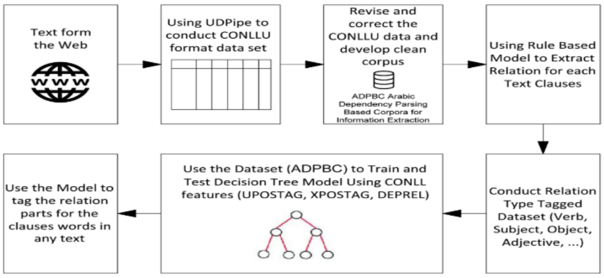

Fig 1: The framework of Creating a structured approach to combine a decision tree model and a rule-based model

To build the proposed model, the Python language and the "sklearn" package have been used to apply the machine learning technique. Following is the code which is used in the decision tree model processes. Before the training step, the data were collected in an Excel sheet and encoded using the "panda" package in Python. In order to conduct the training step, the output from rule-based has been used. Table 1 presents a sentence example has been presented to explain the model steps and the CONLLU file analysis for it.

Table 0. Text example with the CONLLU file

<table>
<tr><td colspan="8">" تعتبر فيروسات كورونا، سلالة واسعة من الفيروسات التي قد تسبب المرض للحيوان والإنسان. ومن المعروف أن عدداً من فيروسات كورونا تسبب أمراض تنفسية ،تتراوح قوتها من نزلات البرد الشائعة إلى الأمراض الأشد وخامة مثل متلازمة الشرق الأوسط التنفسية (ميرس) والمتلازمة التنفسية الحادة الوخيمة (سارس). و يسبب فيروس كورونا، المُكتشف مؤخراً مرض كوفيد-19"</td></tr>
<tr><th>Id</th><th>Form</th><th>Lemma</th><th>UPosTag</th><th>XPosTag</th><th>Feats</th><th>Head</th><th>DepRel</th></tr>
<tr><td colspan="6"># newdoc</td><td></td><td></td></tr>
<tr><td colspan="6"># newpar</td><td></td><td></td></tr>
<tr><td colspan="6"># sent_id = 1</td><td></td><td></td></tr>
<tr><td colspan="8"># text = تعتبر فيروسات كورونا ،سلالة واسعة من الفيروسات التي قد تسبب المرض للحيوان والإنسان</td></tr>
<tr><td>1</td><td>تعتبر</td><td>إعتَبَر</td><td>VERB</td><td>VIIP-3FS--</td><td>Aspect=Imp|Gender=Fem|Mood=Ind|Number=Sing|Person=3|VerbForm=Fin|Voice=Pass</td><td>0</td><td>root</td></tr>
<tr><td>2</td><td>فيروسات</td><td>فيروسات</td><td>NOUN</td><td>N------P1R</td><td>Case=Nom|Definite=Cons|Number=Plur</td><td>1</td><td>nsubj</td></tr>
<tr><td>3</td><td>كورونا</td><td>كورونا</td><td>X</td><td>U---------</td><td>_</td><td>2</td><td>nmod</td></tr>
<tr><td>4</td><td>،</td><td>،</td><td>PUNCT</td><td>G---------</td><td>_</td><td>3</td><td>punct</td></tr>
</table>

5	سلالة	سلالة	NOUN	N------S1I	Case=Nom|Definite=Ind|Number=Sing	1	obj
6	واسعة	وَاسِع	ADJ	A-----FS1I	Case=Nom|Definite=Ind|Gender=Fem|Number=Sing	5	amod
7	من	مِن	ADP	P---------	AdpType=Prep	8	case
8	الفيروسات	فَيرُوس	NOUN	N------P2D	Case=Gen|Definite=Def|Number=Plur	5	nmod
9	التي	اَلَّذِي	DET	SR----FS2-	Case=Gen|Gender=Fem|Number=Sing|PronType=Rel	11	nsubj
10	قد	قَد	PART	F---------	_	11	advmod:emph
11	تسبب	سَبَّب	VERB	VIIA-3FS--	Aspect=Imp|Gender=Fem|Mood=Ind|Number=Sing|Person=3|VerbForm=Fin|Voice=Act	8	acl
12	المرض	مَرَض	NOUN	N------S4D	Case=Acc|Definite=Def|Number=Sing	11	obj
13	للحيوان	للحيوان	X	X---------	Foreign=Yes	14	Nmod
14	والإنسان	والانسان	X	U---------	_	12	Nmod
15	.	.	PUNCT	G---------	_	1	Punct

