

## Design and Implementation of Mongolian Wordnet Management Platform

Hasi

Computer and Information Engineering College  
Inner Mongolia Normal University  
Huhhot, China  
e-mail: hasi@imnu.edu.cn

Tang Enbo

Computer and Information Engineering College  
Inner Mongolia Normal University  
Huhhot, China  
e-mail: tangenboysj@yahoo.com.cn

*Abstract*—With the development of natural language processing technology, a powerful tool containing semantic information is in great need in lexical semantic processing. Aiming at automatic processing of words in machine translation and automatic proofreading, Wordnet mainly provides semantic information in the form of a semantic knowledge database. The Mongolian Wordnet management and application platform includes two parts---the user searching function and the administrator maintaining function. Users can search semantic knowledge online and the administrator can maintain the adding, deleting, revising and searching functions of the database online as well. This article mainly introduces the construction theory of Mongolian Wordnet, the designing frame of the management and application platform, and the designing methods of the main function modules.

*Keywords*- Mongolian Wordnet, management platform, database

### I. INTRODUCTION

Dictionaries have been regarded as the knowledge base of some specific language knowledge. To some extent, dictionaries are reference books on book-shelves in libraries. And in the current computerized time, these visible and touchable dictionaries have been replaced by various kinds of online ones---- computer accessible dictionary databases. Traditional online dictionaries arrange lexical information in alphabetical order. Although these dictionaries are powerful and provide us with convenient ways of searching entries, it is hard to find semantically related words or synonyms in this kind of system, for the system is not arranged according to the semantic order. So what we need is a kind of “built-in” dictionary, which belongs to part of intelligence structure and can be carried anywhere. Thus, a suitable carrier of lexical knowledge becomes a key question. Wordnet is an efficient combination of traditional dictionary information, advanced computer technology and psycholinguistic research.

Wordnet is an online lexical reference system, a machine dictionary based on the principle of psycholinguistics. With the recognized spelling method, the synonym sets (words which can replace one another in some contexts), Wordnet involve both lexical links and semantic links. It also involves hyponymy, antonymy, synonymy and meronymy in sense relation. As a kind of ontology, Wordnet is the psycholinguistic research achievement on lexics. Since the

establishment of English lexical database by the cognitive science research institute in Princeton University in 1985, Wordnet has almost been the most commonly used frame of lexical semantic knowledge base all over the world. Wordnet plays an important role in natural language processing and information searching. Lexical semantic network is established based on the lexical semantic relations between words and will be the basic resource of the future internet---semantic web. Based on the high quality lexical semantic network, the information searching mistakes will be highly reduced, semantic information can be provided to the information searching in unknown fields and will be better understood. Meanwhile, because of the assumption that the cognitive semantic structures of languages have something in common, as the research and application standard of human lexical concept knowledge base, Wordnet will play a more and more important role. Wordnet in different language version is still in research now. The frame of Wordnet has been widely accepted in the fields of lexical semantics and computer dictionary. Establishing multilingual lexical semantic network which expresses semantic relations between languages can increase the accuracy of multilingual information searching, text classification and machine translation. Thus the construction research of Wordnet-based multilingual lexical semantic network including Chinese and Mongolian language has a significant meaning.

Compared with other languages like Chinese and English, the informationization of Mongolian started quite late. In particular, there are many blanks in Mongolian semantic research based on information processing. There is not a complete theory system and a theoretical frame, some basic theoretical questions are still under argument. That is to say Mongolian semantics is an old and young discipline. The lack of basic research in Mongolian semantic analysis and expression, and the slow progress in Mongolian semantic information processing restrict the further research of Mongolian information processing. As an important part of lexicology in traditional linguistics, word meaning is always the necessary important part of Mongolian lexical research, and there have already been many related achievements, which provide the theoretical basis for Mongolian information processing. But traditional research achievements can not meet the needs of information processing. In the field of information processing, due to the fundamental and significant role played of lexical semantic relations, more and more research work has been done in this field. Mongolian Wordnet is a kind of semantic network

which provides lexical semantic information, and it aims at providing semantic knowledge for automatic processing in the application systems such as machine translation, automatic proofreading, text searching. The construction of Mongolian Wordnet is an important part of automatic text processing. It can be applied in language resource management, meet the needs of professional users to search and process word concepts in a large corpus, and will be extensively used in language teaching and bilingual comparison research.

