



**YILDIZ TECHNICAL UNIVERSITY
FACULTY OF ELECTRICAL AND ELECTRONICS
COMPUTER ENGINEERING DEPARTMENT**

SENIOR PROJECT

Mongolian Named Entity Recognition

Project Supervisor: Assoc.Prof.Dr. Banu Diri

Project Group
06011902 Davaadulam Davaakhuu

İstanbul, 2012

CONTENT

SYMBOL LIST	iv
ABBREVIATION LIST	v
FIGURE LIST	vi
TABLE LIST	vii
PREFACE	viii
ABSTRACT	ix
ÖZET	x
1.INTRODUCTION	1
2.MONGOLIAN LANGUAGE	2
2.1.About the Mongolian Language.....	2
2.2.Between Mongolian and Turkish Relationship.....	3
3.FEASIBILITY	5
3.1.Schedule Feasibility.....	5
3.2.Technology and System Feasibility.....	5
4.NAMED ENTITY RECOGNITION	6
4.1.Named Entity Recognition.....	6
4.2.Mongolian NER.....	8
5.USING METHODS	9
5.1.Rule Based System.....	9
6.STAGES OF PROJECT DEVELOPMENT	10
6.1.Rules of Mongolian Named Entity Recognition.....	10
6.1.1.Person Names.....	11
6.1.2.Place Names.....	12
6.1.3.Organization Names.....	13
6.1.4.Other Rules.....	14
7.RESULT	15
7.1.Result of Experiments.....	15
7.2.The System Constraints.....	16
7.2.1.The Constraints For Person Names.....	16
7.2.2.The Constraints For Place Names.....	17
7.2.3.The Constraints For Organization Names.....	17

8.CONCLUSION.....	19
INTERNET RESOURCES.....	20
REFERENCES.....	20
CURRICULUM VITAE.....	21

SYMBOL LIST

P Precision

R Recall

ABBREVIATION LIST

ACE	Automatic Content Extraction
AI	Artificial Intelligence
CRF	Conditional Random Field
HMM	Hidden Markov Model
ME	Maximum Entropy Model
MEMM	Maximum Entropy Markov Model
ML	Machine Learning
MUC	Message Understanding Conference
NER	Named Entity Recognition
NERC	Named Entity Recognition and Classification
NLP	Natural Language Processing
NLU	Natural Language Understanding
RBS	Rule Based System
SL	Supervised Learning
SVM	Support Vector Machine

FIGURE LIST

Figure 2.1	Mongolian National Script.....	3
Figure 2.2	Mongolian Cyrillic.....	3
Figure 3.1	Gantt Chart	5
Figure 6.1	System Print Screen.....	14

TABLE LIST

Table 2.1	Mongolian Cyrillic Letters.....	2
Table 2.2	Similar word of Mongolian and Turkish.....	4
Table 7.1	Analysis of precision and recall.....	15
Table 7.2	Cardinal Success.....	16
Table 7.3	Success by news' type.....	16

PREFACE

This project is about Mongolian Named Entity Recognition and it is performing first time in the Mongolian Language. I believe this project will be beneficial study for Mongolian Natural Language Processing projects. I really thank to my supervisor, Assoc.Prof.Dr. Banu Diri for advised me the project and leaded me.

Sincerely,

Davaadulam Davaakhuu

ABSTRACT

Named Entity Recognition is one of the most important branches of Natural Language Processing. Its purpose is making finder software which is finding proper names such as person name, organization name, and places name by performed rule of language. In early years this type of project was done for languages such as English, Chinese, Hindi, and Turkish etc. However, this type of project has never performed for the Mongolian language. This project is about Named Entity Recognition of Mongolian language. That is why this thesis uses software Mongolian grammar, shapes, forms, and rules.

ÖZET

Varlık İsmi Tanıma, doğal dil işlemedeki çalışma alanlarından biridir. Varlık ismi tanıma ile bir dile ait kişi, yer, kurum isimlerini bulmak gibi zaman, saat, tarih varlık isimlerini tespit etmek mümkündür. Bu alanda İngilizce, Türkçe ve Çince gibi birçok doğal dilde çalışmalar yapılmıştır. Moğolca üzerine bu alandaki ilk çalışma olup, Moğolca yazılmış bir dökümandaki kişi, yer ve kurum isimlerinin tespit edilmesini sağlamıştır.

