Study of NER & Developed System for Development of NER System

Manisha, Mrs. Preethi Jose Dolly

Abstract— **This Paper gives an introduction regarding the Development of Named Entity Recognition (NER) using some approach and implementation using PHP (a Server Side Scripting Language) & MySQL. The developed System will identify different words under different NER Category, i.e Named Entity Tagset by which we will be able to extract data from the system up to great extent. If we will input any sentence into the input box then it break the sentence into words and that word will be categorized into different category.**

Index Terms— **PHP , MySQL , Named Entity Tagset , NER.**

I. INTRODUCTION

Named Entities Recognition (**NER**) means that it involves identification of proper names. Three Universally accepted categories are the names of person, locations and organization etc. Named entity recognition is widely used in Natural Language Processing .The task of Named Entity Recognition is to categorize all Proper Nouns in a document into predefined classes like Person , Organization , Location etc. There are many applications of NER like Machine Translation, Question–Answering Systems, Indexing of identification of Proper Nouns and their classification.

A few of the various Named Entity Classes assumed to be work on are identified in following Categories:

- Person Name
- Organization Name
- Location Name
- Designation
- Abbreviation
- Title Person
- Courses
- Movie Name
- Famous Personality
- Disease
- Book Name
- Bank Name

Manisha, M.Tech Scholar (CSE), MERI College Of Engineering And Technology, Bahadurgarh
Mrs. Preethi Jose Dolly (Asst. Proffesor, CSE) ,MERI College Of Engineering And Technology, Bahadurgarh

Information extraction

Information extraction (IE) is a type of information retrieval whose goal is to automatically extract structured information from unstructured and/or semi-structured machine-readable documents. In most of the cases this activity concerns processing human language texts by means of natural language processing (NLP). Recent activities in multimedia document processing like automatic annotation and concept extraction out of images/audio/video could be seen as information extraction.
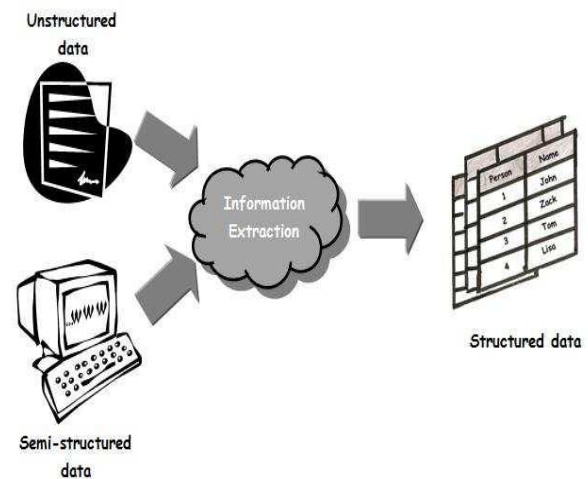

Fig 1, Information Extraction

II. PRIOR WORK

There are some of developed NER System which has been developed in that area are given below:-

ANNIE

ANNIE is one of many Information Extraction systems that have been developed using GATE (General Architecture for Text Engineering). It uses finite state algorithms and the JAPE language.

AFNER

AFNER is a package for named entity recognition. It is written in C++. A combination of approaches is used in order to find named entities. A list of named entities is created and each method adds the found named entities to the list.

Firstly, regular expressions are used to find simple case named entities such as simple dates, times, speed, etc. Secondly, parts of text matching listed named entities are found. The regular expression and list matches are then used in maximum entropy based classifier. Features relating to

individual tokens (including list and regular expression matches) as well as contextual features are used.

BANNER

BANNER is a named entity recognition system, primarily intended for biomedical text. It is a machine-learning system based on conditional random fields and contains a wide survey of the best features in recent literature on biomedical named entity recognition (NER). BANNER is portable and is designed to maximize domain independence by not employing semantic features or rule-based processing steps.

III. PROPOSE WORK

In this research work, we aimed to develop an NER system or application for English and Hindi Language. It was aimed to develop an web based NER System based on LLA approach which can be used with other NLP applications like CLIR, MT etc. The purpose of Developing NER System is to, find out the useful features for NER task. We have concentrated on recognizing given named entities such as NEP (Person), NEPH (Person- Hindi), NEL (Location), NELH (Location-Hindi), NED (Designation), NEBAT (Banking-Terms), NEBAO (Banking-Organization), NEO (Organization), NEOH (Organization-Hindi), NES (Surname), NESH (Surname-Hindi), NEM (Movie), NETP(Title-Person), NEBT(Book-Type), NEBN(Book-Name), NEBA(Book-Author), NECN(College-Name), NECA(College-Abbreviation). NECO(Course), NECOA(Course-Abbreviation),NEDI(Disease), NEFP(Famous Personality),NEMO(Month) and NEW(Week).

IV. RESULT

Fig 2, Home Page of Developed NER

V. DISCUSSION

It can be used to analyze other field data like medicine or solve any word related query, household things, Colleges etc. to fulfill the user needs. This developed tool can be used to identify the named entity with small changes into the system likely by extending some more related field with which we want to extract the information.

VI. CONCLUSION

The system has been developed and we can efficiently use that developed system which we have designed. The developed system is developed using PHP & MySQL on windows platform which is very easy to use.

REFERENCES

[1] Darvinder Kaur and Vishal Gupta. 2010 A survey of Named Entity Recognition in English and other Indian Languages. In Proceedings of (IJCSI) International Journal of Computer Science Issues Vol. 7 Issue 6 (pp 239-245).

[2] Asif Ekbal, Sivaji Bandyopadhyay.2008 Bengali Named Entity Recognition using Support Vector Machine. In the Proceedings of the IJCNLP-08 workshop on NER for South and South East Asian Languages (pp 51-58). Hyderabad, India

[3] Amandeep Kaur, Gurpreet Singh Josan and Jagroop Kaur. 2009 Named Entity Recognition for Punjabi: A Conditional Random Field Approach. In Proceedings of 7th international conference on Natural Language ProcessingICON-09. Macmillan Publishers, India.

[4] Ali ELSEBAI, A Rule Based System for Named Entity Recognition in Modern Standard Arabic ,School of Computing , Science and Engineering University of Salford , Salford , UK, November 2009"Ph.D Thesis".

[5] B Sasidhar, P M Yohan, A Vinaya Babu and A. Govardhan, A Survey on Named Entity Recognition in Indian Languages with particular reference to Telugu, IJCSI,Vol 8,Issue 2 ,pp-438-443,March 2011.

[6] Pramod Kumar Gupta , Sunita Arora , An Approach for Named Entity Recognition System For Hindi :An Experimental Study ,Proceedings of ASCNT-2009,CDAC ,Noida, India ,pp.103-108.