IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 23, 2019, accepted April 21, 2019, date of publication May 8, 2019, date of current version May 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2915352

Information Mining From Criminal Judgments of Lahore High Court

AFNAN IFTIKHAR, SYED WAQAR UL QOUNAIN JAFFRY, AND MUHAMMAD KAMRAN MALIK
Punjab University College of Information Technology, Lahore 54590, Pakistan
Corresponding author: Afnan Iftikhar (afnan.iftikhar@pucit.edu.pk)

ABSTRACT Since the last few years, computers have become a prominent part of the court of law. Courts generate an enormous amount of unstructured text on a daily basis. Extraction of the desired information from this unstructured legal text is one of the major issues. So, there is a need to develop an intelligent system that can automatically find useful and critical information from the available text. Such a system will help judges and lawyers in their judgments and case preparations, common people in understanding law, and finding appropriate lawyer for their legal issues. Therefore, in this research, Punjab University Legal Mining System (PULMS) is developed using three different supervised machine learning algorithms; conditional random field (CRF), maximum entropy (MaxEnt), and trigram N tag (TNT). To train the system, 304 criminal miscellaneous judgments of the Lahore High Court (LHC) of Pakistan are manually tagged for nine named entities (NE). After training, among three machine learning algorithms, the system achieved significant precision, recall, and f-measure using CRF which are 0.97, 0.87, and 0.89, respectively.

INDEX TERMS Legal text mining, information extraction, named entity recognition, legal proceedings, criminal judgments.

I. INTRODUCTION

The judiciary, of any country, is one of the most important body of the governmental organization, where the judges are free to take decisions in just and rational ways, according to the existing laws.[1] Courts of any country are intended to interpret the laid-out laws for settling quarrels and other decision-making tasks. The decisions made by the judges are written down in the form of ''judgments''. ''A judgment is the expression of the opinion of the judge or magistrate arrived at after due consideration of the evidence and arguments, if any, advanced before him''.[2] A legal judgments or proceedings in general, contains brief heading, lawyers and judge information, facts of the case, a reference to some legal text, final decision, date of the decision, and similar related information.

Every year in Pakistan many lawsuits are filed in the courts[3] and the courts of Pakistan made judgments on filed cases. The Lahore High Court (LHC), Lahore, Pakistan is the oldest court of the country with the highest number of filed petitions. Figure 1 shows one of the judgments by the LHC of Pakistan.

These legal judgments are long and complex, which are difficult for a human being to examine, understand and acquire information from. Particularly, it becomes more tiresome and error-prone when one deals with more than one judgment to prepare a case. Therefore, there is a need for an intelligent system which can extract the useful information from these judgments. For this purpose, the process of Named Entity Recognition (NER) can be used.

NER is the process of extracting named entities (NEs) which are generally proper nouns from natural language text and then categorizing these entities into some pre-defined classes [1], [9]. TABLE 1 shows the named entities present in the following example,

''Muhammad Asghar traveled to Narowal in 2016.''

Here Muhamamd Asghar is a Person NE, Norowal (a place) is a Location NE and 2016 is a Time NE.

In this research, a total of nine NEs are defined on LHC judgments that are person name (Per), location (Loc), Date, reference (Ref), organization (Org), Money, case number

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Afzal.

[1]https://www.britannica.com/topic/court-law,[Accessed: January 11, 2019].

[2]https://pakistanilaws.wordpress.com/tag/judgment-writing-in-criminal-trials-in-pakistan/ ,[Accessed: April 29, 2018].

[3]https://www.thenews.com.pk/print/268487-1-87-million-cases-pending-in-pak-courts ,[Accessed: May 1, 2018].

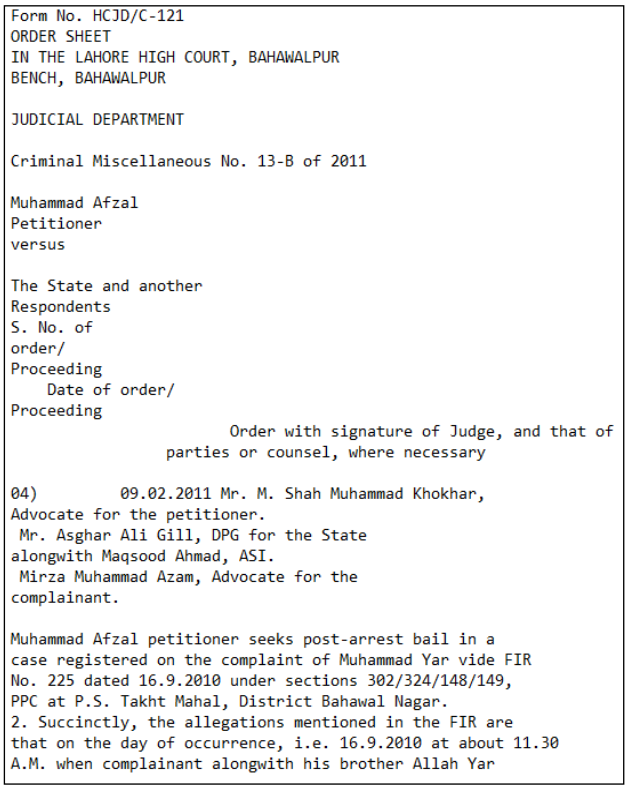


FIGURE 1. Criminal miscellaneous petition of pakistan's LHC.