1. INTRODUCTION

Natural language processing (NLP) is one of important area of computer science. NLP's main goal is designing and making software that can analyze, transform and interpret a language which spoken and written by human.

Although the field of NLP was originally referred to as Natural Language Understanding (NLU) in the early days of AI, it is well agreed today that while the goal of NLP is true NLU, that goal has not yet been accomplished. A full NLU System would be able to:

1. Paraphrase an input text
2. Translate the text into another language
3. Answer questions about the contents of the text
4. Draw inferences from the text [I4]

NLP's one useful subfield is Named Entity Recognition (NER). This thesis is about Mongolian Named Entity Recognition and we will have more information about NER in the next chapters.

2. MONGOLIAN LANGUAGE

Mongolian is the native language of Mongolia and is spoken by 5.2 million people who are all Mongolian residents and some residents of the Inner Mongolia autonomous region of China. In this chapter we will learn about Mongolian and the relationship between Turkish and Mongolian.

2.1. About the Mongolian Language

The Mongolian language is the official language of Mongolia and the best-known member of the Mongolic language family. Before 1946, Mongolian National Script as Uigur alphabets (Figure 2.1) was used. Since 1946, Mongolian people used the Cyrillic alphabets (Figure 2.2) and 90 percent of the population knows writing and reading by Cyrillic. There are 35 alphabets in Cyrillic Mongolian and they are shown by Table 2.1.

Mongolian has vowel harmony and a complex syllabic structure for a Mongolic language that allows clusters of up to three consonants syllable-finally. It is a typical agglutinative language that relies on suffix chains in the verbal and nominal domains. While there is a basic word order, subject–object–predicate, ordering among noun phrases is relatively free, so grammatical roles are indicated by a system of about eight grammatical cases. There are five voices. Verbs are marked for voice, aspect, tense, and epistemic modality [I3].

Table 2.1 Mongolian Cyrillic Letters

	Mongolian Cyrillic	Turkish Counterpart		Mongolian Cyrillic	Turkish Counterpart
1	А а	A a	19	Р р	R r
2	Б б	B b	20	С с	S s
3	В в	V v	21	Т т	T t
4	Г г	G g	22	У у	U u
5	Д д	D d	23	Ү ү	Ü ü
6	Е е	Y y	24	Ф ф	F f
7	Ё ё	-	25	Х х	H h
8	Ж ж	С с	26	Ц ц	--
9	З з	Z z	27	Ч ч	Ç ç
10	И и	I i	28	Ш ш	Ş ş
11	Й й	--	29	Щ щ	--

12	К к	К к	30	Ъ	--
13	Л л	Л л	31	Ы	І і
14	М м	М м	32	Ь	--
15	Н н	Н н	33	Э э	Е е
16	О о	О о	34	Ю ю	--
17	Ө ө	Ӧ ӧ	35	Я я	--
18	П п	Р р			

Modern Mongolian evolved from "Middle Mongolian", the language spoken in the Mongol Empire of the 13th and 14th centuries. In the transition, a major shift in the vowel harmony paradigm occurred, long vowels developed, the case system was slightly reformed, and the verbal system was restructured.

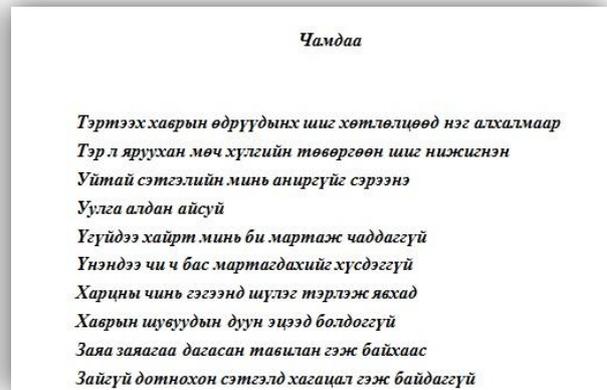
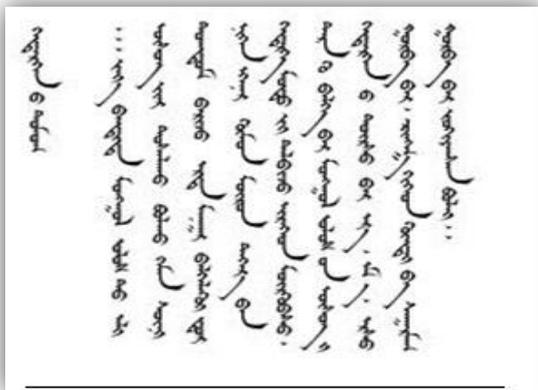