The extracted relation from this example using the AOIE to extract the relation that depends on the rule-based model shown in Table 2

Table 2. extract the relation by AOIE

Verb(V)	Subject(S)	Object(O)	Adjective (A)	Relation Type
تعتبر	فيروسات كورونا	سلالة	من الفيروسات	**VSOA**
تسبب	لفيروسات التي	لمرض		**SVO**
تسبب	عدداً من فيروسات كورونا	أمراض تنفسية		**VSO**
تتراوح	قوة	نزلات	إلى الامراض الاشد وخامه	**VSOA**
يسبب	فيروس	مرض كوفيد-19		**VSO**

The training dataset was collected using the three features selected from the CONLL-U format and the relation tag for each word. The three selected features from the CONLL-U format were POSTAG, XPOSTAG, and DEPREL. The relation tag for each word was obtained from a rule-based model. The purpose of the training dataset was to train a machine learning model, specifically a decision tree, to perform relation extraction. The selected features from the CONLL-U [25]format were considered to be relevant to the position of the words in sentences and their relation type. Each word's relation tag was considered the target for the decision tree model, while the features extracted from the CONLL-U format were used as predictors. The process of collecting the training dataset involved encoding the data into vectors using an encoder and storing the data in an Excel sheet. The training dataset was then used to train a portion of the model, with the remaining portion used for testing. The final results of the model showed a high performance and accuracy, as reported in the text

Table 3. Training data set

Form	POSTAG	XPOSTAG	DEPREL	Relation Tag Rule-Based
تسبب	VERB	VIIA-3FS--	acl	Verb
المرض	NOUN	N------S4D	obj	Obj
للحيوان	NOUN	X---------	nmod	Sup
والإنسان	NOUN	U---------	nmod	Sup
التي	DET	SR----FS2-	nsubj	Subject
تعتبر	VERB	VIIP-3FS--	root	Verb
فيروسات	NOUN	N------P1R	nsubj:pass	Subject
كورونا	NOUN	U---------	nmod	Sup
،	PUNCT	G---------	punct	Sup
سلالة	NOUN	N------S1I	obj	Obj
واسعة	ADJ	A-----FS1I	amod	Sup
من	ADP	P---------	case	Adj
الفيروسات	NOUN	N------P2D	nmod	Sup
التي	DET	SR----FS2-	nsubj	Sup
قد	PART	F---------	advmod:emph	Sup
تتراوح	VERB	VIIA-3FS--	acl	Verb
قوة	NOUN	N------S1R	nsubj	Subject
ها	PRON	SP---3FS2-	nmod	Sup
من	ADP	P---------	case	Sup
نزلات	NOUN	N------P2R	obl:arg	Obj
البرد	NOUN	N------S2D	nmod	Sup
الشائعة	ADJ	A-----FS2R	obl:arg	Sup
الى	ADP	N------P2D	case	Adj
الامراض	NOUN	N------P2R	nmod	Sup
الاشد	NOUN	N------S2D	nmod	Sup
وخامه	NOUN	A-----FS2R	nmod	Sup
يسبب	VERB	VIIA-3MS--	parataxis	Verb

فيروس	NOUN	U---------	nsubj	Subject
كورونا	NOUN	U---------	nmod	Sup
،	PUNCT	G---------	punct	Sup
المُكتشف	NOUN	N------S4D	obj	Sup
مؤخراً	ADJ	A-----MS4I	amod	Sup
مرض	NOUN	N------S2R	nmod	Obj
كوفيد	NOUN	X---------	Sup	Sup

This training data in Table 3 will consist of the features selected from the CONLL-U format and the relation tag for each word, which was obtained from the rule-based model. The features selected from the CONLL-U format were POSTAG, XPOSTAG, and DEPREL, and they were considered relevant to the position of the words in sentences and their relation type. The relation tag for each word was used as the decision tree model's target, while the CONLL-U format features were used as predictors. The training data was processed by encoding it into vectors using an encoder and storing it in an Excel sheet. This data was then used to train a portion of the model, with the remaining portion used for testing. The final results of the model showed high performance and accuracy, as reported in the text. An example of the data to use in the training step would involve selecting a set of sentences and annotating each word in the sentences with its relation tag, along with the POSTAG, XPOSTAG, and DEPREL features from the CONLL-U format. This annotated data would then be used to train the decision tree model. As illustrated in Table 4and Figure 2

