

# HowNet

<http://www.keenage.com>

Zhendong DONG

Qiang DONG

e-mail: [dzd@keenage.com](mailto:dzd@keenage.com)

HowNet is an on-line common-sense knowledge base unveiling inter-conceptual relations and inter-attribute relations of concepts as connoting in lexicons of the Chinese and their English equivalents. We are happy to share it over the internet and expect more users whom we are gladly inviting to help perfect and further develop it.

## 1. Motivation

Dong Zhendong brought to light the following viewpoints in a series of papers published in 1988.

- (a) In the final analysis, natural language processing ultimately requires the support of a powerful knowledge base;
- (b) *Knowledge*, specifically, the form of knowledge that is computer-operable, is a system encompassing the varied relations amongst concepts as well as those amongst the attributes of concepts. As one acquires more concepts, or rather, captures more relations amongst concepts alongside the links between the attributes attached to the concepts, one simply becomes more knowledgeable;
- (c) On the creation of a knowledge base, a common-sense knowledge base constituting a knowledge system should first be constructed. This database shall describe general concepts and map out the relations among them;
- (d) On who should build knowledge base, Dong believes knowledge is owned by all. A meaningful and robust knowledge base is far too vast and profound for a handful few to attempt. On this account, Dong proposed that the knowledge engineers first design the framework and suggest a common-sense knowledge base prototype. Upon this foundation, work can be extended to develop a specialized knowledge base, which rests its weight on professionals in the respective fields. The idea is analogous to the edition of a dictionary for general use and an encyclopedia.

Research and construction of HowNet is a manifestation of the above-mentioned viewpoints.

## 2. Philosophy of HowNet

A profound understanding of the philosophy of HowNet is crucial to mastering and applying it. The philosophy behind HowNet lay ground on its understanding and interpretation of the objective world. The crux is, we state, all matters (physical and metaphysical) are in constant motion and are ever changing in a given time and space. Things evolve from one state to another as recorded in the corresponding change in their attributes. In the case of “human”, it is characterized by the four main states of living: at birth, aging, fall sick and dead. Age (an attribute) catches up in a person, giving the attribute “age” a value, i.e. "old". As a person grows, his/her hair color (an attribute) turns grey (the attribute-value). On the other hand, as a person grows, the character (metaphysical) gradually matures (attribute-value), so is the knowledge (metaphysical product) that will develop wider and deeper (the attribute-values). The above depicts the units for manipulation and description in HowNet being *thing* (sub-divided into physical and mental), *Part*, *Attribute*, *Time*, *Space*, *Attribute-value* and *Event*.

We like to emphasize the significance of *Part* and *Attribute* in the philosophy of HowNet. The way we understand *Part* is that all objects are probably part of something else while at the same time, all objects are also the *whole* of something else. Doors and windows are parts of buildings while the limbs are parts of animals. However, at the same time, buildings form parts of a community and the individual is part of the family or society he/she belongs to. All things can be divided into their respective components. Space can be segmented into “up”, “down”, “left”, “right” while Time can be seen from “the past”, “the present” and “the future”. Nothing can only function as a component and not a whole and the reverse is true. Depending on the system of reference, the same point of reference can either be regarded as a *whole* or a *part*. In HowNet, *Part* is taken as a constituent in a larger whole. The role and function of *Part* in *whole* is analogous to the human body, for instance, “hilltop”, “hillside”, “mountain foot”, “table leg”, “back of chair”, “estuary”. "door" and "window" of buildings are analogous to the relevant parts of the human body such as the eyes, the mouth etc. It is interesting to note that the same analogy applies to different languages. This shows how similar the mankind shares their views on the relations between part and whole.

The way we understand *Attribute* is any one object necessarily carries a set of attributes. Similarities and differences between the objects are determined by the attributes they each carries. There will be no object without attributes. Human beings are attached with natural attributes such as race, color, gender, age, ability to think, ability to use language as well as social attributes such as nationality, class origin, job, wealth etc. Under specific conditions, it is true to say that the attached attributes are even more important than the host itself, a fact most evident in the “next-best alternative” exercises associated with our daily life. For instance, if we want to clamp a nail on the wall but does not have a hammer, what would be the best alternative tool? Obviously, it would be something that carries attributes close to a hammer, where in this case, weight and