Figure 2.1 Mongolian National Script

Figure 2.2 Mongolian Cyrillic

2.2. Between Mongolian and Turkish Relationship

Altay is language family of at least thirty five languages spoken by people who live from Eastern Europe and Mediterranean to Middle East and Middle Asia [I6]. In Altay group Turkish languages, Mongolian, Tungusic, Japanese and Korean. Also Estonian, Finnish and Hungarian are some of relative languages.

Their similar features are:

1. Adding by structure way
2. No prefix
3. Root doesn't changing when it is suffixed or derived
4. First accessory elements then main elements
5. This languages have no grammatical gender
6. They have question supplement
7. Mutual supplements. Especially between Mongolian and Turkish [I5]

Actually, Mongolian and Turkish sentence structure is same like subject–object–predicate. Also, there are almost three hundred words that are almost the same in the two languages. Example is below in Table2.2.

Table 2.2. Similar words of Mongolian and Turkish

Mongolian	Turkish
Tsetseg	Chichek
Sandal	Sandaliye
Khar	Kara
Uls	Ulus
Uzem	Uzum
Jil	Yil
Tumur	Demir
Tavag	Tabak

3. FEASIBILITY

In this chapter we can see our feasibility study. Our feasibility study is trying to show about our project resources and about our planning which is from project beginning to project ending.

3.1. Schedule Feasibility

Time line chart Figure3.1 shows our planning and our project's steps.

ID	TASK	START	END	DURATION	2011			2012	
					Oct	Nov	Dec	Jan	Feb
1	Planning and Organizing	29-Sep-11	06-Oct-11	6d	█				
2	Researching and Brainstorming	29-Sep-11	16-Jan-12	78d	██████████				
3	Logical Designing	04-Oct-11	06-Oct-11	3d	█				
4	Processing Design	05-Oct-11	11-Jan-12	71d	██████████				
5	Coding	05-Oct-11	19-Jan-12	77d	██████████				
6	Analizing	10-Oct-11	19-Jan-12	74d	██████████				
7	Testing	21-Nov-11	20-Jan-12	45d			██████		
8	Project Completion	09-Jan-12	20-Jan-12	10d				█	

Figure 3.1 Gantt Chart

3.2. Technology and System Feasibility

NetBeans is very useful program for projects which are based object oriented programming language. That's why we choose NetBeans for our project with Java programming language. Here is also our technical specification:

Windows 7 Ultimate

System:

Manufacturer: Sony

Processor: Intel(R) Core(TM)2 Duo CPU

Installed memory(RAM): 4.00 GB (2.87 GB useable)

System type: 32-bit Operation System

4. NAMED ENTITY RECOGNITION

Named Entity Recognition is our topic of this thesis. This part explains about NER and Mongolian NER system.

4.1. Named Entity Recognition

NER is an important subtask for information extraction tasks such as machine translation, information retrieval, question answering and so on. Also it involves the identification and classification of named entities: expressions that refer to people, places, organizations, products, companies and dates, times, or monetary amounts.

Early work in NER systems in the 1990s was aimed primarily at extraction from journalistic articles. And then it is getting turned to processing of military dispatches and reports. Since about 1998, there has been a great deal of interest in entity identification in the molecular biology, bioinformatics, and medical natural language processing communities [11].

Different kind of NER systems were evaluated as a part of the Sixth Message Understanding Conference in 1995 (MUC). The target language was English. The participating systems performed well. However, many of them used language-specific resources for performing the task and it is unknown how they would have performed on another language than English [4].

