NER for Hindi Language using Association Rules

Conference Paper · August 2014

DOI: 10.1109/ICDMIC.2014.6954253

CITATIONS
14

READS
1,421

3 authors:

Dr Arti Jain
Jaypee Institute of Information Technology
34 PUBLICATIONS **272** CITATIONS
SEE PROFILE

Divakar Yadav
Indira Gandhi National Open University (IGNOU)
136 PUBLICATIONS **990** CITATIONS
SEE PROFILE

Devendra Kumar Tayal
IGDTUW
57 PUBLICATIONS **567** CITATIONS
SEE PROFILE

NER for Hindi Language using Association Rules

Arti Jain
CSE/IT
JIIT
Noida, India
arti.jain@jiit.ac.in

Divakar Yadav
CSE/IT
JIIT
Noida, India
divakar.yadav@jiit.ac.in

Devendra Kr. Tayal
CSE
IDTUW
New Delhi, India
dev_tayal2001@yahoo.com

***Abstract*—In this paper, we propose a state-of-art association rule mining algorithm for Hindi NER. Association rules are one of the key components of the data mining. Mined rules are of – TYPE 1, TYPE 2 and Type 3 i.e. dictionary, bi-gram and feature rules respectively. We consider corpus of news articles (100 training and 50 test sets) from leading Hindi newspapers. Hindi NER shows significant increase in performance when TYPE 2 rules are combined with TYPE 1 or with TYPE 3.**

Keywords— data mining; association rule; named entity; natural language processing;

I. INTRODUCTION

Named Entity Recognition (NER) [3-5] [18] [29-31] is defined as a task to identify and then to classify proper nouns or Named Entities (NEs) in a text document into some pre-defined classes. These classes can be of the types, for example person names (e.g. first name, last name, middle name etc. of a person), organization names (e.g. government, private bodies, company names etc.), location names (e.g. village, town, city, state, region names etc.), miscellaneous names (e.g. date, time, number, monetary expression etc.) and none-of-above (i.e. Not a Named Entity (NNE)). Also, the text document can be any written text in an Indian language (say Hindi) or in any other language. For example, consider a sentence in the Hindi as: *"प्रधानमंत्री मनमोहन सिंह को अपने साउथ ब्लाक पीएमो कार्यालय से भावभीनी विदाई मिली।"* Here, NER process can identify the named person as *"मनमोहन सिंह"*, location as *"साउथ ब्लाक"* and organization as *"पीएमो"*. In this particular case, the NER system may have applied simple rules for the following observations. Words directly following the term *"प्रधानमंत्री"* refer a name of a person, and the words preceding the terms *"ब्लाक"* and *"कार्यालय"* are the name of location and organization respectively. Thus the NE recognition can be based on a variety of document features and syntax patterns of the terms within the text.

In 1995, Message Understanding Conference (MUC) [1] [19] [21] has reflected the creation of subtask for NER in an Information Extraction (IE) field of the Natural Language Processing (NLP) [2]. Since then, NER is a precursor to the wide NLP applications such as Question Answering (QA) [3 - 5], Machine Translation [6] and so on. In the biomedical domain [7-9], again NER is a vital subtask in the organizing and retrieval of information.

We have observed that NER systems for languages such as English have come up with high accuracy [10-12]. Such system takes advantage of huge amount of annotated corpora, and other language resources. However, for our Indian languages there is a scarcity of resources (e.g. lack of annotated corpora, comprehensive gazetteer lists, morphological analyzers, coarse and fine grained part-of-speech tags etc). So, developing a NER system with remarkable evaluation measures for such resource poor languages is a difficult task.

In our current work, we propose a NER method for Hindi language using Association Rule Mining (ARM). The association rules [35-36] [38] can potentially define the term features and syntax patterns within the text documents which aid the NE classes to be mined (please refer Section IV).