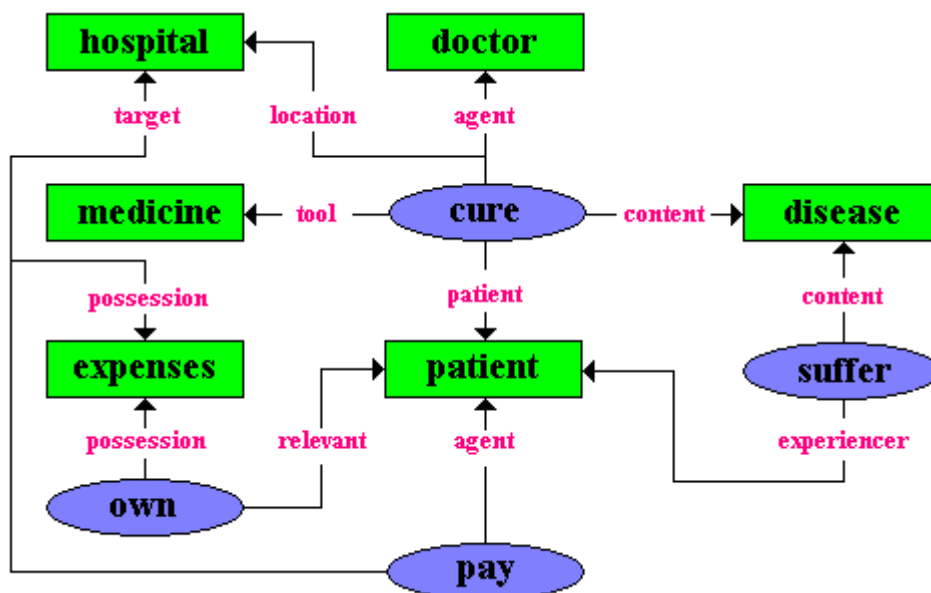
hardness would be the key attributes. The relationship between the attributes (e.g. weight and hardness, etc.) and their host (a hammer) is unbending. The attributes simply come with the host and vice versa. The attribute-host relation differs from the part-whole relation. HowNet reflects this difference by way of coding specifications such that attributes are necessarily defined in terms of the possible classes of host. In this connection, HowNet also requires pointers to indicate the relevant attributes when defining attribute-values.

### 3. Characteristic of HowNet

Fully computational is the characteristic of HowNet. It is a system by the computer, for the computer, and expectantly, of the computer.

As a knowledge base, the knowledge structured by HowNet is a graph rather than a tree. It is devoted to demonstrate the general and specific properties of concepts. For instance, “human being” is the general property of “doctor” and “patient”. The general properties of “human being” are documented in Main Features of Concepts. Being the *agent* of *cure* is the specific attribute of "doctor" while being the *experiencer* of *unwell* is the specific attribute of "patient". Be it the millionaire or the poor; the beauty or the ugly, being a *human being* is the general property they all share though each take a distinct attribute-value, namely, *rich*, *poor*, *beautiful* and *ugly*.

HowNet spares no effort in mirroring complicates of inter-concept relations as well as inter-attribute relations. HowNet teaches the following knowledge graph to the computer so that they are computer-operable.



Pic. 1

In sum, HowNet explicates the following relations:

- a. Hypernym-Hyponym (implied by main features of concepts, see "HowNet Management Tool")
- b. synonym (by means of "SACR")
- c. antonym (by means of "SACR")
- d. converse (by means of "SACR")
- e. part-whole (coded with pointer %, e.g. "heart", "CPU", etc)
- f. attribute-host (coded with pointer &, e.g. "color", "speed", etc)
- g. material-product (coded with pointer ?, e.g. "cloth", "flour", etc)
- h. agent-event (coded with pointer \*, e.g. "doctor", "employer", etc)  
(may also be "experiencer" or "relevant", depending on the type of event)
- i. patient-event (coded with pointer \$, e.g. "patient", "employee", etc)  
(may also be "content" or "possession", etc. depending on the type of event)
- j. instrument-event (coded with pointer \*, e.g. "watch", "computer", etc)
- k. location-event (coded with pointer @, e.g. "bank", "hospital", "shop", etc)
- l. time-event (coded with pointer @, e.g. "holiday", "pregnancy", etc)
- m. value-attribute (coded without pointer, e.g. "blue", "slow", etc)
- n. entity-value (coded without pointer, e.g. "dwarf", "fool", etc)
- o. event-role (coded with role-name, e.g. "wail", "shopping", "bulge", etc)
- p. concepts co-relation (coded with pointer #, e.g. "cereal", "coalfield", etc)

A notable characteristic of HowNet is that synonyms, antonyms and converse relations can be generated by the users themselves based on the rules specified in "Synonymous, Antonymous and Converse Relations" instead of coding each of them overtly on every concept as WordNet does.