After 1995 NER systems have been developed for some European languages and a few Asian languages. There have been at least two studies that have applied one NER system to different languages. Palmer and Day [4] have used statistical methods for finding named entities in newswire articles in Chinese, English, French, Japanese, Portuguese and Spanish. They found that the difficulty of the NER task was different for the six languages but that a large part of the task could be performed with simple methods. Cucerzan and Yarowsky [5] used both morphological and contextual clues for identifying named entities in English, Greek, Hindi, Rumanian and Turkish. With

minimal supervision, they obtained overall F measures between 40 and 70, depending on the languages used [12].

The first study on NER that is in Turkish texts is the one by Cucerzan and Yarowsky where the authors present a language-independent NER system. A statistical NER system for Turkish was presented by the authors report evaluation results comparable to the results of similar systems for English.

The sequences of tokens matching the items in the lexical resources or matching the patterns in the pattern bases are annotated accordingly by the NER system using the named entity tags (ENAMEX, TIMEX, and NUMEX) proposed in the Message Understanding Conference series [2].

The entity tags are shown below:

1. Names (enamex)
 - Person
 - Place
 - Organization
2. Times (timex)
 - Date
 - Time
3. Numbers (Numex)
 - Money
 - Percent

MUC working type is in sentence kind of XML type. For example:

<ENAMEX TYPE="PERSON">**Davaa**</ENAMEX> worked for <ENAMEX TYPE="ORGANIZATION">**Workcube Inc.**</ENAMEX> from <TIMEX TYPE="DATE">**2006**</TIMEX> to <TIMEX TYPE="DATE">**2009**</TIMEX> in <ENAMEX TYPE="PLACE">**Turkey**</ENAMEX> with <NUMEX TYPE="MONEY">**2000\$**</NUMEX> salary a <TIMEX TYPE="TIME">**month**</TIMEX> .

There are a variety of techniques for NER. Two broadly classified approaches to NER are:

- Linguistic approach and
- Machine learning based approach.

The linguistic approach is the classical approach to NER. It typically uses rules manually written by linguists. The recent Machine Learning (ML) techniques make use of a large amount of annotated data to acquire high-level language knowledge. ML based techniques facilitate the development of recognizers in a very short time. Several ML techniques have been successfully used for the NER task. ML approaches like Support Vector Machine (SVM), Conditional Random Field (CRF), and Maximum Entropy Markov Model (MEMM) are also used in developing NER systems [3].

Also these techniques are Supervised Learning (SL) and they have been used many times before in NER systems. Here are some of the examples:

D. Bikel used SL techniques include Hidden Markov Model (HMM) in 1997, S. Sekine used Decision Trees in 1998, also same year A. Borthwick used Maximum Entropy Models (ME), M. Asahara and Matsumoto used SVM in 2003 and same year A. McCallum and Li used CRF [1].

4.2. Mongolian NER

Mongolian NER system is being done by rule based system which helps to recognize person names, place names and organization names. And the system uses additional Mongolian language grammars with java programming language. In addition it is followed by MUC. Also the system is giving influential outcomes and easy to use. We can just upload Mongolian text file as input then get the names with program's rules. We will see about the rules in detailed by sixth chapter.

5. USING METHODS

This chapter covers methods to Mongolian NER.

5.1. Rules Based System

In computer science, rule-based systems are used as a way to store and manipulate knowledge to interpret information in a useful way. They are often used in artificial intelligence applications and research [I7].

Rule based systems main aim is filtering the high amount of data with specified rules and performing lexical analysis to compile or in NLP. Useful areas are: Credit Card Fraud Detection, Telecom Fraud, Intrusion Detection Systems-IDS and Named Entity Recognition etc. A typical RBS has four components.

1. A list of rules or rule base
2. An inference engine
3. Temporary working memory
4. A user interface [I8]

In this thesis we purpose to make a NER system using a RBS and our rules are explained with more detail in the below.

6. STAGES OF PROJECT DEVELOPMENT

The purpose of our project is making NER system for Mongolian. From the beginning until now our project is getting more effective.