Table 4. The normalized training data set

POS TAG	XPO STA G	DEP REL	Relati on Tag	POS TAG	XPO STA G	DEP REL	Rel atio n Tag
1	53	8	6	0	7	4	5
7	46	31	7	4	25	9	5
12	108	31	1	0	7	11	5
0	22	26	3	11	28	30	5
1	50	8	4	4	25	9	5
7	35	22	5	4	25	20	5
7	37	22	5	5	66	15	5
5	68	15	5	7	38	11	5
7	44	22	5	0	5	4	5
12	94	0	1	1	50	8	4
0	7	26	3	0	4	4	5

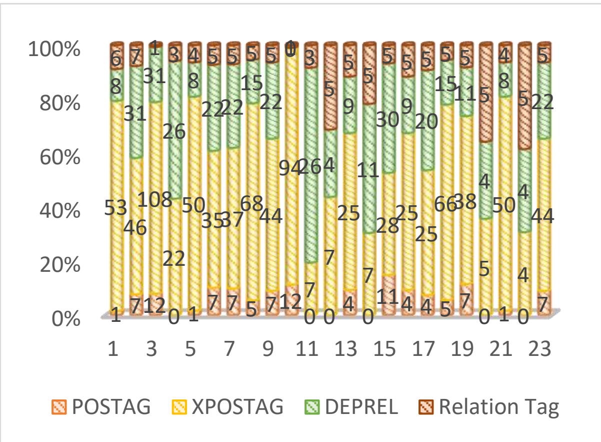

Fig 2. The percentage of training data

3.2 Result and Evaluation

referring to the process of training a machine learning model. It is stated that 70% of the collected data has been utilized for the training step. The remaining 30% of the data will likely be used for validation or testing.

Additionally, it is mentioned that the training data contains 2810 words, which are divided into different relation types. The relation types likely refer to the categories or labels into which the data is organized. For example, if the data is related to text classification, the relation types could be different categories like "sports," "politics," "entertainment," etc. The division of data into different relation types is important for training machine learning models, as it helps the model learn and make predictions based on the relationships between different categories of data and detect the different types of sentences. The noun phrase and the verb phrase, and the number of percentages are illustrated in the following table

Table 5. Training data set characterization

Word Relation Type	Number	Word Relation Type	Number
Verbs	235	Adjective	181
Subject	168	I	45
Object	162	E	44
Supplement Word		1763	
Total		2598	

The training has been implemented using a max depth with 20 layers while different depths have been tested but didn't give satisfactory results. Table4 shows a part of the decision tree illustration after the training step. Also, Table 5 shows the importance of UPOSTAG, XPOSTAG, and DEPREL, which ensure that the chosen features significantly impact the relation extraction step.

Table 6. Traning features importance

Feature	Importance
UPOSTAG	0.20089015
XPOSTAG	0.50512599
DEPREL	0.29398386

For the test step, the rest of the data representing 30% of the dataset, has been used to predict the relation type of each word. The number of words in the test set was 1118 words. Table 7 shows the output for the decision tree model and the previous model for the example sentence.

Table7. The output for the decission tree model and the previous model

Form	Relation Tag Rule-Based	Relation Tag DT	Form	Relation Tag Rule-Based	Relation Tag DT
تسبب	Verb	Verb	من	Sup	Adj
المرض	Obj	Obj	نزلات	Obj	Adj
للحيوان	Sup	Sup	البرد	Sup	Sup
والإنسان		Sup	الشائعة	Sup	Sup
التي	Subject	Subject	الى	Adj	Sup
تعتبر	Verb	Verb	الامراض	Sup	Sup