HowNet is a knowledge system, not a semantic dictionary although we termed the general knowledge base upon which HowNet operates as the Knowledge Dictionary. All documentation on HowNet, including the Knowledge Dictionary forms an organic knowledge system. To name a few, the Main Features of Concepts, the Secondary Features of Concepts, Synonymous, Antonymous and Converse Relations (SACR) and Event Relatedness and Role-shifting (ERRS) are fundamental components of the system and not merely coding specifications. We expect them to be used in conjunction with the Knowledge Dictionary.

#### **4. Methodology**

As a knowledge system that describes relations between concepts as pictured above, HowNet is not a thesaurus. HowNet attempts to construct a graph structure of its knowledge base from the inter-concept relations and inter-attribute relations. This is the fundamental distinction between

HowNet and other tree-structure lexical databases. The philosophy of HowNet and its very nature underlined its unique method of building.

#### 4.1. Extraction of Sememe

Defining sememes is as difficult as defining morpheme. However, just as morpheme, sememes, though labourious defining, are easily used and understood. Broadly speaking, a sememe refers to the smallest basic semantic unit that cannot be reduced further. Take for instance “human being”, despite being a most complex concept encompassing a set of attributes, it can be regarded as a sememe. We hypothesise that all concepts can be reduced to the relevant sememes. We deem further that there exist a close set of sememes, from which, composes an open set of concepts. If we can manage the close set of sememes to describe inter-concept relations as well as inter-attribute relations, an ideal knowledge base would be conceivable. Using the Chinese language to search for this close set of sememes is really trying a short cut. The Chinese characters (including simple word) is a close set that can be exploited to express both simple and complex concepts, as well as the inter-concept and inter-attribute connections.

We like to highlight an important method used in the extraction of sememes: the set of sememe is established on meticulous examination of about 6000 Chinese characters. Take the Event class for instance, we ever extracted as much as 3200 sememes from Chinese characters (simple morpheme).

After the necessary merger, 1700 sememes are derived for further classification that finally resulted in about 700 sememes. Note that up till this point, no polysyllabic words (in Chinese) are involved. These 700-odd sememes then served as a tagging set to tag polysyllabic words, and in the process we made necessary adjustment and extension when the set cannot satisfy the requirements. Finally the process arrived at a set of over 800 sememes we are now using in HowNet.

To illustrate the point to our English-speaking counterparts, imagine going through the motion using English. We would extract a common event sememe, "treat1" (provide medical treatment for) from the following English word: doctor, patient, hospital, medicine, therapy...

In sum, the building of HowNet is a bottom-up grouping approach. The first step is to form a tagging set of sememe through detail studying of all fundamental sememes and then apply tests to perfect the sememe list.

#### 4.2 Examination and Confirmation of Sememes

At the formation of an initial list of sememes grouped to serve as a basic tagging set, the issues of examination and confirmation arise.

First, we should check the coverage of the list of sememes against an extended scope of corpus annotation. We have set a rule for this process. When there exist a word with multiple concepts, say eight, and if the existing list of sememes failed to classify all the eight concepts, then we will

have to adjust the tagging set. We expect this to be the case at large. There are instances where we should exercise judgment to determine if we owe a certain concept the merit to stand on its own.

Next, examine the status of specific sememes in the concept network. If a sememe stands out among other concepts in either the same or a different category, then, it is a stable sememe that must be kept. Take the event “treat1” for instance, it appears under “medical treatment”, “to treat”, “to seek treatment” and the like. It also appears under terms like “doctor”, “hospital”, “medicine”, “clinic”, and “disease” among others. As such, the sememe “treat1” is stable and shall be retained.

The extraction of sememes and their examination are most crucial and detrimental to HowNet. It is a consistent process in the building of HowNet. We can therefore conclude the characteristics of the methodology employed by HowNet to be bottom-up and involve interaction between a tagger set and the final knowledge dictionary.

## 5. Preview to HowNet Knowledge System

### 5.1 Database and Documentation of HowNet knowledge system

The HowNet knowledge system includes the following database and documentation:

(01) Chinese-English Bilingual Knowledge Dictionary

(02) HowNet Management Tool

(03) Some Guidelines

- I Event Role and Features (ERF)
- I Word Class
- I Synonymous, Antonymous and Converse Relations (SACR)
- I Event Relatedness and Role-shifting (ERRS)
- I Pointers and their usage

The scale of HowNet depends on the size of its Chinese-English Bilingual Knowledge Dictionary. Given that it has gone online, amendments are made convenient. The size of HowNet is measured on the base of number of word/phrase entries and the concept entries. The size of version 2.0 is as follows:

Language Type	Total Word Entries	N Category	V Category	A Category
Chinese	050220	026037	016657	09768
English	055422	028876	016706	10716

Language Type	Total Concept Entries	N Category	V Category	A Category
Chinese	062174	029787	020468	011173

English	072994	036770	021203	014339
---------	--------	--------	--------	--------

Table 1

Note: The N, V, A categories referred to herein are not the Noun, Verb and Adjective equivalents.