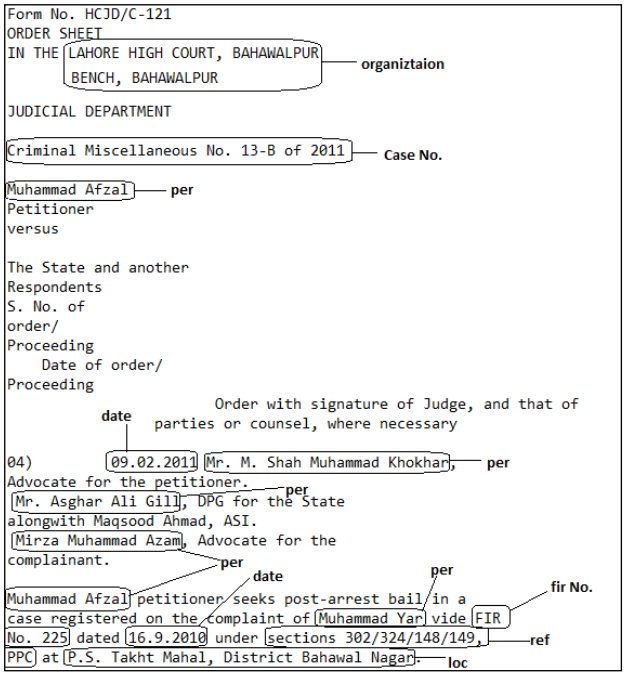


FIGURE 2. Name entities shown in criminal miscellaneous petition.

(Case No.), First Investigation Report number (FIR No.), and miscellaneous name (Misc. name). The purpose is to develop a good information extraction (IE) system which works well on the judgments. Figure 2 shows named entities marked

TABLE 1. NER example.

Word	Categories
Muhammad	Person
Asghar	Person
Travelled	Non-entity
To	Non-entity
Narowal	Location
In	Non-entity
2016	Time

on a criminal miscellaneous judgment during the judgment tagging process. The main contributions of the work are listed below:

- A corpus of 292,502 tokens from 304 criminal miscellaneous judgments is tagged for nine NEs and verified using multi annotator agreement. To date, no such work has been done on the legal proceedings ofPakistan.
- Three different machine learning algorithms namely CRF, MaxEnt and TNT are trained and tested on the corpus.
- Comparison of the results generated from three learning algorithms are reported.

The rest of the paper is organized as follows; Section II describes the related work and background required to understand NER extraction from legal documents. Section III explains the process of data acquisition and data annotation. In Section IV and Section V the methodology used for the experiments is explained and the experimental results are illustrated. Finally, Section VI provides discussion, and Section VII concludes the paper with future research directions.

II. BACKGROUND AND RELATED WORK

Since last few years, computers have become a prominent part in the court workings. As a result, an enormous amount of unstructured text is generated day by day by each court. These unstructured texts are valuable sources of information. Automatic extraction of desired and useful information from these unstructured legal text documents is a very significant and open problem. So, there is a need to develop a software tool that finds out useful and critical information from available data to help lawyers, judges, and common people.

In [2]–[6], information such as the name of the lawyer or the judge, city, court, law firms, state etc. were extracted from unstructured text of different legal judgments, proceedings, case law, court dockets, law reviews, Medline abstracts, and legal briefs. The process of extraction of these entities was done by locating the representative paragraphs (i.e. a paragraph in the legal text containing the required information) in the document. After locating the representative paragraphs, the desired information was extracted using different approaches such as lookup, contextual, statistical model or regular expression parsing. The extracted information was then stored in a structured form [4]. After the creation of structured data, it was used in creating the profile of lawyers,

judges, and experts which resulted in 95% precision[4] and 60% recall[5] [5]. The extracted individual names such as attorney and judge name were then matched to the personal biography (i.e. personal profile containing information like person name, address, phone number etc.) records by using the Naïve Bayesian Network. The proposed approach linked the biographic profile to the specific person name in the text by hyperlinks, which resulted in average 98.5% precision and 91% recall [3].

Dozier and Zielund [6] used the information such as court, judge, lawyer, document type and title for the creation of repository of expert witnesses, lawyers, and judges. The repository was created to show the summaries exhibiting the relationships among individual entities based on their document co-occurrence and cross-document co-references e.g. relationship summary can show that which expert witnesses were hired by which lawyer, and which lawyers have appeared before which judges etc. Information in the database can also help in trend analysis. It resulted in 95% precision and 60% recall. Information was also extracted for the creation of litigation history database which can help in a different type of trend analysis [2].

Poudyal *et al.* [7] present the results of eight different machine learning algorithms for automatic recognition and extraction of different NEs such as person, organization, date, and regulation laws. For experimentation with Minorthird (an open source software tool), eight different algorithms with their default parameters were considered. Algorithms include CRF, Voted Perceptron Semi-Markov Model (VPSMM), Voted Perceptron Conditional Markov Model (VPCMM), Semi-CRF, Maximum Entropy Markov Model (MEMM), Voted Perceptron Hidden Markov Model (VPHMM), Support Vector Machine Conditional Markov Model (SVMCMM), and Voted Perceptron Semi-Markov Model 2 (VPSMM2). Algorithms were evaluated to identify the best algorithm for each entity and to compare the number of manually tagged entities to the number of entities tagged by the machine. Thus, the algorithm with highest f-score was selected against each entity. Hidden Semi-Markov Models algorithm gave highest f-measure of 0.910 for a date, Support Vector Machine (SVM) gave f-measure of 0.538 for Organization and 0.865 for Person and CRF has highest f-measure of 0.853 for Regulation Law.

Dernoncourt *et al.* [8] have developed a tool named "NeuroNER" by using Artificial Neural Network (ANN). The proposed tool was able to identify four NEs such as Person, Organization, Location and Miscellaneous names from freely and publically available data sets of CoNLL 2003 and i2b2 2014.