The remainder of this paper is organized as follows. Section II gives a brief survey of different NER techniques, across languages and domains. Section III illustrates various NER challenges for Hindi. Section IV describes NER using ARM and is divided into three sub-sections. Section IV.A discusses association rule mining. Section IV.B proposes our NER algorithm. Section IV.C discusses NER for Hindi using ARM. Then Section V is about experimentation and results of our work. Finally, Section VI concludes our paper.

II. LITERATURE SURVEY AND RELATED WORK

Named entity recognition uses a variety of techniques. Broadly, NER approaches can be classified as Machine Learning (ML) [13] [15] [20] based NER, Linguistic NER [23] [25] and Hybrid NER [24] [26]. Machine Learning (ML) based NER typically takes care of model generation (using patterns and relations) from annotated text. Some of the successful ML techniques for NER are Hidden Markov Model (HMM) [10], Maximum Entropy (MaxEnt) [11-12], Conditional Random Field (CRF) [13], Decision Tree (DT) [14], and Support Vector Machine (SVM) [15]. NYMBEL system [10] is based on HMM. MaxEnt [12] discovers diverse knowledge sources for NER. Tagging of unknown proper names is done using DT [14], whereas SVM [15] is emphasized on NER as binary classification task. The major disadvantage of ML based NER is that they require large amount of NE annotated training data to acquire higher F-

978-1-4799-4674-7/14/$31.00 ©2014 IEEE

measure. On the other hand, linguistic NER typically uses hand-crafted rules that are manually written by linguists. Such rule-based systems incorporate lexicalized grammar. Lexical grammar includes Part-of-Speech Tagging (POST) [16], gazetteer lists (e.g. dictionaries and available thesauri) [25], and other features such as trigger words, syntax patterns (e.g. word precedence), and other orthographic features (e.g. digit information) etc. One such linguistic system is the NYU system [19]. NYU uses handcrafted rules. In [20], highly sophisticated linguistic analysis is used, and [21] has extensive specialized dictionaries, and white and yellow pages. Also, FASTUS named identification system [22] is based on regular expressions that are carefully handcrafted. FASTUS considers recognition of phrases, patterns and merging incidents. An example pattern is- *"If a proper noun is followed by person's title, then the proper noun is a person's name"*. It is noted that linguistic NER highly rely on manually compiled corpora and manually coded rules. And so, there is an exorbitant cost of development and maintenance of these corpora and respective rules. Furthermore, they are neither domain nor language independent. So, incorporating the data mining and knowledge discovery techniques such as association rules mining for language based NER system is must.

In addition, hybrid NER uses the combination of linguistic and ML based approaches. [23] Uses a hybrid approach to create high precision patterns for NE extraction. In [24], there is a combination of MaxEnt, HMM and handcrafted rules for NER system. In addition, we can see the influence of some external resources of information- namely thesaurus; gazetteers and dictionary for NER system [25]. Both the linguistic approaches [18-19] and ML based approaches [11] [24] used the gazetteer lists. NEs can be classified using gazetteers but impossible to cover all proper names into dictionaries since new proper names keep on generated.

Language based NER includes English NER, Hindi NER etc. English NER [17-18] has 88% - 92% F-measure. However for Hindi NER F-measure is still not at par with English. NER task for Hindi is explored by [26]. They have used language independent morphological and contextual features. Their system has achieved 41.70% F-value, very low recall of 27.84% and about 85% precision for Hindi. More successful Hindi NER system is developed by [27]. They have used CRF with feature induction and have achieved 71.50% F-value on training data of 340k words. In [28] Maximum Entropy Markov Model (MEMM) has given 79.7% F-value. While [29] has used the seed data and annotated corpus to find the lexical pattern learning for Indian languages (Hindi and Bengali NER). Also, genetic algorithm based weighted ensemble [31] is applied on NER for Indian languages (Hindi, Oriya, Telugu etc.) with F-measure as 92.20%, 84.59%, 89.26% respectively.

III. HINDI NAMED ENTITY RECOGNITION CHALLENGES

NER for humans appear to be straightforward as most of the NEs are the proper names. But for a machine to learn and understand NER is too hard especially for Indian languages such as Hindi. Several Hindi NER challenges [28] [30-32] are mentioned here.