## 5.2 Record Format in HowNet Knowledge Dictionary

The HowNet Knowledge Dictionary is the heart of the whole system. In this Dictionary, every concept of a word or phrase and its description form one entry. Regardless of the language types, an entry will comprise four items. Every item is made up of two portions joined by the “=” sign. To the left of the “=” sign is the data field, while that on the right is the data value. The items are arranged in the following sequence:

W\_X= word / phrase form

G\_X = word / phrase syntactic class

E\_X = example of usage

DEF = concept definition

### 5.2.1 Selection of Words and Phrases and their Concepts

As it is known that the knowledge dictionary of HowNet is based on Words and Phrases and their concepts. How do we select words and phrases and their concepts?

Firstly, we do not believe that the Chinese language has words in as strict sense as that in European languages. We select words and phrases mainly from a 80,000 words and phrases with usage frequency out of a very large corpus with 400 million Chinese characters, rather than from any current Chinese dictionary. Much attention has been paid to those currently popular in usage, such as "Internet", "Euro", "dioxin", and "download", "click" or " hacker " in computer subject.

Secondly, for the selection of concepts or meanings, we do not just follow any ready-made Chinese dictionary. Our careful attention has been paid to the popularity of any meaning of a word or phrase. We usually only choose those meanings which are still in use and discard those obsolete ones.

Thirdly, the knowledge dictionary of the current version is a Chinese-English bilingual one. The purpose of doing so is not to provide an ordinary Chinese-English dictionary but to check if the description of meanings will fit both languages.

### 5.2.2 Examples for Words and Phrases

We mainly provide examples for those words and phrases which have more than one meaning. The emphasis is given to the capability of disambiguation rather than its explanatoriness. To take

two of meanings of the Chinese word "打" for example, one meaning is:"buy|买", and the other is: "weave|辫编". They are found in the knowledge dictionary as:

NO.=000001

W\_C=打

G\_C=V

E\_C=~酱油, ~张票, ~饭, 去~瓶酒, 醋~来了

W\_E=buy

G\_E=V

E\_E=

DEF=buy|买

NO.=015492

W\_C=打

G\_C=V

E\_C=~毛衣, ~毛裤, ~双毛袜子, ~草鞋, ~一条围巾, ~麻绳, ~条辫子

W\_E=knit

G\_E=V

E\_E=

DEF=weave|辫编

Suppose we come across a sentence as follows: "我女儿给我打的那副手套哪去了". The comparison between the semantic distance calculation of "手套"with"酱油" and "手套" with"毛衣" will help us tell which should be the correct choice in the given context. This method has two advantages: first, in most cases the disambiguation is to be done without rules on specific words and phrases; second, in most cases the algorithm is language-independent.

The compilation of examples is taken as a project named [97@YY001](#) funded by State Language Commission of China and implemented by the staff and students of Peking University. In HowNet Version 2 we show the examples for those in A, B, and C three letters.

## 6. Defining Concepts and the Rules

Description of concepts in HowNet is an attempt to present the inter-relation between concepts and that between their attributes. As such, the description is necessarily complex and unless a clear set of rules is installed, consistency cannot be guaranteed. Description of concepts includes both general and particular aspects.

At the same time, the method of description and the concerning rules must ensure that the inter-concept relations and inter-attributes relations are expressed thoroughly. In this connection,



the building of HowNet is also the design and building of such mark-up language. To date, the Knowledge Dictionary Mark-up Language (KDML) comprises the following components:

- (1) approximately 1500 features and event roles;
- (2) pointers and punctuation;
- (3) word order.

All the 1500 features are marked in bilingual to avoid ambiguity and ensure their readability, for example:

compile|编辑, software|软件...

## 6.1 General Rules

- (1) DEF shall not be left blank.
- (2) DEF shall include at least one feature. There is no limitation to the number of features in any DEF, only if the definition is reasonable in content and acceptable in terms of formats.
- (3) The first item in the DEF shall be a main feature as shown by "HowNet Management Tool". However, in the case of functional words such as prepositions, conjunctions, sentential adverbs etc., a secondary feature can be used for the first item, but it should be enclosed within {}.
- (4) A comma is used to separate the items, should there be more than one in any DEF. There should not leave a space between the comma and the next item.
- (5) Beside the first item, other items in the same DEF can also be a main feature. Note however that a main feature not placed in the first position shall lose its ability to inherit features in the hypernym-hyponym association.
- (6) all items in the DEF can be used with a pointer, even the first item.

