ISSN: 0000-0000 (print), 0000-0000 (online)

AIMS Journal

Journal homepage: www.jdconline.net/aims

Regular Paper

Intelligent Business Rules Analysis using Named Entity Recognition

Uzmah Irshad[a], Rafaqut Kazmi[b]

[a]*University Technology Malaysia, Johar, Malaysia*

ARTICLE INFO

Article history:
Received 7 February, 2017
Revised 19 April 2017
Accepted 30 April, 2017

Keywords:
Business Process
Business Rules
Named Entity Recognition

ABSTRACT

This study investigates that the analysis phase in business process modelling is a major aspect due to its impact on a whole business process modelling and running phase. The Analysis phase is difficult and complex, one reason of difficulty of business process model is absence of background knowledge required to completely understand a business requirement. Since, business requirements are gathered from multiple business stakeholders, various aspects and angels are covered by various stakeholders. Here a business analyst can perform a better analysis of a business process if a background information is also an accompanied with business requirements. However, this is not a typical practice simple business requirements are handed over to a business analyst per analysis purposes. Such scenario makes nothing more difficult for the business analyst. Here, either a business analyst has to interview the stakeholders again and again for the sake of getting clarity on intermittent business requirements or to take blind design in finalizing meanings of the vocabulary used in business requirements. In this paper, we present the results of the study for attaching shared meanings with the used vocabulary that will not only help in simplifying the analysis phase of business process modelling but also may provide more accuracy.

1. Introduction

Named Entity Recognition is also defined as entity identification, entity pieces and entity extraction. Named entities are the categories: such as the name of a person, locations, organizations, expression of the times, quantities, financial values, percentages, etc. Named Entity Recognition (NER) is the mission of progressing text to recognize and arrange names, an essential part in various Natural Language Processing (NLP) operations, permissive the taking out of valuable information from instruction [1]. Name Entity Recognition is a main subtask of information origin. MUC-7 classifies named entities into the following categories and subcategories these are [2]:

- Entity: Person, organization, location
- Time expression: date and time.
- Numeric expression: money, percent.

A set of business processes is a complimentary part of a business enterprise. Analysis of a business process is a major phase in the development of a business process model [3] The modelling of a business process is one of the techniques used for understanding that how a particular business process works and how individuals from different work-groups interact with each other. A typical business process model is alternate of the use cases typically used in software engineering to depict the way a user can interact with a system and what functionality the user can perform [4]. However, for technology focused business personals, a business process model can be more helpful in understanding the behaviour of a business process. In simple words, a business process model is a step by step illustration of the way different business users achieve a specific goal [5]. Here, creation of a business process model is a formal procedure in typical business process modelling activity. To create a business process model, a business analyst should be aware of the scope of the process and its inputs and outputs. The main segment of business process modelling is description of an activity or flow of a process. To enforce a typical business policy, a set of business rules is used in combination with the business process models. Often, a workflow diagram is also drawn for the sake of visual model that shows the basic steps of a particular activity and possible exception as well.

* *Corresponding author.* Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: rafaqutkazmi@gmail.com
http://dx.doi.org/.....

In the development of a business process, analysis of a business process is a premier and a significant phase [6]. In a typical business process analysis phase, it is examining that how a set of inputs is transformed into a set of complacent outputs. Since, such business processes significantly impact the performance of a business activity, the phase of business process analysis becomes more significant. A business process analysis phase is not only employed for designing a new business process, but also used for improvement of the existing business processes in an iterative manner. However, such a business process analysis phase is a challenging task in the design of a complete, accurate and robust business process model. Analysis of business process requirements becomes more difficult, especially dealing with a set of complex business requirements. In a typical analysis of a business process, a business analyst has to thoroughly understanding activities; apprehend the relationship among the intermittent activities, etc. Once, basic information regarding a business analyst needs to define boundary of a process along with the inputs and outputs of the process [7]. Finally, a process flow diagram is designed to represent the activities of a process along with the relationships.