1) Rare occurrence of many named entities (NEs) in a corpus.

2) NE Ambiguity.

2.1) Multiple ways of mentioning the same NE.

2.1.1) Variation in Person name: E.g. *"*महात्मा गाँधी", "मोहन दास करम चन्द्र गाँधी", "गाँधी", etc. all refer to the same person.

2.1.2) Variation in Location name: E.g. *"*नई दिल्ली", *"*दिल्ली" both refer to the same location.

2.2) Person s. Location: E.g. *"*बहादुर शाह ज़फर" can refer to a name of a person or a location name.

2.3) Person vs. NNE: E.g. *"*खुशी कक्षा छः मे पढ़ती है।", *"*मुझे आज बहुत खुशी हुई।" Here *"*खुशी" refers to a person and a NNE respectively.

2.4) Location vs. NNE (verb): "गया मे पिंडदान होता है।", *"*क्या वो कल घूमने गया था?" Here "गया" refers to location and a NNE respectively.

2.5) Acronyms: lack of particular standard for forming acronyms in Hindi. E.g. both *"*बीजेपी"and*"*भाजपा" are the acronyms of *"*भारतीय जनता पार्टी"

3) Lack of Capitalization: English language uses capitalization as the discriminating feature for classifying tokens as NEs. On the other hand, Hindi does not have the concept of capitalization at all. For example, person name in English "R. K. Gupta" is represented as *"*आर. के. गुप्ता" in Hindi.

4) Agglutinative property: Agglutinative property makes NE identification even more difficult. For example: *"*इलाहबाद", *"*इलाहबाद को" etc. all refer to the place *"*इलाहबाद", where "को" is a postposition marker in Hindi.

5) Free Word order: Hindi has a relatively free-word order, i.e. the words (NEs) can occupy any place in the sentence. E.g.*"*राम ने रावण को मारा ।", *"*रावण को राम ने मारा ।"

Due to these wide NE variations, machine may fail to identify the presence of many NEs. This means no chance of entity classification and the probabilistic graphical models result in very less recall. So there is a need of an association rules based approach for Hindi NER.

IV. NER USING ASSOCIATION RULES MINING

This section is divided into three sub-sections Section IV.A, Section IV.B and Section IV.C respectively. Section IV.A describes association rules mining (definition and applications). Section IV.B proposes our NER algorithm for Hindi using ARM. Section IV.C discusses mined rules (TYPE 1, TYPE 2 and TYPE 3) with illustrative examples.

A. Association Rules Mining

Association rule mining [33-34] is one of the most important techniques of data mining. Association rules aim to

extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the given databases. Association rules mining has wide applications [35-38] in the areas of medical diagnosis, web caching, query expansion in information retrieval, homeland security, inventory control, market and risk management, telecommunication networks and so on.

An association rule is defined as a relationship of the form: $A => B$, where A *and* B are the sets of items from the given dataset. Here, each association rule is assigned two factors: support factor and confidence factor. Support factor is the ratio of the total number of items in both A and B over the total number of items in the database. Confidence factor is the ratio of the total number of items in both A and B over the total number items in A. In other words,

$$Support = \frac{transcations\ having\ items\ A\ and\ B\ together}{total\ no.\ of\ transcations\ in\ the\ datbase}$$

$$Confidence = \frac{transcations\ having\ items\ A\ and\ B\ together}{transcations\ having\ item\ A\ in\ the\ datbase}$$

Mining association rules means extraction of rules where Support (S) and Confidence (C) are greater than (or equal to) user-specified (or pre-defined) Minimum Support (MSup) and Minimum Confidence (MConf) in the database.

B. Proposed Hindi NER Algorithm

This section represents our proposed NER algorithm (please refer Fig. 1) for Hindi using ARM.