## II. CONSTRUCTION OF MONGOLIAN WORDNET

### 2.1 Basic Idea of Construction

Wordnet organizes nouns, verbs, adjectives and adverbs into synonym sets, each set stands for a basic word concept, and there are many lexical semantic relations between these sets including hyponymy, antonymy, synonymy and meronymy. In Mongolian semantic research based on information processing, many scholars use the method of seme analysis to make semantic classification of Mongolian words, especially nouns, verbs and adjectives. Based on this achievement, we hope to draw on the designing methods and principles of other languages' semantic network such as Chinese Wordnet, English Wordnet and Chinese Concept Dictionary (CCD), and combine them with the characteristics of Mongolian language to describe and reflect the semantic relations of Mongolian such as hyponymy, antonymy, synonymy and meronymy.

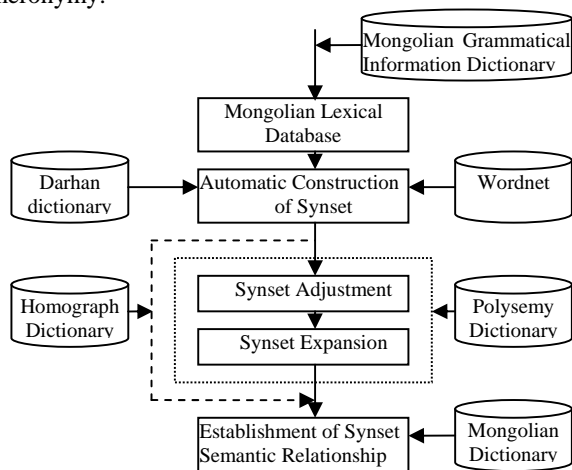


Figure 1: Flow Chart of Semantic Network Establishment

### 2.2 Homograph Sets

Wordnet's frame is based on synset, so synset-based semantic web, establishing Mongolian language synset is the premise of the compatibility between Mongolian Wordnet and Wordnet in other languages. In order to make full use of the research results of other scholars, the thesis labeled the Mongolian words from Mongolian Grammatical Information Dictionary with synset ID. The main approach is to search for the corresponding Chinese word for each word of

Mongolian Grammatical Information Dictionary in Chinese Wordnet, and mark with the corresponding Chinese synset ID. Considering the incompleteness or inaccuracy of the labeling, the thesis finds these words from Darhan dictionary first, and then from the Chinese Wordnet. Through the above method more than 8000 nouns' synset ID labeling has been completed. The corresponding Chinese words like SIYANBEI do not exist in Chinese Wordnet, the synset ID of these words are manually set. If the synset ID of two words are the same, they are homographs.

The data table of Mongolian Wordnet synsets is as follows:

ID	UGSAIMAG	GALIG	MONGGOL	synset_id
2	Ne	NINGCEU	ᠨᠢᠩᠴᠡᠭᠦ	
3	Ne	ARANDAG	ᠠᠷᠠᠨᠳᠠᠭ	102415195
4	Ne	SIDELTE	ᠰᠢᠳᠡᠯᠲᠡ	701467904
5	Ne	NERMEGESU	ᠨᠡᠷᠮᠡᠭᠡᠰᠤ	600062789
6	Ne	NERBEGDEL	ᠨᠡᠷᠪᠡᠭᠢᠳᠡᠯ	700132263
7	Ne	HVDALDVGACID	ᠬᠢᠪᠳᠠᠯᠳᠠᠩᠳᠠᠭᠴᠢᠳ	
8	Ne	NEREDULGE	ᠨᠡᠷᠡᠳᠤᠯᠭᠡ	605948118
9	Nt1	HVDAGVI	ᠬᠢᠪᠳᠠᠭᠢ	
10	Ne	HVDAGAI	ᠬᠢᠪᠳᠠᠭᠠᠢ	
11	Nn1	SIYANBEI	ᠰᠢᠶᠠᠨᠪᠡᠢ	
12	Ne	HVDHVMSI	ᠬᠢᠪᠳᠠᠬᠤᠪᠮᠰᠢ	
13	Ne	SILGARAL	ᠰᠢᠯᠭᠠᠷᠠᠯ	600152253
14	Ne	HVRICAL	ᠬᠢᠪᠷᠢᠴᠠᠯ	604741650
15	Ne	NIGULESHUI	ᠨᠢᠭᠤᠯᠡᠰᠤᠬᠡᠰᠢ	101009911
16	Ne	NIGULESUL	ᠨᠢᠭᠤᠯᠡᠰᠤᠯᠤ	101009911
17	Nn1	SILBI	ᠰᠢᠯᠪᠢ	101790426
18	Ne	AYVNGG_A	ᠠᠶᠢᠨᠭᠭᠠ	100085128
19	Ne	AYVMSIG	ᠠᠶᠢᠮᠰᠢᠭ	607060236
20	Ne	HVRAMDVLG_A	ᠬᠢᠪᠷᠠᠮᠳᠠᠯᠭᠠ	600955185

Figure 2: Table of Mongolian Wordnet Synsets

### 2.3 Semantic Relations

Semantic relations are the core of semantic network, and plentiful semantic relations can be established according to Wordnet.

A noun normally has only one direct superior word, so the editor of the dictionary can define the noun using this superior; and a noun normally has more than one inferior words, so the editor seldom displays all of them. Nouns in Wordnet is just organized according to this hyponymy, so hyponymy is the main organizing basis in semantic network and has a very important application value. Some verbs, similar to meronymy among nouns, have another meaning hidden inside. For example, **ᠠᠶᠢᠨᠭᠭᠠᠷᠢ** (to snore) includes **ᠠᠶᠢᠨᠭᠭᠠᠷᠢ** (to sleep). If this idea is accepted, this pair of verbs can be thought to have meronymy between each other, for instance, **ᠠᠶᠢᠨᠭᠭᠠᠷᠢ** (to snore) and **ᠠᠶᠢᠨᠭᠭᠠᠷᠢ** (to dream) are a part of **ᠠᠶᠢᠨᠭᠭᠠᠷᠢ** (to sleep), because the former two actions must happen during the time span of latter action.

After the establishment of Mongolian synset, the hyponymy, antonymy and meronymy relations of Chinese Wordnet will be added to Mongolian synset in an automatic transformation way so as to efficiently form the main framework of Mongolian lexical semantic web. Although differences exist between languages, people's interpretation of the world is interlinked, similar or even identical at the level of basic concept. Therefore using the semantic relations in Chinese Wordnet to construct the main framework of Mongolian lexical semantic web is a better path. Then combining with the research results of traditional Mongolian

lexical semantics, in a way of manual intervention, linguistics experts complete the work of further adjusting and improving the semantic relations. For example, because the word  $\text{ᠰᠢᠶ᠋ᠠᠨᠪᠡᠢ}$  SIYANBEI does not have a corresponding node in other Wordnet, after establishing the node in Mongolian lexical semantic network, this node can be added to the hyponym position of Ethnic Minority according to Mongolian dictionary. Therefore, the semantic relationship that  $\text{ᠰᠢᠶ᠋ᠠᠨᠪᠡᠢ}$  SIYANBEI is a minority is established.