In Table 2 different techniques applied for the extraction of information from different type of documents have been represented. The results of these techniques in terms of precision and recall have also been shown in Table 2.

[4]Precision = true positive / (true positive + false positive)

[5]Recall = true positive / (true positive + false negative)

TABLE 2. Approaches and results for IE.

Paper	Data Set	Approach	Result	
			Precision	Recall
[2]	3 million case law opinions and 40 million dockets	Combination of NE extraction, co-reference resolution, document classification techniques, record linkage, and relationship extraction among NEs,	90%	88%
[3]	600 case law documents	extraction by MUC-style and linking by naïve Bayesian inference network	98%	90%
[4]	400 legal documents	lookup, context rules, and statistical models for NER	96%	89%
[5]	300,000 jury verdict and settlement documents	regular expression parsing, Bayesian network	95%	60%
[6]	300,000 jury verdict and settlement documents	Message Understanding Conferences, Bayesian network	95%	60%
[7]	20 judicial decisions	VPSMM, VPCMM, SVMCMM, CRF, SemiCRF, MEMM, VPHMM, VPSMM.	65%	47%
[8]	CoNLL 2003 and i2b2 2014	ANN	86.3%	84.6%

III. DATA ACQUISITION

There are different types of legal judgments such as civil, revenue, writ petition, criminal judgments etc. A particular type of legal judgments can further be divided into sub-classes. For example, criminal judgments are of different sub-types such as criminal appeal, criminal miscellaneous, criminal revision etc. In year 2016, around 116,347 criminal miscellaneous cases were registered in the Punjab province of Pakistan.[6] This huge number has led to the selection of criminal miscellaneous judgments for this research. As in these judgments the name of the police station where FIR has been registered is mentioned. This information can help to find the type of cases occurring in different geographical areas and thus provide information to take related safety measures.

The judgments used in this research are available on LHC's website.[7] First a scrapper for the extraction of judgments from LHC is developed. In total 3,503 judgments were extracted. From this large collection, criminal miscellaneous judgments were separated which resulted in 304 judgments. These judgments have 292,502 tokens. The length of each judgment varies from 2 to 30 pages. Each judgment can be partitioned into header, body and footer which contains some specific information.

[6]https://punjabpolice.gov.pk/crimestatistics, [Accessed: January 16, 2018].

[7]http://data.lhc.gov.pk/reported_judgments/judgments_approved_for_reporting, [Accessed: May 1, 2018]

TABLE 3. Frequency of most representative tokens.

Tokens	Frequency
Petitioner	4,355
Complainant	1,666
Muhammad	1,253
Counsel	1,141
FIR	1,121

TABLE 4. Statistics about training and testing data.

Name Entity (NE)	Training Tokens	Testing Tokens
Person	11,657	2,435
Location	2,272	428
Date	1,751	409
Organization	1,611	467
Money	767	189
Misc.name	1,292	356
FIR No.	645	168
Case No.	3,763	938
Reference	4,275	888
Name entities	28,033	6,578

- The header of the judgment contains information about the lawyer, date of hearing, petitioner and respondent's name.
- The body contains brief facts of the case, reference considered, and the decision of the case.
- The footer of the judgment gives information of judge and the minute taker.

These partitioning helped in identification of particular information e.g. if one wants to know about the lawyers involved in the case this information can be found in judgment header, or for the name of judge only footer is required to get the desired information. The partitions of the judgment into header, body and footer is shown in Figure 3. In this research, task of NER is applied on whole judgment without any partition whose example was shown in Figure 2. Annotation guidelines are used for manual annotation of the judgment, so that quality tagged data can be produced.

A. ANNOTATION GUIDELINE

The corpus for the training of NER classifiers consists of 292,502 tokens from 304 criminal miscellaneous judgments. After the removal of stop words from 292,502 tokens the top 5 most representative tokens along with their frequencies are shown in Table 3. After consulting with law specialists, the tokens were manually tagged using a software tool as shown in Figure 4. The tool is similar to work mentioned in [9]. Each token is assigned one of nine NEs, and if, a token is not an NE otherwise O (others) tag is assigned. In Table 4, the total number of each NE in both training and testing data set is shown.

While annotating the tokens, three things were taken under consideration that is what should be annotated? Which token should not be annotated as NE? and how much sequence to be annotated? Considered NEs along with their examples are described in Table 5.

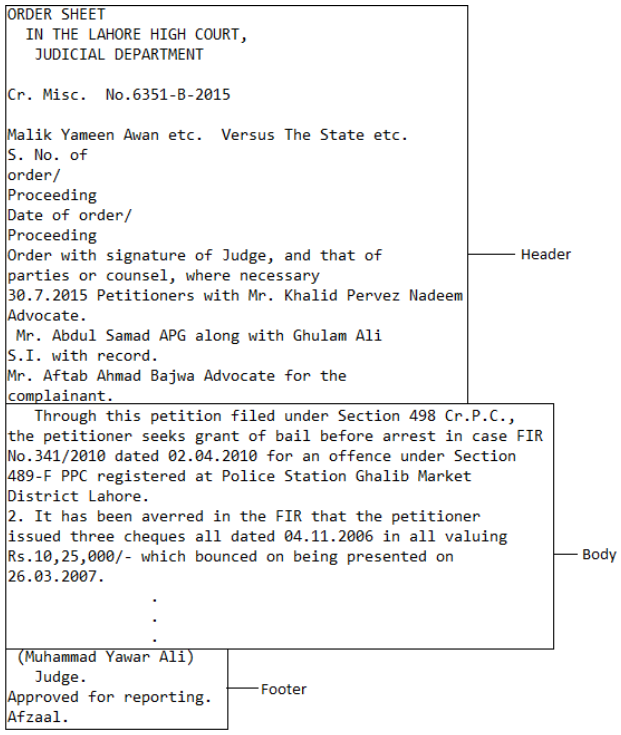


FIGURE 3. Partitioned judgment.