1.1. NER for Business Rule Analysis

In the previous section, it is discussed that the analysis phase in business process modelling is a major aspect due to its impact on a whole business process modelling and running phase. The analysis phase is difficult and complex. One reason of difficulty of business process model is the absence of background knowledge required to completely understand a business requirement [8]. Since, business requirements are gathered from multiple business stakeholders, various aspects and angels are covered by various stakeholders. Here, a business analyst can perform a better analysis of a business process if a background information is also an accompanied with business requirements. However, this is not a typical practice simple business requirements are handed over to a business analyst per analysis purposes. Such scenario makes nothing more difficult for the business analyst. Here, either a business analyst has to interview the stakeholders again and again for the sake of getting clarity on intermittent business requirements or to take blind design in finalizing meanings of the vocabulary used in business requirements [9].

In the related work discussed in chapter 2, it has been found that no significant research work has been done to attach background information with the used vocabulary in business rules. To address the same problem, a standard named Semantics of Business Vocabulary and Rules (SBVR) [1] were introduced by the OMG in 2008. Though, SBVR provides a concept of shared meanings for the used vocabulary so that all stakeholders may consider single meaning of vocabulary. However, associating such shared meanings with the used vocabulary is in itself a challenging task.

2. Related Work

NER approach was used in text corpora. In text n-gram model was used to derive contexts [6]. NE was also used in information extraction. NER can be name of person, place and organization. Some task of NER was to retrieve correct personal names in documents, translation of NE'S from one language to another language. It was also used in clustering, string matching, database entries mapping, database cleaning and data mining. Two methods were used in NER. One was Semi-supervised that is a recent term called "bootstrapping" bootstrapping means to start a business with stretching resources. Second was UN Supervised in this method, clustering was used.

A semi supervised method was presented for the classification of business entities [7]. There was a huge amount of data shared on the internet by users there for organizations would like to explore and extract Business intelligence. Analysis of this data can improve marketing effectiveness and organization, business decisions making. Tools were available for named entity recognition (NER) but still there were problems for identification of organization names, co-references and abbreviation. This claimants that their proposed methods were more efficient than other methods and nominal human interference was involved [8].

NER was discussed in common research in information extraction was to propose a system to origin stable types of information from papers in a limited speech and domain (Maynard et al., 2011). This system told about name entity recognition to extend suitability for end user application by using dispute field, arrange and type However, in many applications the type of file and field may be unidentified, or an organization may be demanded, which procedure was divide types of papers without the necessity for tuning [9].

The Semi-supervised examination was studied to name entities, related entities, and prepositional expression attachment with a tell-web framework [10]. The meaning of an entity was invented and updated faster in the internet globe than printed relation. Web provided a lot of information and data. Billion web pages were used to read automatically by web pages [11]. To access, fast web data Google and yahoo web search engine were used. BOW, was designed to capture language distributions of any kind of NLP terms on the web, such as named entities, related tuples, and expression. Different methods were composed a BOW vector to illustrate and was expanded smoothly to designate the semantic object of other language elements [12]. For high accuracy automatically generated- training data were required. Automatically join named entities (NE) with high accuracy and remember claim a huge amount of skill-glossy data, which was costly to gain. Same NEs required distinct domains mainly need distinct tag data [13-16].

3. Materials and Method

In the business rules analysis phase, it is explained that the business goals of a deployment and the situation the business requirements that obtained those goals. For starting the business requirements, business constraints are examining that might act on the ability to obtain the business goal. Without particular business analysis, imperfect explanations are discussed. During business analysis phase business requirements are composes. There are some Instructions that you will later use as input to the technical requirements phase.

3.1. Business Rules specification

Business Process is the steps that perform to create value for customers. A business process has three components; these are activities, inputs and outputs. In this process input may include materials, labour, energy, and controlling equipment, etc. and outputs may be any physical product.

The process specification is the process that explains how input change into outputs and what must be done in direction to change input into products. It is a set of instruction that outlining a business progress in which each initial level business activity is hoping to move out. In a system growth, process specifications are usually included as a whole component of the requirement.