STEP 1: START
STEP 2: SET $FG = 0$
STEP 3: FOR every pair of terms $<term_1, term_2>$
FIND the Set S_R of Rules $A => nmcl$
s.t. $<term_1, term_2>$ MATCH A
(WHERE $S >= MSup$ AND $C >= MConf$)
GO TO Step 4
STEP 4: IF (S_R is not *NULL*) THEN
CHOOSE in S_R, Rule $A => nmcl_A$
$nmcl_A$ with highest confidence among rules
Rules *TYPES: 1* OR (AND) *2* OR (AND) *3*
AND SET $FG = 1$
ELSE
Name-Class ($term_2$) = *Not-Name* AND $FG = 0$
STEP 5: IF $FG = 1$THEN
Name-Class ($term_2$) = $nmcl_A$ AND $FG = 0$
STEP 6: REPEAT STEPS 3 TO 5 UNTIL LOOP ENDS
STEP 7: STOP

Fig. 1. Hindi NER algorithm.

C. Mining Association Rules for Hindi NER

Named entity recognition is the NE extraction task from the datasets. Datasets are the nothing but the given text documents. Documents are nothing but the sequences of terms. And, the occurrences of terms are defined as items. Each item has some feature value and some name class. However, the sets of items A and B be defined as terms, sequences of terms, features and NE classes. Then we can say that B is an NE class that we wish to predict.

Let us see some terminology that we use in our algorithm. $<term_1, term_2>$ be defined as the sequence of terms, $feat_2$ as the feature of $term_2$, $nmcl_2$ as the name class of $term_2$, S is support and C is confidence. After series of manual computations and informal experiments, we consider three baseline rules for A *with* MSup = 30% and MConf = 80%. Association rules are stated in the form of TYPE 1, TYPE 2 and TYPE 3 rules (please refer Fig. 2). Among the mined rules, we observe that rules with highest confidence can significantly increase the system performance.

- TYPE 1: TYPE 1 rule is a dictionary rule i.e.
  $<term_2>$ => $nmcl_2$, (S and C)
- TYPE 2: TYPE 2 rule is a bi-gram rule i.e.
  $<term_1, term_2>$ => $nmcl_2$, (S and C)
- TYPE 3: TYPE 3 rule is a feature rule i.e.
  $<term_1, feat_2>$ => $nmcl_2$, (S and C)

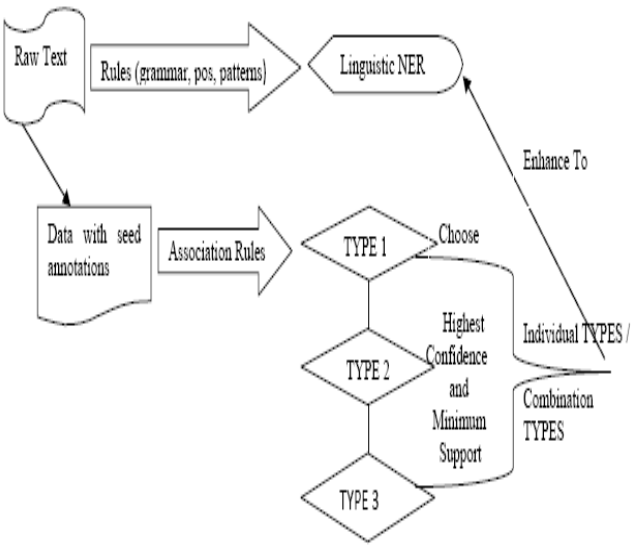

Fig. 2. NER system using association rules.

We reconsider an example sentence: *"प्रधानमंत्री मनमोहन सिंह को अपने साउथ ब्लाक पीएमो कार्यालय से भावभीनी विदाई मिली।"* Named classes are given in the training corpus where an annotation symbolizes the term *"मनमोहन सिंह"* as person named class.