First of all our system needs text input. After get text, pre-processing starts that is separating words from symbols (- . ‘ < ’ > etc), making the text sentence by sentence and detecting spaces to make one space if spaces are more than one. Then our Mongolian NER system is firstly finding all potential words which are uppercase words and some lowercase accessory words. For example: uppercase word- Монгол (Mongolia), lowercase accessory word- улс (country). Then the system saves picked words in a candidate list. When the program saves the words to the list, the list needs five features of that word. The five features are the own word, next word of the word in the text, previous word of the word, the word's offset address, the word's type which needs to understand word is person name or place name or organization name. After getting the candidate words, our system tries to separate person names, organization names and place names from the list with rules.

The next step is finding proper names from the candidate list. This step is materialized under certain rules. The list words compared one by one with person name's rules then place name's rules then organization name's rules. If candidate word is equal with person name's rule its type should be one, if its equal with place name its type is two, if its equal with organization name its type is three. We will know about the rules below.

After find proper names the text should be colored by the type. Person name is red, place name is blue and organization name is green. If candidate word is not proper name, its color is black.

6.1. Rules of Mongolian Named Entity Recognition

In this chapter we can see rules for Mongolian Named Entity Recognition one by one.

6.1.1 Person Names

Our system firstly compares the candidate words with person name's rules. Now here are the person name's rules.

1. We have a string array which includes some common person names. Our system compares candidate words with that array then if it finds equal word, makes candidate word a proper name. Some parts of the array: "Чингис", "Кадафи", "Чаплин" (Chinggis, Kadafi, Chaplin).
2. We prepared a string array. The array has an accessory words that usually comes in front of the person names. And the system compares candidate word's front word with this array word. If it finds equal word, candidate word's type will be one which means this candidate word is a person name. Here is part of the array: "Гишүүн", "дарга", "хуульч", "охин" and "аав" (deputy, commander, lawyer, daughter, and father).
3. Also same operation for back of person names. Example of found words: "Гуай", "охин", "аав", "ээж", "эгч" and "авгай" (mister, daughter, father, mother, sister, and lady).
4. If the candidate word's first character is uppercase and second character is dot, the candidate word should be a person's name. For example: Р.Бадамдамдин and Д.Энхбат. Because the uppercase character is reporting Surname's first character. If surname is Jordan and first name is David the name can be like J.David.
5. Also system is detecting connector words. If two or more candidate words are connected "ба" or "болон" ("ба" means "and", "болон" means "or") and also these words are existing in front or back of person names, system makes person names to other candidate words automatically. For example: Даваа ба Долгор (Davaa and Dolgor).

6. Also same operation available for comma with candidate words. For example: охин Дулам, Бат, Болд (Dulam, Bat, Bold).

6.1.2. Place Names

Our system compares candidate word with place name's rules after person name comparing.

1. We have a string array named by selfPlace. The array includes some broad place names. And the system compares candidate words with this array and if it finds equal word set the candidate word's type to two. Which means the candidate word is a place name. Here is some parts of the array: "Монгол", "Япон", "Улаанбаатар", "Англи", "Европ" (Mongolia, Japan, Ulan-Bator, England, and Europe).
2. Another prepared string array includes words to describe place names by existing in front of place names. For example: "муж", "хот", "Өмнөд", "Дундад" and "тосгон" (state, town, southern, middle, and village).
3. Also same operation for back of place names. Example of found words: "сумтай", "хот", "тосгон" and "орон" (small group place, town, village, country).
4. If the all the word's characters are uppercase and last character is "У", which means this word is a place's name. For example: БНСУ and АНУ (KOR-Korea, USA)
5. Also system is detecting connector words. If two or more candidate words are connected "ба" or "болон" ("ба" means "and", "болон" means "or") and also these words are existing in front or back of the place names, system make place names to other candidate words automatically.

6. Also same operation available for comma with candidate words.

6.1.3. Organization Names

After place name's rule, our system compares candidate word with organization name's rules.