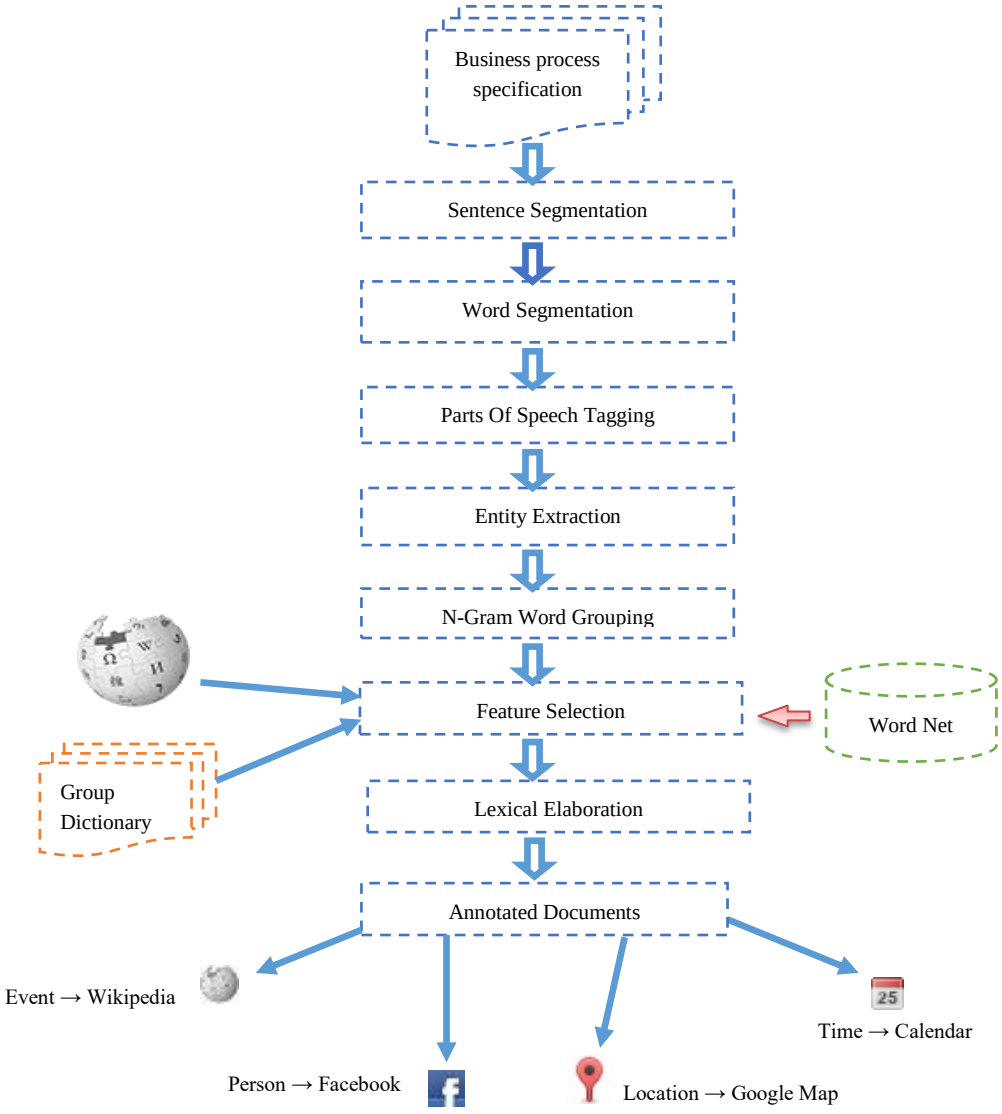

Fig 1. Used Framework for Business Rules Analysis

3.2. Sentence Segmentation

Dividing something into smaller unit is called segmentation. Sentence Segmentation is the procedure that decides the longer progress units consisting of one or more words. This task is distinguished sentence boundaries into words indistinct sentences. When procedure simple text, tables of abbreviations that holding periods can serve stop faulty assignment of sentence boundaries as shown in Table 1.

Table 1: Sentence segmentation

Input	The administrator of a department must belong to that department.
Sentence segmentation	The administrator of a department must belong to that department.

3.3. Word Segmentation

Word division is the problem of separating a string of written language into its integral words. Word segmentation is the procedure of distinguishing the different boundaries between words, articulate, or phonemes in spoken natural languages. In this process intellectual process is used by human and artificial procedure for essential language procedures. Breaking up the sequence of character in a text by placing the word boundaries are called tokenization, the steps where one word terminates and other start. An example of word segmentation is shown in Table 2.