- Generation of TYPE 1 dictionary rule is of the form <*"मनमोहन सिंह"*> => $nmcl_2$ or *NE class as person* (*"मनमोहन सिंह"*), where *S* depends upon number of occurrences of *"मनमोहन सिंह"* as tem, and *C* depends upon number of occurrences of *"मनमोहन सिंह" as* labeled term in the person NE class.
- Generation of TYPE 2 bi-gram rule is of the form <*"प्रधानमंत्री", "मनमोहन सिंह"*> => $nmcl_2$ or NE class as person (*"मनमोहन सिंह"*), where *S* depends upon number of occurrences of "*प्रधानमंत्री मनमोहन सिंह"* as expression (or sequence of terms), and *C* depends upon number of occurrences of "*मनमोहन सिंह" as* labeled term in the person NE class.
- Generation of TYPE 3 feature rule is of the form < *"प्रधानमंत्री", sur_per*(P_M)> => $nmcl_2$ or NE class as person (P_M) where *S* depends upon number of occurrences of *"प्रधानमंत्री* P_M*"* as sequence of terms, and *C* depends upon number of occurrences of the term P_M labeled in the feature *sur_per*(P_M) and person NE class (*sur_per()* is defined as surname of a person, i.e. *sur_per("मनमोहन") = "सिंह"* means *"सिंह"* is a surname of the person *"मनमोहन"*).

V. Experimentation And Results

We obtain data from (http://en.wikipedia.org/wiki/List_of_newspapers_in_India_by_readership) three leading Hindi newspapers-*"दैनिक जागरण"* (Dainik Jagran), "हिन्दुस्तान" (Hindustan) "दैनिक भास्कर" (Dainik Bhaskar) dated 15th March 2014 to 15th May 2014. Our corpus consists of variety of news articles in the Hindi language. Our training set consists of 100 articles for learning the association rules and testing set consists of another 50 articles. Our algorithm is implemented in the Java based Python code with version 3.3.0 (www.nltk.org).

We use two evaluation measures- Precision (P) and Recall (R). These measures are defined as follows.

$$P = \frac{no.\ of\ correct\ responses}{total\ no.\ of\ responses}$$

$$R = \frac{no.\ of\ correct\ responses}{total\ no.\ of\ answers}$$

Where, response is a term with NE class as labeled by our algorithm while answer is term with NE class as labeled by the corpus. Response is stated to be correct only of it matches with an answer.

TABLE I. Hindi NER Results

Rule TYPE(s)	Rule Name	Evaluation Measures	
		P(%)	R (%)
TYPE 2 only	bi-gram	84.76	47.37
TYPE 3 only	feature	75.64	34.63
TYPE 2 and TYPE 1	bi-gram + dictionary	91.10	67.89
TYPE 2 and TYPE 3	bi-gram + feature	89.72	57.92

We note that TYPE 2 (bi-gram rules) can be combined with TYPE 1(dictionary rules) or with TYPE 3 (feature rules) to significantly increase the performance of NER for Hindi using ARM (please refer Table I). Also, if for some term no rule is observed then NNE is assigned to that term.

VI. Conclusion

We present a state-of-art baseline association rules mining for named entity recognition in Hindi language. We consider three types of rules- TYPE 1 is dictionary based, TYPE 2 is bi-gram based, and TYPE 3 is feature based. We consider corpus of news articles from three leading Hindi newspapers from 15th March to 15th May 2014 and coded in Python version 3.3.0. Minimum support and confidence values are taken as 30% and 80% respectively. Rules with high confidence values are observed as of TYPE 2 in combination with TYPE 1 or TYPE 3 which significantly improve the performance of NER system. But our system suffers from low recall. In future, we can incorporate some more association rules to increase recall.

References

[1] Chinchor, N., "MUC-7 Named Entity Task Definition, version 3.5, 17", Proceedings of the Seventh Message Understanding Conference (MUC-7), Morgan Kaufmann Publishers, September 1997.

[2] Bandyopadhyay, S., Naskar, S. K., Ekbal, A., "Emerging Applications of Natural Language Processing: Concepts and New Research", IGI Global, Information Science Reference, 2013.