FIGURE 4. Screenshot of GUI for tag verification.

Following guidelines have been considered while doing manual annotation of NEs.

- **Person Name: (Per)** includes the name of a person e.g. Muhammad, Umer, Usman etc., nickname of a person e.g. Kaloo, Billa, Dittu etc. and surname such as Mian, Malik, Chaudhry (ch.), Khan, Dar, Wattoo etc. placed either before or after the person name. These surnames helps in distinguishing two persons for e.g. Malik Iftikhar Ahmed and Mian Iftikhar Ahmed both are going to be distinguished as the names of two different persons. But in sentence, "Ali Ahmed of Malik family", the word Malik is not a person name but is a family name.

TABLE 5. Name entity tags.

Name Entity Tags	Description	Example
Per	Name of person	Ch. Zain ul Abidin, Salman Wattoo
Loc	Name of any place/ location	Pakistan, Police station Fort Abbas, Bahawal Naggar
Date	Any date	10/31/2016, 10th September 2016
Ref	Reference to different section of "Criminal CODES"	Section 497 PPC, u/s 330 Cr.P.C
Org	Name of any organization/business	FIA, Lahore high court
Money	Any money	Rs. 15000/. , Rupees ten lac.
Case No.	Any Case number	Criminal Miscellaneous No. 4431-B of 2010, Crl. Misc. No. 1615-B of 2014
FIR No.	First Investigation Report (FIR) number	FIR.No.287/2010, FIR No.342/11 etc.
Misc. name	Miscellaneous name to identify respondent	The state, Session judge Daska

```
IN THE <LAHORE HIGH COURT, BAHAWALPUR
BENCH, BAHAWALPUR ---- org>

JUDICIAL DEPARTMENT

<Criminal Miscellaneous No. 13-B of 2011 ---- case No.>
 <Muhammad Afzal ---- per>
Petitioner
versus

<The State ---- Misc. name> and another
Respondents
S. No. of
order/
Proceeding
    Date of order/
Proceeding
                        Order with signature of Judge, and that of
                parties or counsel, where necessary

04)        <09.02.2011 ---- date> <Mr. M. Shah Muhammad Khokhar ---- per>,
Advocate for the petitioner.
 <Mr. Asghar Ali Gill ---- per>, DPG for the State
alongwith <Maqsood Ahmad ---- per>, ASI.
 <Mirza Muhammad Azam ---- per>, Advocate for the
complainant.

 <Muhammad Afzal ---- per> petitioner seeks post-arrest bail in a
case registered on the complaint of <Muhammad Yar ---- per> vide <FIR
No. 225 ---- FIR No.> dated <16.9.2010 ---- date> under <sections 302/324/148/149,
PPC ---- ref> at <P.S. Takht Mahal, District Bahawal Nagar ---- loc>.
2. Succinctly, the allegations mentioned in the FIR are
that on the day of occurrence, i.e. <16.9.2010 ---- date> at about 11.30
A.M. when complainant alongwith his brother <Allah Yar ---- per>
```

FIGURE 5. Final tagged judgment.

- **Location tag: (Loc)** includes names of countries, cities, places within the city e.g. (Narang Moor, Thing Moor etc.) and some addresses (as Cyber Crime Circle, Lahore). Words like Pakistani, Lahori etc. which shows nationality or as in ''Lahore High Court'' Lahore is a part of an organization and not a location, thus these tokens were not marked as location.
- **Date tag: (Date)** includes all different date formats, for example, 10/30/2016, 10th April 2010 or April 10, 2010 etc. The tokens like mid of February, end of year or 81 days were not marked as date.
- **Organization: (Org)** tag includes the name of all organizations, business, and companies either of government or private. For example, Lahore High Court, Nao Bahar Bottlers, FIA etc. The token like brands or products name such as 7up, Coca Cola in sentence like ''bottle of Coca Cola were broken'' was not marked as Org.
- **Miscellaneous name: (Misc. name)** the respondent tag includes all proper nouns for referring to the respondent of case stated as ''Mian Junaid vs. the State''. Here the State was marked as Misc. name.
- **Reference: (Ref)** within a judgment there are references to different petitions, judgments, acts, laws, etc. In this work, only the reference to the ''Criminal CODES'' i.e. the Pakistan Penal Code (PPC) and the Criminal Procedure Code (Cr.P.C.) was considered. For example, section 173 of The Code of Criminal Procedure, 1898 was marked as a single reference. The tokens like 1991 SCMIR 322 or PLD 2009 Supreme Court 427 were not considered as reference.
- **Money tag: (Money)** includes all the tokens referring to amount of money either in the form of numbers or words along with currency.
- **First Investigation Report Number: (FIR No.)** tokens referees to some FIR No. like FIR.No.287/2010. In sentence like, ''delay of 81 days in lodging the FIR'', the word FIR doesn't specify any FIR No. and thus was not marked.
- **Case Number: (Case No.)** tokens giving the information about case type, number and year were considered as a single Case No. E.g. Criminal Miscellaneous No. 4431-B of 2010.