Table 2: Word Segmentation

Input	The administrator of a department must belong to that department.
Word segmentation	[The] [Administrator] [of] [a] [department] [must] [to] [that] [department] [.]

3.4. PoS Tagging

The PoS tagging is the problem that determines the POS tag for a specific instance of a word. There are 8 parts of speech that introduced by Thrax of Alexandria (c 100 BCE). For PoS tagging, the Java API of Stanford PoS tagger is used that tags with accuracy about 97%. A sample output of PoS tagging is shown in Table 3.

Table 3: POS Tagging

Input	The administrator of a department must belong to that department.
POS Tagging	[The/ DT] [administrator/NN] [of/ IN] [a/DT] [department/NN] [must/MD][Belong/VB] [to/TO] [that/DT] [department/NN]./.

3.5. Entity Extraction

Procedure of automatically origin document metadata from unstructured text documents called Entity Extraction. Person, locations, title, dates, are extracting entities .These entities such as location, dates, specialized terms and production terminology from free-form text can enable organizations to not only correct keyword inquire, but also open the passage to semantic search, faceted search and document re purposing. The entity extraction procedure can be gracefully customized to be watchful of single objects in your organization. The Entity Extraction software supports grammar-supported techniques as well as statistical standard. Example is given bellow.

Table 4: Entity Extraction

Example	The administrator of a department must belong to that department.
Entity extraction	administrator, department, belong

In the example shown in Table 4, there are three entities. In which Administrator is a Person, Department is a place and belong is a verb

3.6. N-Gram Word Grouping

N-gram is a neighbouring of n items from a given arrangement of text or language. The items can be phonemes, syllables, letters, term or low-minded set according to the application. An n-gram design is a style of probabilistic language design for predicting the next item in such a sequence in the form of an arrangement Markov design. N-gram models are extensively used in statistical natural language processing. In speech recognition, the phonemes and sequences of phonemes are modelled using a n-gram arrangement. For example

Table 5: N-Gram Word Grouping

Example	EU-Rent has a loyalty club.
N-Gram word grouping	[EU-Rent]

In the example discussed in Table 5, N-gram word grouping is shown. EU-Rent has a loyalty club in this sentence EU-Rent has two syllables that is joined by [-].

3.7. Lexical Elaboration

A simple case study of narratives generates two set in sequence. Lexical elaboration was selected for these two purposes are: relationship between the text modifications and vocabulary gain. The other object is that it is painful to prove the result of modification of other linguistic terms, such as syntax, at the sentence level. It seems hopeless to produce a comparable syntactically shorten and elaborated sentence set.

3.8. Feature Selection

Feature Selection is also assumed as attribute selection, variable selection. The procedure of choosing a subset of relevant characteristics is called Feature selection and it is used for designing construction. In this process main assumption is that the data may include many redundant or unrelated features. The redundant feature cannot supply more information only common feature chooses; unrelated features are those that supply no beneficial instruction in any context. Feature Selection is a method that replaces complicated classifier into simpler ones. Feature Selection Algorithm has three classes these are: wrapper methods filter methods and embedded methods.

3.9. Annotated Documents

The Comments, Notes, explanation, or other symbol of exterior remarks is called Annotation. These remarks can be attached to a web document. It is possible to annotate any web document independently, without editing these documents because these are external. Annotation is mainly seen as metadata because they give additional information around an existing part of the data. An annotation can be stored locally. An annotation has many properties contain: Physical places; an annotation that stored regional file system. The other is scope: it combined to a whole document or exact to a fragment. Last one is Annotation type: Annotation, any query and comment are included.

4. Experiments and Results

A set of examples of SBVR were taken such as EU-Rent is a famous case study given in SBVR 1.0 [1] document. 10 sentences are used in this case study. The largest sentence that was used in this case study has 32 words and smallest sentence have 4 words in this case study. Correct, Incorrect and missing words are shown below. The individual is the synonyms of person that cannot be picked up. The group is an adjective that is missing and not be picked up. Day and model are characteristics that have been picked as an object. Branch and model are organized, but picked up as a person. One verb is start that has been picked as characteristics specify is also a verb that is missing. Accumulate is a verb that is picked up as an organization.