[3] Greenwood, M. A. and Gaizauskas, R., "Using a Named Entity Tagger to Generalize Surface Matching Text Patterns for Question Answering", In EACL03: 10th Conference of the European Chapter of the Association for Computational Linguistics, Budapest, Hungary, 2003.

[4] Toral, A., Llopis, F., Munoz, R., and Noguera, E., "Reducing Question Answering Input Data using Named Entity Recognition", In Proc. 8th International Conference on Text, Speech & Dialogue, 2005.

[5] Molla, D., Zaanen, M., and Smith, D., "Named Entity Recognition for Question Answering", In Proc. ALTW 2006.

[6] Babych, B., Hartley, A., and Atwell, E.,"Statistical Modelling of MT output corpora for Information Extraction", In Proc. Corpus Linguistics conference, Lancaster University (UK), pp. 62-70, 28 - 31 March 2003.

[7] Tsai, R. T. H., Sung, C. H., Dai H. J., Hung, H. C., Sung,T. Y., and Hsu, W. L., "NERBio: Using Selected Word Conjunction, Term Normalization, and Global Patterns to Improve Biomedical Named Entity Recognition", BMC Bioinformatics, 7(Suppl 5):S11, 2006.

[8] Zhou, G., Zhang, J., Su, J., Shen, D., and Tan, C., "Recognizing Names in Biomedical Texts: a Machine Learning Approach", Bioinformatics, vol. 20, no. 7, pp. 1178- 1190, 2004.

[9] Sikdar, U. K., Ekbal, A., Saha, S., "Modified Differential Evolution for Biomedical Name Recognizer", Computational Linguistics and Intelligent Text Processing, pp. 225-236, 2014.

[10] Bikel, D. M., Miller, S., Schwartz, R., and Weischedel, R., "Nymble: a High Performance Learning Name Finder", In Proc. ANLP-97, pp. 194–201, 1997.

[11] Borthwick, A., Sterling, J., Agichtein, E., and Grishman, R., "Exploiting Diverse Knowledge Sources via Maximum Entropy in Named Entity Recognition", In Proc: Sixth Workshop on Very Large Corpora, New Brunswick, New Jersey, Association for Computational Linguistics, 1998.

[12] Curran, J. R. and Clark, S., "Language independent NER using a Maximum Entropy Tagger", In Daelemans,W. and Osborne, M., editors, Proc: CoNLL- 2003, Edmonton, Canada, 2003.

[13] Peng, F., Feng, F., and McCallum, A., "Chinese Segmentation and New Word Detection using Conditional Random Fields", COLING, 2004.

[14] Bechet F., Nasr A., Genet F., "Tagging Unknown Proper Names Using Decision Trees", In Proc. 38 ACL Conference, Hong- Kong, pp. 77-84, 2000.

[15] Wu. T. F., Lin, C. J., Weng, R. C., "Probability Estimates for Multi-class Classification by Pair-wise Coupling", The Journal of Machine Learning Research, vol. 5, pp.975-1005, 2004.

[16] Gupta, J. P., Tayal, D. K. and Gupta, A., "A TENGRAM Method Based Part-of-Speech Tagging of Multi-Category Words in Hindi Language", Expert Systems With Applications, Elsevier Science Publication, vol. 38, no. 12, pp. 15084-15093, Nov-Dec 2011

[17] McDonald, D. D., "Internal and External Evidence in the Identification and Semantic Categorization of Proper Names, Corpus Processing for Lexical Acquisition", MIT Press, Cambridge, MA, 1996.

[18] Wakao T., Gaizauskas R. and Wilks Y., "Evaluation of an Algorithm for the Recognition and Classification of Proper Names", In Proc: COLING-96, 1996.

[19] Grishman,R., " The NYU System for MUC-6 or Where's the Syntax?", In Proc. Sixth Message Understanding Conference (MUC-6), Columbia, MD, November 1995.

[20] Burger, J. D., Henderson, J. C., and Morgan, W. T., " Statistical Named Entity Recognizer Adaptation", CoNLL, Taipei, Taiwan, pp. 163-166, 2002.