As for how much to annotate, it was taken under consideration that NE tags assigned to single words are not appropriate as in some cases NE may consist of more than one token, for example, Malik Nadeem Awan is a name of a single person consisting of three tokens and it should be assigned a single tag as <Malik Nadeem Awan—-per> instead of assigning separate entity tag for each single token i.e. <Malik—- per> <Nadeem—- per> <Awan—- per>.

In addition to this problem, single tokens such as ''station'' cannot be classified to any entity tag whereas ''police station Faisal Town, Lahore'' will be assigned to location class, in the same way any tag cannot be assigned to the token ''Allah'' as it does not belong to any of the considered class, whereas a sequence of token ''Allah Dita'' is a person's name. In Figure 5, a judgment tagged with considered annotation guideline is shown.

IV. METHODOLOGY

For the Named Entity Recognition, a dataset of the criminal miscellaneous judgments was used to train three different classifiers with specific features. Three classifiers were used so that in future voting technique can be applied to get tag/class of each token, so that better results can be achieved. Following are the classifiers used in this research:

- Conditional Random Field (CRF), an implementation of Stanford NLP Group [10].

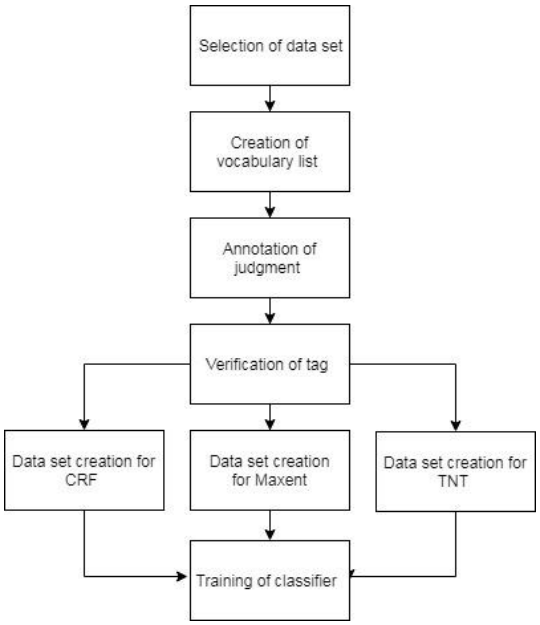


FIGURE 6. Process flow diagram of methodology.

```
Crl.Misc.No.1799-B  caseNo.
of  caseNo.
2010    caseNo.
.   O
Mehboob per
Ahmad   per
.   O
The Misc.name
State   Misc.name
etc.    O
12.10.2010  date
.   O
Sardar  per
Mehmood per
Iqbal   per
Khakwani    per
,   O
Advocate    O
for O
the O
petitioner. O
```

```
caseNo. Crl.Misc.No.1799-B
caseNo. of
caseNo. 2010
O   .
per Mehboob
per Ahmad
O   .
Misc.name   The
Misc.name   State
O   etc.
date    12.10.2010
O   .
per Sardar
per Mehmood
per Iqbal
per Khakwani
O   ,
O   Advocate
O   for
O   the
O   petitioner.
```

FIGURE 7. a) Training data for CRF and TNT classifiers, b) Training data for MaxEnt classifier.

- Maximum Entropy classifier (MaxEnt), an implementation of Stanford NLP Group [11].
- Trigrams 'N' Tags (TNT) [12].

NER is a sequence labelling task which can be modelled by a sequence model. All three classifiers selected for this study are suitable for sequence labelling problem [13], [14]. These classifiers have been widely used for NER [15], [16], [17] and have provided promising results. The process flow of the whole work is shown in Figure 6.

A. CLASSIFIERS DATA SET

The total 304 Criminal miscellaneous judgments were divided into 80% (i.e. 244 judgments) and 20% (i.e. 60 judgments) ratio for training and testing datasets respectively. Training dataset comprises of 244,290 tokens and testing dataset contains 48,211 tokens. To train model using CRF and TNT, the data is prepared in a format where first word is token and second word is NE as shown in Figure 7 a) and for MaxEnt, first word represents NE and second word represents token as shown in Figure 7 b.

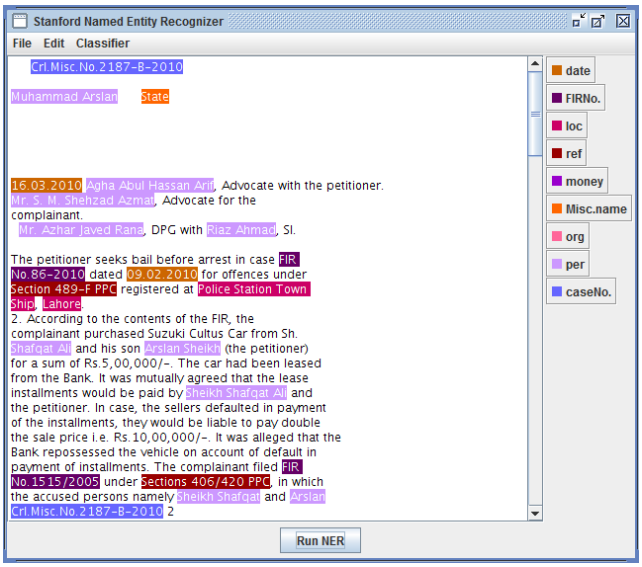


FIGURE 8. CRF NE tagging tool.

TABLE 6. Features used by CRF for NER.

CRF Feature	NER
Previous Word	Yes
Current Word	Yes
Next Word	Yes
Length of N-gram	Size 6
Current Word Character n-gram	Yes
Class feature	Yes
Class combination	Yes
Previous class combination	Yes

TABLE 7. Features used by MaxEnt for NER.