فيروسات	Subject	Sup	الاشد	Sup	Sup
كورونا	Sup	Sup	وخامه	Sup	Sup
،	Sup	Sup	يسبب	Verb	Verb
سلالة	Obj	Obj	فيروس	Subject	Subject
واسعة	Sup	Sup	كورونا	Sup	Sup
من	Adj	Adj	،	Sup	Sup
الفيروسات	Sup	Sup	المُكتشف	Sup	Adj
التي	Sup	Subject	مؤخراً	Sup	Sup
قد	Sup	Sup	مرض	Obj	Sup
تتراوح	Verb	Verb	كوفيد	Sup	Sup
قوة	Subject	Obj			
ها	Sup	Sup			

Table 8. Decision tree model test results

Field	Training word number	Testing word number		Accuracy
Economic and Social	297	126		75.40%
Sports and Health	202	86		81.40%
Sports	472	200		73.50%
Biomedical	400	160		66.5%
Weather	284	120		86.67%
Economic	414	180		83.33%
News	505	111		52.80%
Total	2810	1118		68.5%

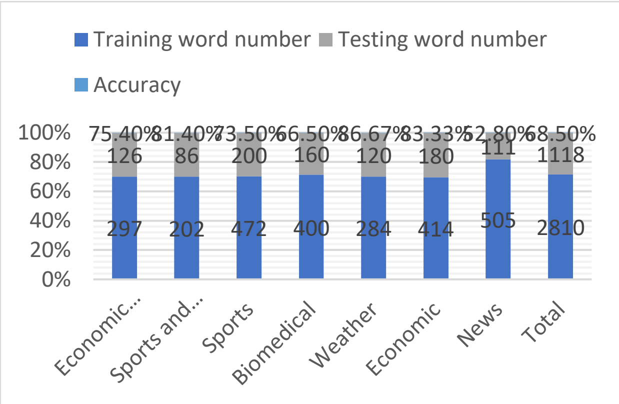

Fig 3 The Percentage of Decision tree model test results

The system has achieved satisfactory results in most fields except the news and biomedical as illustrated in table 8 and figure3 . Regarding the above results, an investigation has been done to reveal the reason for the lower accuracy of the DT model in Biomedical and News. For the Biomedical, the high number of specialized terminologies used in this field causes this situation. Also, for the news, a lot of nominal sentences face a problem to be discovered in the DT model because of the large extent of convergence in features with the verbal corrections, which is one of the complexities of the Arabic language.

4. CONCLUSION

The decision tree model implemented in this study has demonstrated its potential to enhance the utilization of the proposed relation extraction model on a larger scale, particularly in web-based text sources. The model effectively incorporates CONLLU format features and integrates them with the outputs generated by the rule-based model, presented in the preceding section. The trained model exhibited exceptional performance and accuracy when applied to texts from various domains. Its robustness and versatility are well-suited for improving relation extraction in extensive Arabic datasets. By leveraging this model, researchers and practitioners can extract valuable insights and knowledge from large-scale Arabic text sources, contributing to advancements in various fields.

Furthermore, successfully integrating the decision tree model with the rule-based model showcases the potential for combining different approaches in natural language processing. This hybrid methodology allows for harnessing the strengths of both models, resulting in a more comprehensive and effective solution for relation extraction tasks. Overall, the proposed model presented in this study represents a significant step forward in relation to extraction for Arabic text. Its demonstrated high performance, accuracy, and scalability make it a valuable tool for researchers, data scientists, and professionals working with large Arabic datasets. The findings of this study contribute to the development of natural language processing techniques and open new avenues for exploring and extracting knowledge from Arabic text sources on a large scale.

5. REFERENCES

[1] S. Ali, H. Mousa, and M. Hussien, “A Review of Open Information Extraction Techniques,” IJCI. Int. J. Comput. Inf., vol. 6, no. 1, pp. 20–28, 2019.

[2] A. Téllez-valero, M. Montes-y-gómez, and L. Villaseñor-pineda, “A Machine Learning Approach to Information Extraction,” no. February, 2005.

[3] M. Al-Ayyoub, A. Nuseir, K. Alsmearat, Y. Jararweh, and B. Gupta, “Deep learning for Arabic NLP: A survey,” J. Comput. Sci., vol. 26, no. November, pp. 522–531, 2018.

[4] R. Bemthuis et al., “ScienceDirect Business Business rule rule extraction extraction using using decision decision tree tree machine machine learning learning techniques : study into smart returnable transport techniques : A case study into smart returnable transport items it,” Procedia Comput. Sci., vol. 220, pp. 446–455, 2023.