[21] Iwanska, L., Croll, M., Yoon, T., and Adams, M., "Wayne state university: Description of the UNO Natural Language Processing System as used for MUC-6", In Proc. Sixth Message Understanding Conference (MUC-6), Columbia, Morgan-Kaufmann Publishers, 1995.

[22] Hobbs, J., Appelt, D., Bear, J., Israel, D., Kameyama, M., Stickel, M., and Tyson, M., "FASTUS: Extracting Information from Natural Language Texts", In E. Roche and Y. Schabes, editors, Finite State Devices for Natural Language Processing, MIT Press, 1996.

[23] Talukdar, P. P., Brants, T., Liberman, M., and Pereira, F., "A Context Pattern Induction Method for Named Entity Extraction", In Proc: Tenth Conference on Computational Natural Language Learning (CoNLL), 2006.

[24] Srihari, R., Niu, C. and Li, W., " A Hybrid Approach for Named Entity and Sub-Type Tagging", In Proc. Sixth conference on Applied Natural Language Processing, 2000.

[25] Mikheev, A., Moens, M., and Grover, C., "Named Entity Recognition with Gazetteers", In Proc: EACL, Bergen, Norway, 1999.

[26] Cucerzan S., and Yarowsky, D., "Language Independent Named Entity Recognition Combining Morphological and Contextual Evidence", In Proc. Joint SIGDAT Conferenceon EMNLP and VLC, pp.90-99, 1999.

[27] Li, W., and McCallum, A., "Rapid Development of Hindi Named Entity Recognition using Conditional Random Fields and Feature Induction", ACM Transactions on Computational Logic, 2004.

[28] Kumar, N., and Bhattacharyya, P., "Named Entity Recognition in Hindi using MEMM", Technical Report, IIT Bombay, India, 2006.

[29] Ekbal, A., and Bandyopadhyay, S., "Lexical Pattern Learning from Corpus Data for Named Entity Recognition", In Proc. International Conference on Natural Language Processing (ICON), 2007.

[30] Ekbal, A., Haque, R., Das, A., Poka, V., and Bandyopadhyay, S., " Language Independent Named Entity Recognition in Indian Languages", In Proc: IJCNLP workshop on NERSSEAL, 2008.

[31] Ekbal, A., and Saha, S., " Weighted Vote Based Classifier Ensemble for Named Entity Recognition: A Genetic Algorithm Based Approach", ACM Transactions on Asian Language Information Processing, 2011.

[32] Ekbal, A., Saha, S., Singh, S., "Active Machine Learning Technique for Named Entity Recognition", In Proc. International Conference on Advances in Computing, 2012.

[33] Agrawal, R., Imielinski, T., and Swami, A. N., "Mining Association Rules between Sets of Items in Large Databases", In Proc: ACM SIGMOD International Conference on Management of Data, pp. 207–216, May, 1993.

[34] Budi, I., and Bressan, S., "Association Rules Mining for Name Entity Recognition", WISE, pp. 325-328, 2003.

[35] Liu, B., Ma, Y., Wong, C. K., and Yu, P. S., "Scoring the Data Using Association Rules", Applied Intelligence, vol. 18, no. 2, pp. 119-135, 2003.

[36] Wu, X., Zhang, C., and Zhang, S., "Efficient Mining of Both Positive and Negative Association Rules", ACM Transactions on Information Systems, Vol. 22, no. 3, pp. 381-405, July, 2004.

[37] Budi, I., Bressan, S., Wahyudi, G., Hasibuan, Z. A., and Nazief, B., "Named Entity Recognition for the Indonesian Language: Combining Contextual, Morphological and Part-of-Speech Features into a Knowledge Engineering Approach", Discovery Science, pp. 57-69, 2005.

[38] Kotsiantis, S. and Kanellopoulos, D., "Association Rules Mining: A Recent Overview", GESTS International Transactions on Computer Science and Engineering, vol. 32, no. 1, pp. 71-82, 2006.