These semantic relations are established between synsets, these relations are expressed by establishing links between two synset IDs in Wordnet database. For example, the table of hyponymy is as follows:

wn_hyponymy : 表	
synset_id_1	synset_id_2
100002056	100001740
100002342	100002056
100002452	100002056
100002560	100002056
100002645	100016236
100003009	100016236
100003226	100003009
100004358	100003226
100004483	100003226
100004609	100003009
100004740	100003009
100004824	100003009
100005598	100001740
100006026	100003226

Figure 3: Table of Mongolian Wordnet Hyponymy

So, all information of Mongolian Wordnet is stored in database, management and application are actually operating these tables.

### III. MONGOLIAN WORDNET'S MANAGEMENT AND APPLICATION PLATFORM

As a language knowledge database, Mongolian Wordnet, aims at provision of assistance of language knowledge users. Users can be divided into two types: normal users and administrators. Normal users have the right to search for word information in Mongolian Wordnet database. Administrators can complete all kinds of maintaining work of the Wordnet database after log on the management and application platform, such as adding, deleting, revising, searching and counting, to dynamically update the database.

The platform is designed according to B/S mode, which has the advantage that the platform can be operated anywhere and no specific software is needed. That is to say, as long as a computer with the internet access is available, installation and maintenance is not needed. Meanwhile, the expansion of the system is very easy. The system is developed mainly using jsp technology.

All kinds of applications can be realized on the platform, such as database management and maintenance by the administrators, setting synset IDs and hypernyms. Normal users can register and log on the system and then search in

Chinese, Mongolian and English Wordnets. In order to help people to better understand the system and popularize it, information on Mongolian Wordnet is displayed on the main page. The platform uses jsp technology to design the main interface and sql language to achieve the operation of data.

#### 3.1 Designing Frame of Mongolian Wordnet System

Main page of the system is a visual communication tool, so the designer of it must concern about the overall arrangement. Although its design is not equal to graphic design, there is still much similarity between them. The format design uses combination of words and graphs to display harmony and beauty. The page design of Multi-page sites needs to display the organic links of related pages, especially relations between order and content between pages and within a page. Most of all, in order to achieve the best visual effect, a reasonable overall arrangement is necessary.

The overall arrangement of the system is as follows:

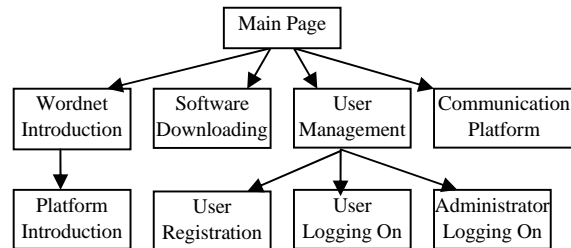


Figure 4: Overall Arrangement of the System

The main page is divided into two parts. The first part mainly introduces related information, services, construction of Mongolian Wordnet and semantic relations. Another part is user management part which includes user registration and user logging on. Registration, which is mainly about user classification and management, is managed in backstage system. Design of main page mainly uses jsp language, and uses functions such as css for background design.

#### 3.2 Introduction to System Function

The Mongolian Wordnet management and application platform has two functions: Mongolian Wordnet foreground searching program and the backstage database management program of Mongolian Wordnet searching system. The former is used to search words online, manage users and for the logging on of administrators. The latter is used to maintain and manage the backstage database, for example, to add, delete, correct and search data in the data tables.

##### 3.2.1 Design of Searching Function

This function is provided to normal users. After registering and logging on to the system, normal users can enter the searching interface, entering keywords, and search the corresponding content in the database. This is achieved by sql language embedded in java. With reference to

Wordnet 2.1 when designed, searching function can meet the basic needs of normal users.