Maxent Feature	NER
N-Gram	Yes
Class feature	Yes
Maximum N-Gram length	4
Minimum N-Gram length	1
Prefix of the string	Yes
suffix of the string	Yes

B. CLASSIFIERS TRAINING

Datasets created in previous section is used to train and test three selected classifiers. Different features used for the training of CRF, MaxEnt and TNT are described in Table 6, Table 7 and Table 8 respectively. Figure 8, shows the screen shot for the CRF tool trained for NER.

V. RESULTS

For the training of classifiers, combination of various features such as useWordPairs, useTags and wordShape were tried.

TABLE 8. Features used by TNT for NER.

TNT Feature	NER
N-Gram	Yes
N-Gram length	3
suffix trie	Yes
Suffix trie length	up to 10 characters
Data smoothing	linear interpolation

TABLE 9. CRF confusion matrix.

		ACTUAL									
		O	Per	Case No.	Org	FIR No.	Misc. name	Loc	Ref	Date	Money
PREDICTED	O	41996	15	35	48	0	7	16	14	9	2
	Per	27	2361	0	3	0	2	6	0	0	1
	Case No.	0	0	903	0	0	1	0	0	0	0
	Org	6	0	0	400	0	0	9	0	0	0
	FIR No.	0	0	0	0	168	0	0	0	0	0
	Misc. name	5	0	0	0	0	344	0	0	0	0
	Loc	8	15	0	16	0	2	385	0	0	0
	Ref	7	0	0	0	0	0	12	873	0	0
	Date	1	9	0	0	0	0	0	0	321	0
	money	9	0	0	0	0	0	0	0	0	184

Among those combination of feature, the combination of features which showed promising results in terms of precision, recall and f-measure are presented in Table 6, Table 7 and Table 8.

For the evaluation of trained models precision, recall, and f-measure for each NE is calculated. In the following sections results of various experiments have been discussed briefly.

A. RESULT OF CRF CLASSIFIER

During the testing phase, CRF tagged 48,211 words of 60 randomly selected documents in 7.7 seconds. Table 9 represents the confusion matrix for nine NEs and the O (Others) tag. In the Table 9, rows represent the predicted entity and columns represent the actual entity.

Table 10 represents the precision, recall, and f-measure of the CRF classifier. From Table 10, is could be seen that the FIR number resulted in highest f-measure of 1.00 and location resulted in lowest f-measure of 0.901. The average precision, recall, and f-measure of the CRF classifier are 0.97, 0.95 and 0.96 respectively.

B. RESULT OF MAXENT CLASSIFIER

In Table 11, confusion matrix is shown, which illustrates the result of MaxEnt classifier which has tagged 48,211 words 60 randomly selected criminal miscellaneous judgments in 2.3 seconds. In the Table 11, rows represent the predicted entity and columns represent the actual entity.

TABLE 10. Evaluation results for CRF classifier.

Entity	Precision	Recall	F-measure
FIR no.	1.000	1.000	1.000
Per	0.983	0.983	0.983
Loc	0.903	0.899	0.901
Org	0.963	0.856	0.907
Date	0.969	0.972	0.971
Ref	0.978	0.984	0.981
Misc. name	0.985	0.966	0.975
Case no.	0.998	0.962	0.980
Money	1.000	0.983	0.991

TABLE 11. MaxEnt confusion matrix.

		ACTUAL									
		O	Per	Case No.	Org	FIR No.	Misc. name	Loc	Ref	Date	Money
PREDICTED	O	41705	64	158	106	93	178	31	92	8	24
	Per	26	2313	0	2	0	0	27	0	1	1
	Case No.	106	0	773	0	19	1	0	10	4	0
	Org	36	1	0	328	0	0	21	3	0	0
	FIR No.	6	0	7	0	51	0	1	2	0	0
	Misc. n.	29	2	0	2	0	169	0	0	0	0
	Loc	26	18	0	29	0	8	340	2	0	0
	Ref	105	0	0	0	3	0	8	779	0	0
	Date	5	2	0	0	2	0	0	0	317	0
	money	6	0	0	0	0	0	0	0	0	162

TABLE 12. Evaluation results for MaxEnt classifier.

Entity	Precision	Recall	F-measure
FIR no.	0.761	0.303	0.434
Per	0.975	0.963	0.969
Loc	0.803	0.794	0.799
Org	0.843	0.702	0.766
Date	0.972	0.960	0.966
Ref	0.870	0.877	0.873
Misc. name	0.836	0.474	0.605
Case no.	0.846	0.824	0.835
Money	0.964	0.866	0.912

Table 12 represents the precision, recall, and f-measure of MaxEnt classifier. According to which, person resulted in highest f-measure of 0.969 and FIR number resulted in lowest f-measure of 0.434. The average precision, recall, and f-measure of the MaxEnt classifier are 0.87, 0.75 and 0.79 respectively.

C. RESULT OF TNT CLASSIFIER

To test TNT trained classifier, a labeled data set of 60 judgments comprising of 48,211 tokens have been used. The result of the testing phase is shown in Table 13.

Table 14, shows the precision, recall and f-measure of TNT model. According to which the entity "date" resulted in highest f-measure of 0.950 and entity "location" resulted in lowest f-measure of 0.858. The average precision, recall, and f-measure of trained TNT model are 0.89, 0.94 and 0.91 respectively.

TABLE 13. TNT confusion matrix.