1. We have a string array named by selfOrganization. The array includes some common organization names in Mongolian texts. And the system compares candidate words with this array and if it finds an equal word set the candidate word's type to three. Which means the candidate word is an organization name. Here is some part of array: "Тавантолгой", "Корпораци", "Монцамэ" and "Google".
2. String array includes words to describe organization names by existing in the back of organization names. Here is "цэнгэлдэх", "агентлаг", "үйлдвэр", "компани" etc. (station, agent, factor, and company)
3. If all the word's characters are uppercase which means this word is an organization's name. For example: УИХ and МНН.
4. Also the system is detecting connector words. If two or more candidate words are connected "ба" or "болон" ("ба" means "and", "болон" means "or") and also these words are existing in front or back of organization names, the system makes organization names to other candidate words automatically.
5. Also same operations are available for comma with candidate words.
6. Organization name can sometimes be acronym words. That is why our system finds long meaning of organization name. For example: **Монголын хөрөнгийн бирж (МХБ)**. (Mongolian)

6.1.4. Other Rules

1. If proper names are first and the next word is a candidate word, the system makes that word's proper name too. It depends with which proper name is nearby that candidate word.
2. Also our system has another string array which is detecting wrong words from proper names then changes that word's type to zero. Which means the system makes gives a black color to that word and the word looks like normal word. For example: Индонезийн Ерөнхийлөгч Сусило Бамбанг (Indonesia's president Susilo Bambang). System finds the first word as a place name then finds the person name (Susila Bambang) with president word. But normal program makes president is the person name too. Because of another rules. But detecting rule finds president word from them then changes that word as normal word.

Here is in Figure 6.1, Mongolian NER System is made by Java programming language.



Figure 6.1 System Print Screen

7. RESULT

This chapter is about Mongolian NER system's experimental results and the system's some constraints.

7.1. Result of Experiments

In our evaluation of the result, the proper names were considered to be recognized correctly, if they were an exact match of the similar entities in the data. Our classification results are performed by well known performance metrics that are precision and recall. Precision is defined as the ratio or percentage of named entities found by the system that is correct. Recall is the ratio or percentage of named entities present in the corpus that were actually found by the system. Their formats are shown by Table 7.1 and below formulas.

Table 7.1 Analysis of precision and recall

	Correct	Incorrect	
Retrieved	a	c	a+b=m
Un retrieved	b	d	c+d=N-m
	a+c=n	b+d=N-m	a+b+c+d=N

$$P = \frac{a}{n} \text{ (Precision)} \quad (1)$$

$$R = \frac{a}{m} \text{ (Recall)} \quad (2)$$

$$F \text{ measure} = \frac{2 \times P \times R}{P + R} \quad (3)$$

We used 50 text files to our result which are type of economy, sport, health, political, and current gazette news etc. The files have total around 22600 words and found 1652 proper names. Cardinal success value of the system is shown by Table7.2. Other successes which are listed by type of news are shown by Table7.3.

Table7.2 Cardinal Success

Proper Names	# name	# Correct name	# Incorrect name	Precision	Recall	F measure	Success rate
Person	470	414	73	0.85	0.88	0.86	86%
Place	519	477	105	0.82	0.92	0.87	87%
Organization	663	546	82	0.87	0.82	0.84	84%

Table7.3 Success by news' type

Type	Person	Place	Organization
Political	82%	78%	93%
Current	85%	80%	76%
Economy	87%	95%	92%
Magazine	82%	82%	83%
Health	96%	85%	79%
Scientific	76%	83%	-
Sport	78%	75%	87%
Art & Culture	86%	89%	65%

7.2. The System Constraints

In despite of our system works with good success, still it has some constraints.

7.2.1. The Constraints For Person Names

1. If candidate word comes front or back of person name, our system makes the candidate word as person name. For example: Гишүүн **О. Магнайг Тагнуулын Ерөнхий Газрын** албан хаагчид дудаж уулзжээ (FBI's coworkers called member O.Magnai and met him). In this sentence system makes О. Магнайг as person name and plus makes Тагнуулын Ерөнхий Газрын (FBI) as person name too.
2. If candidate word connected with person name by connector words (and, or) or comma, the system makes the candidate word as person name. For Example:

Жүжигчин **С. Өлзийхүү** , **УДЭТ** - ын жүжигчин **М. Түвшинхүү** , жүжигчин **А. Алтантуяа** оролцон тоглосон юм (Actress S.Ulziikhuu, SOT's actress M.Tuvshinkhuu and actress A.Altantuya acted all together). In this sentence the system makes С. Өлзийхүү as person name then makes УДЭТ (SOT- State Opera Theatre) person name too because of comma.