Table 6: Results of the business to named entity translated by Tool.

Sr #	Types	N_{Sample}	$N_{Correct}$	$N_{Incorrect}$	$N_{Missing}$
1	Person	10	6	3	1
2	Place	1	1	0	0
3	Object	9	7	2	0
4	Characteristics	2	1	1	0
5	Organization	7	6	1	0
6	Verb	8	7	0	1
7	Supporting fact type	8	8	0	0
8	Adjective	13	12	0	1
9	Currency	1	1	0	0
	Total	59	49	7	3

The Table 6 consists of 10-person type, place 1, object type 9, Characteristics 2, Organization 7, verb types 8, Supporting fact type 8, Adjective 13 and currency type is 1 only. Separate output is also showed. But the total number of samples is 59, in which 49 a correct and 7 are incorrect, and 3 are missing.

In second case study, 14 sentences are used in this case study. The largest sentence that was used in this case study has 28 words and smallest sentence have 8 words in this case study. Correct, Incorrect and missing words are shown in this table. One verb type is proposed that is missed. Receive is a verb that has been picked as an object. The update is also a verb picked as an adjective. Or is a supporting type fact that cannot be picked.

Table 7: Results of the business to the named entity is Translated by Tool

Sr #	Types	N_{Sample}	$N_{Correct}$	$N_{Incorrect}$	$N_{Missing}$
1	Person	10	9	1	0
2	Place	6	6	0	0
3	Object	9	8	1	0
4	Organization	3	3	0	0
5	Verb	22	21	0	1
6	Time	1	1	0	0
7	Adjective	5	4	1	0
8	Supporting fact type	10	9	0	1
	Total	66	61	3	2

This table consist of 10 person type, place 6; object type 9, time type is 1, Organization 3, verb types 22, Supporting fact type 10, Adjective 5. Separate output is also showed. But total numbers of samples are 66, in which 61 correct and 3 are incorrect, and 2 are missing.

Table 8: Recall and precision for case study 1 and case study 2 is also given in this table.

Types	N_{Sample}	$N_{Correct}$	$N_{Incorrect}$	$N_{Missing}$	Rec%	Pre%

Business rule 1	59	49	7	3	83.00	87.50
Example 2	66	61	3	2	92.42	95.31

In above table, in example 1 Recall and Precision is calculated. In this table, total samples are 59, correct 49 and 7 are incorrect and 3 are missing. In this table, recall is calculated 83 % and precision is calculated 87.5%. In example 2 total sample are 66, correct are 61, incorrect 3, missing are 2. Recall is also calculated, that is 92.42%. Precision is 95.31%. In this table after calculating recall and precision, f-value is also calculated

Table 9: F-value for case study 1 and case study 2 is also given in this table.

Types	Rec%	Pre%	F-value
Business Rule set 1	83.00	87.50	85.10
Business Rule set 2	92.42	95.31	93.84

In case study 1 and 2 F-values are calculated. In first example F-value is 85.10. For other case study F-value is 93.84.