		ACTUAL									
		O	Per	Case No.	Org	FIR No.	Misc. name	Loc	Ref	Date	Money
PREDICTED	O	41506	13	30	21	13	7	15	15	7	6
	Per	34	2367	0	2	0	0	6	0	0	1
	Case No.	167	0	904	1	13	1	2	1	0	0
	Org	82	0	0	437	0	3	18	0	0	1
	FIR No.	8	0	2	0	139	0	0	2	0	0
	Misc. n.	7	0	0	1	0	342	0	0	0	0
	Loc	50	15	0	5	0	3	378	2	0	0
	Ref	166	0	0	0	1	0	9	864	0	0
	Date	15	5	2	0	2	0	0	3	323	0
	Money	15	0	0	0	0	0	0	0	0	179

TABLE 14. Evaluation results for TNT classifier.

Entity	Precision	Recall	F-measure
FIR no.	0.920	0.827	0.871
Per	0.982	0.986	0.984
Loc	0.834	0.883	0.858
Org	0.807	0.935	0.867
Date	0.922	0.978	0.950
Ref	0.830	0.974	0.896
Misc. name	0.977	0.960	0.968
Case no.	0.830	0.963	0.891
Money	0.922	0.957	0.939

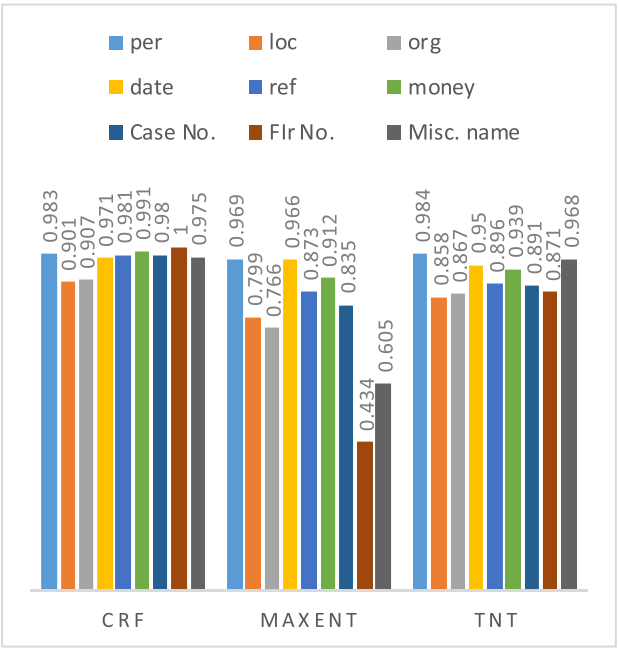


FIGURE 9. F-measure of each entity.

VI. DISCUSSION

In this part, the behavior of all trained classifiers according to their results have been summarized. In Figure 9, bar chart of f-measure of each entity according to the CRF, MaxEnt and TNT have been presented. From this figure it could be observed that, the person name has the highest value in the TNT with f-measure of 0.98. For location, organization, date, reference, money, case number, FIR number and Misc. name the CRF gave the highest f-measure value which are 0.901, 0.907, 0.971, 0.981, 0.991, 0.98, 1.00 and 0.975 respectively. For most of the entities, results of the CRF classifier are better than the other two classifiers in comparison.

Out of the different types of judgments of LHC of Pakistan, such as civil petition, criminal revision, criminal appeal, etc., currently in this research, classifiers have been trained to NE's from criminal miscellaneous judgments. The approach proposed in this research can be applied on other types of judgments too as well as the extraction of entities other than the nine mentioned in this research.

VII. CONCLUSION

Computers and computing are becoming ubiquitous in all spheres of life and so as in courts. Courts are producing an enormous amount of text data in the form of legal proceedings for public awareness and guidance. Processing of this huge amount of data is practically impossible for human. Hence, various machine learning techniques could be applied on this data to make it human consumable. In order to achieve this task, first the legal data should be tagged to apply machine learning algorithms and later the trained machine learning models could be used for extracting meaningful information from the legal data.

As a pioneering work on legal data of Pakistan, in the current study, 304 criminal miscellaneous judgments of the Lahore High Court, Pakistan, are tagged with nine named entities (NEs), namely Person name, Location, Organization, Money, Date, Reference, First Investigation Report (FIR) number, Case number and miscellaneous name. The tagged data is divided into training and testing sets. Three widely used sequence labeling algorithms including Conditional Random Field (CRF), Maximum Entropy (MaxEnt) and Trigram's N Tag (TNT) are trained and tested. It is found that TNT algorithm has outperformed the other two algorithms for three NEs including Person name, Location and Organization with f-measure of 0.984, 0.858 and 0.867 respectively. The CRF; resulted into the highest f-measures for remaining six NEs including Date, Reference, Money, Case number, FIR number and Miscellaneous name, with values of 0.97, 0.975, 0.99, 0.973, 1 and 0.98 respectively.

In comparison to various reported results in literature on different datasets, these initial results seem promising. As current results are obtained on only "criminal miscellaneous judgments", hence, to increase confidence on reported results, more experiments on different types of judgments are needed. In future, other type of judgments including civil, bail etc. of the Lahore High Court would also be tagged and used in training as well as testing of algorithms. Incorporation of different judgment types in the dataset will require to introduce new NEs which will increase the NE count. Hence, preparation of a comprehensive dataset, carrying variety of court judgments is a prospective future work. Besides revisiting dataset, as recently deep learning frameworks have

generated state-of-the-art results on NER, it is also planned to apply various deep learning algorithms for NER on legal text. Furthermore, pre-trained word embeddings are being widely employed for application of deep learning algorithms on textual data. Hence, preparation of specialized pre-trained word embedding for legal text is also a prospective research activity.