[5] S. M. A. El-Morsy, M. Hussein, and H. M. Mousa, “Arabic open information extraction system using dependency parsing,” Int. J. Electr. Comput. Eng., vol. 12, no. 1, pp. 541–551, 2022.

[6] S. Larabi Marie-Sainte, N. Alalyani, S. Alotaibi, S. Ghouzali, and I. Abunadi, “Arabic natural language processing and machine learning-based systems,” IEEE Access, vol. 7, pp. 7011–7020, 2019.

[7] M. Luna, E. D. Gennatas, L. H. Ungar, E. Eaton, E. S. Diffenderfer, and S. T. Jensen, “Building more accurate decision trees with the additive tree,” vol. 116, no. 40, 2019.

[8] S. El-Morsy, “, ‘Arabic Open Information Extraction -model by Machine Learning to extract relation’ Github.com. https://github.com/salsama/Implementation-of-arabic-open-information-extraction/blob/master/DT_model.py,” 2023.

[9] H. Yu, “Relation Extraction with BERT-based Pre-trained Model,” pp. 1382–1387, 2020.

[10] A. R. Extraction and L. Framework, “REEL :,” 2014.

[11] S. Mohamed, M. Hussien, and H. M. Mousa, "ADPBC: Arabic Dependency Parsing Based Corpora for Information Extraction," Int. J. Inf. Technol. Comput. Sci., vol. 13, no. 1, pp. 54–61, 2021.

[12] M. Dragoni, M. Federici, and A. Rexha, "An unsupervised aspect extraction strategy for monitoring real-time reviews stream," Inf. Process. Manag., vol. 56, no. 3, pp. 1103–1118, 2019.

[13] Z. Q. Geng, G. F. Chen, Y. M. Han, G. Lu, and F. Li, "Semantic relation extraction using sequential and tree-structured LSTM with attention," Inf. Sci. (Ny)., vol. 509, pp. 183–192, 2020.

[14] K. M. Almunirawi and A. Y. A. Maghari, "A Comparative Study on Serial Decision Tree Classification Algorithms in Text Mining," vol. 7, no. 4, pp. 754–760, 2016.

[15] D. Han, Z. Zheng, H. Zhao, S. Feng, and H. Pang, "Span-based single-stage joint entity-relation extraction model," PLoS One, vol. 18, no. 2 February, pp. 1–14, 2023.

[16] A. Elnagar, R. Al-Debsi, and O. Einea, "Arabic text classification using deep learning models," Inf. Process. Manag., vol. 57, no. 1, p. 102121, 2020.

[17] M. A. R. Abdeen, S. AlBouq, A. Elmahalawy, and S. Shehata, "A closer look at arabic text classification," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 11, pp. 677–688, 2019.

[18] R. Zhou, H. N. Nguyen, and I. Sasase, "Packet scheduling for cellular relay networks by considering relay selection, channel quality, and packet utility," J. Commun. Networks, vol. 11, no. 5, pp. 464–472, 2009.

[19] D. Hutchison et al., "Computational Linguistics and Intelligent Text Processing 18th," in 18th International Conference, CICLing 2017, 2017.

[20] R. D. Brown, LNAI 8082 - Text, Speech, and Dialogue, no. September. 2013.

[21] I. Boujelben, S. Jamoussi, A. Ben Hamadou, and A. Ben Hamadou, "A hybrid method for extracting relations between Arabic named entities," J. King Saud Univ. - Comput. Inf. Sci., vol. 26, no. 4, pp. 425–440, 2014.

[22] S. E. Kramdi et al., "Approche générique pour l ' extraction de relations à partir de textes To cite this version : HAL Id : hal-00384415 Approche générique pour l ' extraction de relations à partir de textes," 2009.

[23] I. Boujelben, S. Jamoussi, and A. Ben Hamadou, "Enhancing Machine Learning Results for Semantic Relation Extraction," pp. 337–342, 2013.

[24] A. Ibm and T. A. Road, "Mining Association in Large Databases," pp. 207–216, 1993.

[25] F. Peng and A. McCallum, "Information extraction from research papers using conditional random fields," Inf. Process. Manag., vol. 42, no. 4, pp. 963–979, 2006.