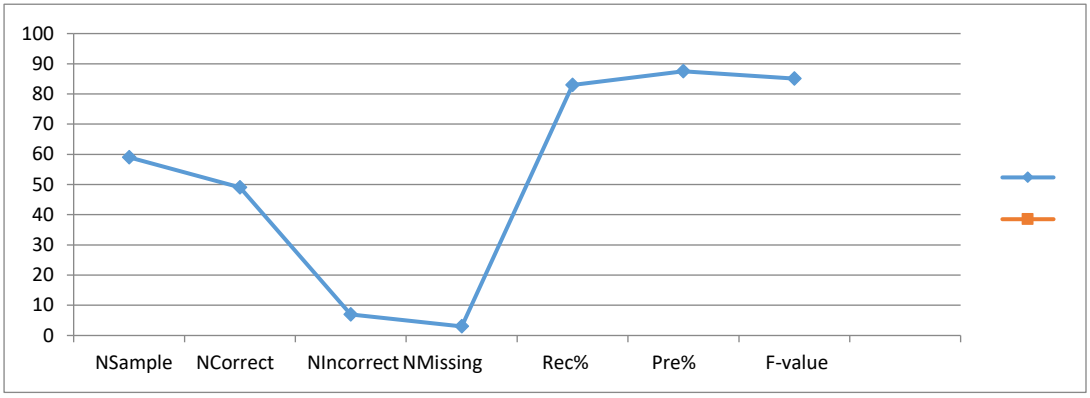

Fig 2. Evaluation of recall, precision and F-value

This graph display the results of total samples, correct, incorrect, and missing, recall, precision and F-values. There values are showed in this graph: samples are 59, correct 49 and 7 are incorrect and 3 are missing. In this graph, recall is 83 % and precision is 87.5%. F-value is 85.1 that showed in the graph a colour line.

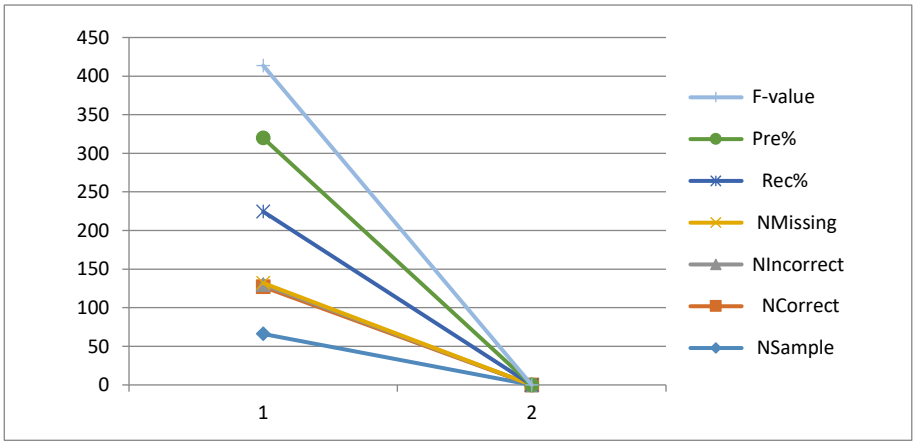

Fig 3. Evaluation of recall, precision and F-value

This graph displays the results of total samples, correct, incorrect, and missing, recall, precision, and F-values. There values are showed in this graph: samples are 66, correct 61 and 3 are incorrect and 2 are missing. In this graph, recall is 92.42 % and precision is 95.31%. F-value is 93.84 that showed in graph a colour line. All values are shown in graph after this evaluation we will evaluate both case studies by a graph.

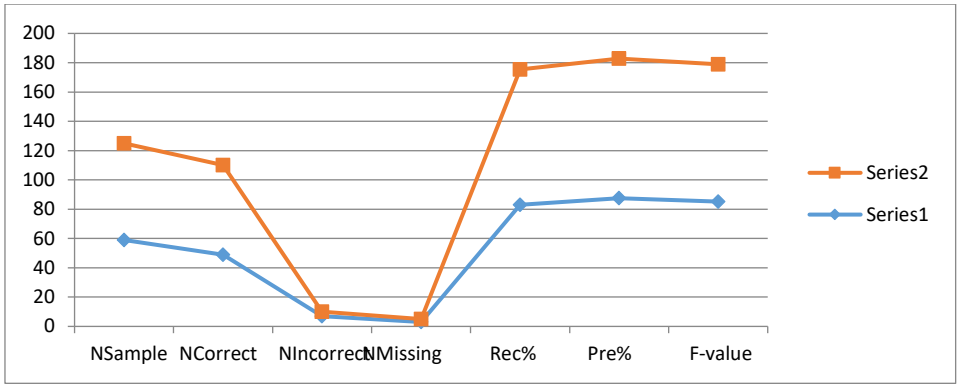

Fig 4. Results of both examples

According to our results demonstrate that Recall of case study 1 is 83%, Precision 87.5 % and F-value is 85.1 for case study 2 results of Recall 92.42%, Precision is 95.31% and F-value is 93.84. After these results these values are evaluated by a graph .It also shows this approach is easy and time saving.