3. Some place and organization names are named by popular person name. That is why sometimes system makes some organization or place name tags person name.
4. Sometimes person name's key word existed front or back of candidate. In this situation system tags wrong words as person name.

7.2.2. The Constraints For Place Names

1. If candidate word comes front or back of place name, our system makes the candidate word as place name. For example: **Канад** , **Хятадын “ Оюу Толгой ”** уул уурхайн компани (Canada, Chinese “Оюу-Толгой” company). Our system tags Оюу Толгой (company name) as place name. Because Хятадын (Chinese) word comes front of the word.
2. If candidate word connected with place name by connector words (and, or) or comma, the system tags the candidate word as place name.
3. Sometimes place name's key word existed front or back of candidate word. In this situation system tags the words as place name.

7.2.3. The Constraints For Organization Names

1. In Mongolian gazette news, usually organization names come as abridgement. However, sometimes words which not organization names are written by abridgement. For example: “ **LV** ” брэнд (“LV” brand).
2. If candidate word comes front or back of organization name, our system tags the candidate word as organization name. For example: **ЭНЭ MCS** компани... (This MCS company...). System tags ЭНЭ (This) as organization name because of MCS компани.

3. If candidate word connected with organization name by connector words (and, or) or comma, the system makes the candidate word as organization name.
4. Some organization names start with place names. That is why sometimes our system finds organization name like half is place name and half is organization name. For example: **Монголын Хөрөнгийн Бирж** (Mongolian Stock Exchange Place).
5. Sometimes organization name's key word existed front or back of candidate word. In this situation system tags wrong words as organization name.

8. CONCLUSION

Named Entity Recognition is wide area of Natural Language Processing. NER's one of important method is Rule Based system which is getting data by lexical rules. In this project our system recognizing proper names which are person, place and organization names for Mongolian language. Our Mongolian NER system was tested mixed text files. The test result success was for person name 86%, for place name 87%, and for organization name 84%. This project can be developed by using additional rules. As a future work, we aim to develop by adding new proper name areas such as date and money.

INTERNET RESOURCES

- [I1] http://en.wikipedia.org/wiki/Named_entity_recognition
- [I2] <http://www.cnts.ua.ac.be/conll2002/ner/>
- [I3] http://mn.wikipedia.org/wiki/Кирилл_МОНГОЛ_бичгийн_дүрэм
- [I4] <http://www.cnlp.org/publications/03nlp.lis.encyclopedia.pdf>
- [I5] http://tr.wikipedia.org/wiki/Altay_Dilleri
- [I6] http://en.wikipedia.org/wiki/Altay_language
- [I7] http://en.wikipedia.org/wiki/Rule-based_system
- [I8] <http://www.j-paine.org/students/lectures/lect3/node5.html>

REFERENCES

- [1] David Nadeau, Satoshi Sekine “A survey of named entity recognition and classification.pdf”
- [2] Dilek Küçük, Adnan Yazıcı (2009) “Rule-based Named Entity Recognition from Turkish Texts-DilekKucuk.pdf”
- [3] Named Entity Recognition in Hindi using Maximum Entropy and Transliteration.pdf
- [4] David D. Palmer, David S. Day (1997) “A Statistical Profile of the Named Entity Task”.
- [5] Silviu Cucerzan, David Yarowsky (1999) “Language independent named entity recognition combining morphological and contextual evidence”.
- [6] Zeynep Banu Özger (2011) “Türkçe Metinler için Kural Tabanlı Varlık İsmi Tanıma”

CURRICULUM VITAE

Name Surname : Davaadulam Davaakhuu

Birthday : 10/ 08/ 1986

Place of birth : Mongolia

High school : Sukhbaatar's Second High School

Practical training : Workcube e- business Inc, İstanbul (7 weeks for internship, 8 weeks for practice)

Mongolian Police Department, Mongolia (4 weeks for internship)

Nettapps Inc, Mongolia (4 weeks for internship)