Once the desirable results of NER from legal text are achieved, variety of applications could be built for its effective utilization. For example, extracted NEs can help in the creation of a question-answer system where questions could be answered through extracted NEs and their relationships. These NEs can also be used in the creation of various bibliographic profiles. Hence provision of a variety of applications on extracted NEs is also future work.

REFERENCES

[1] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Lingvisticæ Investigationes*, vol. 30, no. 1, pp. 3–26, Jan. 2007.

[2] M. Chaudhary *et al.*, "Mining legal text to create a litigation history database," in *Proc. IASTED Int. Conf. Law Technol.*, Cambridge, MA, USA. 2006.

[3] C. Dozier and R. Haschart. "Automatic extraction and linking of person names in legal text," in *Proc. RIAO*, 2000, pp. 1305–1321.

[4] C. Dozier, R. Kondadadi, M. Light, A. Vachher, S. Veeramachaneni, and R. Wudali, "Named entity recognition and resolution in legal text," in *Semantic Processing of Legal Texts*. Berlin, Germany: Springer, 2010, pp. 27–43.

[5] C. Dozier, P. Jackson, X. Guo, M. Chaudhary, and Y. Arumainayagam, "Creation of an expert witness database through text mining," in *Proc. 9th Int. Conf. Artif. Intell. Law*, New York, NY, USA, 2003, pp. 177–184.

[6] C. Dozier and T. Zielund, "Cross document co-reference resolution applications for people in the legal domain," in *Proc. Conf. Reference Resolution Appl.*, Barcelona, Spain, 2004, pp. 1–18.

[7] P. Poudyal, L. Borrego, and P. Quaresma, "Using machine learning algorithms to identify named entities in legal documents: A preliminary approach," 2011.

[8] F. Dernoncourt, J. Y. Lee, and P. Szolovits. (May 2017). "NeuroNER: An easy-to-use program for named-entity recognition based on neural networks." [Online]. Available: https://arxiv.org/abs/1705.05487

[9] M. K. Malik, "Urdu named entity recognition and classification system using artificial neural network," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 17, no. 1, pp. 2:1–2:13, Nov. 2017.

[10] J. R. Finkel, T. Grenager, and C. Manning, "Incorporating non-local information into information extraction systems by Gibbs sampling," in *Proc. 43rd Annu. Meeting Assoc. Comput. Linguistics*, Stroudsburg, PA, USA, 2005, pp. 363–370.

[11] C. Manning and D. Klein, "Optimization, maxent models, and conditional estimation without magic," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol., Tuts.*, Stroudsburg, PA, USA, vol. 5, 2003, p. 8.

[12] T. Brants, "TnT: A statistical part-of-speech tagger," in *Proc. 6th Conf. Appl. Natural Lang. Process.*, Stroudsburg, PA, USA, 2000, pp. 224–231.

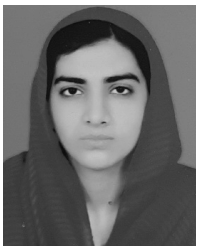
[13] P. Liu and H. Meng, "SeemGo: Conditional random fields labeling and maximum entropy classification for aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, Dublin, Ireland, 2014, pp. 527–531.

[14] N. Nguyen and Y. Guo, "Comparisons of sequence labeling algorithms and extensions," in *Proc. 24th Int. Conf. Mach. Learn.*, New York, NY, USA, 2007, pp. 681–688.

[15] V. C. Tran, N. T. Nguyen, H. Fujita, D. T. Hoang, and D. Hwang, "A combination of active learning and self-learning for named entity recognition on Twitter using conditional random fields," *Knowl.-Based Syst.*, vol. 132, pp. 179–187, Sep. 2017.

[16] S. K. Saha, S. Sarkar, and P. Mitra, "A hybrid feature set based maximum entropy Hindi named entity recognition," in *Proc. 3rd Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2008, pp. 1–7.

[17] A. McCallum and W. Li, "Early results for named entity recognition with conditional random fields, feature induction and Web-enhanced lexicons," in *Proc. 7th Conf. Natural Lang. Learn. (HLT-NAACL)*, Stroudsburg, PA, USA, vol. 4, 2003, pp. 188–191.



AFNAN IFTIKHAR received the B.S. degree (Hons.) in computer science (CS) with a major in information technology from the Forman Christian College, in 2014, and the M.Phil. degree in CS from the Punjab University College of Information Technology (PUCIT), University of the Punjab, Lahore, Pakistan, in 2016. She is currently a Lecturer with the University of Management and Technology (UMT), Lahore. She has been a Visiting Lecturer with PUCIT. Her research interests include natural language processing, data mining, and machine learning.

SYED WAQAR UL QOUNAIN JAFFRY received the Ph.D. degree in computer science from the VU University Amsterdam, The Netherlands. He is currently an Assistant Professor and the Founding Head of the Artificial Intelligence and Multidisciplinary Research Laboratory (AIM-RL), Punjab University College of Information Technology (PUCIT), University of the Punjab, Lahore, Pakistan. He has more than 17 years of teaching and research experience at University level. At AIM-RL, he is employing techniques of modeling and simulation, data mining, information retrieval, and serious games to address various multidisciplinary research questions.

MUHAMMAD KAMRAN MALIK received the Ph.D. degree in computer science. He is currently an Assistant Professor with the Punjab University College of Information Technology (PUCIT), University of the Punjab, Lahore, Pakistan, in 2018. He has more than 15 years of teaching and development experience. He has provided Consultancy to many multinational firms on the above-mentioned areas. He has authored 25 journal and conference papers and holds one U.S. patent. His research interests include natural language processing, machine learning, and data science.