5. Conclusion and Future Work

The main objective of this research work was to design a new business process and improve existing business process by providing background information that relates to business requirements. In this way a business analyst can perform a better analysis of a business process. Shared meaning is also used for vocabulary identification. During business analysis phase business requirements are composes. In our framework different approaches are discussed that is involved word segmentation, sentence segmentation, part of speech tagging, lemmatization and n-gram word grouping. Our tool translates lexical analysis of business rule and named entity recognition business rule and evaluate the results. For this purpose 2 case studies are taken and have made successful experiments. After Part of speech tagging interpretation named entity recognition, classification is performed. We performed an analysis of syntactic translation that convert it into English text and get vocabulary using inputs person, place, verb, adjective, characteristics, organization, time and supporting fact type. The result is shown in result section, Recall of case study 1 is 83%, Precision 87.5 % and F-value is 85.1 for case study 2 results of Recall 92.42%, Precision is 95.31% and F-value is 93.84. After these results these values are evaluated by a graph .It also shows this approach is easy and time saving.

REFERENCES

[1] Baisley, D. E., Hall, J., & Chapin, D. (2005). Semantic Formulations in SBVR. In Rule Languages for Interoperability.
[2] Black, William J.; Rinaldi, F.; Mowatt, D. 1998. Facile: Description of the NE System used for Muc-7. In Proc. Message Understanding Conference.
[3] Borthwick, A. (1999). A maximum entropy approach to named entity recognition (Doctoral dissertation, New York University).
[4] Daland, R. (2009). Word segmentation, word recognition, and word learning: A computational model of first language acquisition (Doctoral dissertation, Northwestern University).
[5] Dingare, S., Nissim, M., Finkel, J., Manning, C., & Grover, C. (2005). A system for identifying named entities in biomedical text: How results from two evaluations reflect on both the system and the evaluations. Comparative and Functional Genomics, 6(1-2), 77-85
[6] Ekbal, A., & Bandyopadhyay, S. (2010). Named entity recognition using appropriate unlabeled data, post-processing and voting. Informatica (Slovenia), 34 (1), 55-76.
[7] GU, B. (2008). Recognizing named entities in biomedical texts (Doctoral dissertation, School of Computing Science-Simon Fraser University).
[8] Guo, J., Xu, G., Cheng, X., & Li, H. (2009, July). Named entity recognition in the query. In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval (pp. 267-274). ACM.
[9] Kamada, A., Governatori, G., & Sadiq, S. W. (2010). SBVR based Business Contract and Business Rule IDE. In RuleML Challenge.
[10] Kuksa, P. P., & Qi, Y. (2010, April). Semi-supervised Bio-named Entity Recognition with Word-Codebook Learning. In SDM (pp. 25-36).
[11] Li, C., Weng, J., He, Q., Yao, Y., Datta, A., Sun, A., & Lee, B. S. (2012, August). Twiner: named entity recognition in targeted twitter stream. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (pp. 721-730). ACM.
[12] Mayfield, J., McNamee, P., & Piatko, C. (2003, May). Named entity recognition using hundreds of thousands of features. In Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4 (pp. 184-187). Association for Computational Linguistics.
[13] Ritter, A., Clark, S., & Etzioni, O. (2011, July). Named entity recognition in tweets: an experimental study. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 1524-1534). Association for Computational Linguistics.
[14] Send, A., Agrawal, A. J., & Kakde, O. G. (2012). Domain Specific Named Entity Recognition Using Supervised Approach. International Journal of Computational Linguistics (IJCL), 3 (1), 67-78.
[15] Wang, M., Che, W., & Manning, C. D. (2013, June). Effective Bilingual Constraints for Semi-Supervised Learning of Named Entity Recognizers. InAAAI.
[16] Wang, S. (2009). Data Analysis Project: Semi-Supervised Discovery of Named Entities and Relations from the Web.

[17] Zhou, G., & Su, J. (2002, July). Named entity recognition using an HMM-based chunk tagger. In proceedings of the 40th Annual Meeting on Association for Computational Linguistics (pp. 473-480). Association for Computational Linguistics.

[18] Zhu, J., Uren, V., & Motta, E. (2005). ESpotter: Adaptive named entity recognition for web browsing. In Professional Knowledge Management (pp. 518-529). Springer Berlin Heidelberg.