### 3.2.2 Design of Maintenance Function

Considering the operation process of database administrators, easy operation, friendly, flexible, practical and safe interface, the management and maintaining program of Mongolian Wordnet database is divided into four function modules: searching records, adding records, revising records and deleting records. The searching records function can be achieved either by word searching or ID searching. All function modules are as follows:

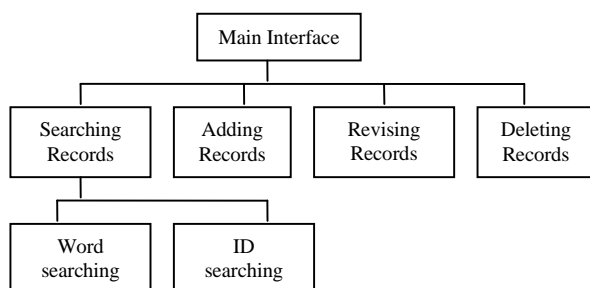


Figure 5 Function Modules

The `wm_mongol` table management page mainly uses tables, forms, iframes, the connection to the database as well as the searching language of sql.

The connection language to the database and the searching language of sql are as follows:

```

String spath = "data/MongolianWordnet.mdb";
String dbpath = application.getRealPath(spath);
String url = "jdbc:odbc:Driver={Microsoft Access Driver (*.mdb)};DBQ="+dbpath;
Class.forName("sun.jdbc.odbc.JdbcOdbcDriver");
Connection conn= DriverManager.getConnection(url);
Statement
stmt=conn.createStatement(ResultSet.TYPE_SCROLL_SENSITIVE,ResultSet.CONCUR_UPDATABLE);
String sql = "select * from wn_mongol";
ResultSet rs = stmt.executeQuery(sql);
    
```

## IV. CONCLUSION

The establishment of Mongolian Wordnet management and application platform has already been completed now, the testing of many functions has been completed, which can basically satisfy users' requests. With the further research work, the application requests of Mongolian Wordnet will be even richer, the function of the platform will be stronger. The content of the system database will be dynamically updated continuously, so, more lexical semantic knowledge will be provided to users, and this is the main purpose of establishing the system.

## ACKNOWLEDGMENT

This paper is supported by Humanity and Social Science Research Project of Ministry of Education (No.11YJC740033) ,and The Natural Science Foundation of Inner Mongolia (No.2011MS0905) .

## REFERENCES

- [1] G.A.Miller.An online lexical database.International Journal of Lexicography , 1990 ,3 (4) :235 - 244.
- [2] C.Fellbaum. Co-occurrence and antonymy . International Journal of Lexicography , 1995 ,8 (4) :281 - 303.
- [3] C.Fellbaum.Wordnet-An Electronic Lexical Database.MIT Press, in 1998.
- [4] J.R.Huang,S.K.Xie,J.F.Hong,Y.Z.Chen,Y.L.Su,Y.X.Chen,S.W.Huang,Chinese Wordnet:Design,implementation,and Application of an Infrastructure for Cross-Lingual Knowledge Processing, Journal of Chinese Information Processing, February in 2010.
- [5] Nasun-urtu, Semantic Research for the Mongolian Language Oriented to Information Processing,Journal of Inner Mongolia University(Humanities and Social Sciences),2002.5.
- [6] Q.X.Chen, An online lexicon of semantic class:Wordnet, applications of languages and words, 1998.2.
- [7] C.R.Huang,R.Y.Chang, B.L.Shiang,SinicaBow(Bilingual Ontological Wordnet):Integration of Bilingual Wordnet and SUMO [C].//Proceedings of LREC2004/Lisbon, 2004:1553-1556.
- [8] Z.D.Dong,Q.Dong.OntologyandHowNet[OL],[2006-04-23].http://www.keenage.com/html/e\_index.html.
- [9] L.Zhang,J.J.Li,M.H.Hu,T.S.Yao,Implementation of Chinese Wordnet, Journal of Northeastern University(Natural Science), the fourth issue in volume 24, 2003(4).
- [10] Y.D.Bi, Y.P.Yan, Constructing a Wordnet-Based Multilingual Lexical-semantic Net:A Semi-automatic Method, Journal of PLA University of Foreign Languages, 2008.9.
- [11] S.Wang, C.G.Cao, A Method for Automatic Translation of Wordnet Concepts into Chinese, Journal of Chinese Information Processing, 2009.7.