## 6.2 Rules in Detail

### 6.2.1 Rules on Defining Event

- (1) DEF shall only begin with a main feature as shown by "HowNet Management Tool";
- (2) Complex event concepts shall be defined in accordance to the following rules:
  - (a) Use event roles for complex event concepts. This is because the complexity probably involves at least one event roles, for instance:

program: includes a event role -- PatientProduct

extemporize : includes a event role -- content

profiteer: includes a event role -- possession

graverobbery: includes a event role -- source

- (b) Event roles should be expressed in this format: class of event role = main / secondary features, for example, the word "program" should be coded as follows:

DEF=compile|编辑,ContentProduct=software|软件

#### 6.2.2 Rules on Defining Attribute-value and Numerical-value

- (1) "attribute-value" is the only main feature for concepts involving attributes. "numerical-value" is the only main feature for concepts relating to numerals. In this connection, they should take the first position in the relevant definitions.
- (2) In the definitions of concepts involving attributes and numerical, the second item states the property of the attribute/numeral as represented by the attribute-value / numerical-value concerned.
- (3) In most cases, define the specific value in the third position. e.g.

delicious: DEF=aValue|属性值,taste|味道,good|好,desired|良

crooked1: DEF=aValue|属性值,form|形状,curved|弯,

crooked2: DEF=aValue|属性值,behavior|举止,sly|狡,undesired|莠

#### 6.2.3 Rules on Defining Attribute and Numeral

- (1) The main feature for concepts involving attributes is "attribute"; while "numerals" is the main feature for numerical concepts. These will occupy the first position in the relevant definition.
- (2) All concepts of attributes and numerical must necessarily involve the use of the pointer "&" to indicate the host. For example:

taste: DEF=attribute|属性,taste|味道,&edible|食物

shape: DEF=attribute|属性,form|形状,&physical|物质

bearing: DEF= attribute|属性, behavior|举止,&human|人

The two sections on 6.2.2 and 6.2.3 illustrated specifically the relation network governing the thinking of HowNet. To put it simply, things carry some attributes and are in turn the host of those attributes while at the same time, each attribute necessarily carries a value. The above-listed examples show that: "An *edible thing* is the host of the attribute *taste*, and one of the values of the attribute *taste* is *delicious*." This is the way HowNet builds its graph of inter-concept and inter-attribute relationships.

#### 6.2.4 Rules on Defining Unit

- (1) “meter”, “kilometer”, “ton” and the like are what we refer to as Units. In the Chinese language, it also refers to the “noun classifier”(NounUnit) and “verb classifier”(ActUnit) that are unique to the language.
- (2) As with the attribute class, coded in the first position of the definition of any Unit must be “unit”, “NounUnit” or “ActUnit”. For example:

meter: DEF=unit|单位,&length|长度  
 round: DEF=ActUnit|动量,event|事件  
 dose: DEF=NounUnit|名量,&medicine|药物

#### 6.2.5 Rules on Defining “Thing”

- (1) “Thing” includes the following concept categories: “material”(including living and non-living things), “spiritual”(including sentiments, desires, thoughts and experience), “time”, “space”, “fact” and their component parts. It should be stressed that “fact” as described in HowNet is really “events”. This will be discussed further in section 7.
- (2) The rules HowNet has set for defining the concept class “Thing” are varied as different categories of concepts have different requirements. As a general guide, there are two points to note: first, the use of appropriate pointers and secondly, the order of pointers when more than one are used in one definition.
- (3) In defining concepts with specific attribute-value, this value (underlined> is used without a pointer, for example:

man: DEF=human|人,male|男  
 expert: DEF=human|人,able|能,desired|良  
 poser: DEF=problem|问题,difficult|难,undesired|莠

#### (4) Rules on Defining “Parts”

The second item in the definition will have to carry the pointer “%”to denote the *whole* in which the Part belongs to. The definition should try as much as possible to describe the position or function of the Part in the Whole. For example:

heart: DEF=part|部件,%AnimalHuman|动物,heart|心  
 CPU: part|部件,%computer|电脑, heart|心

The above definitions mean that “heart” and “CPU” are the parts for “Animal Human” and “Computer” respectively, while AnimalHuman and Computer are the respective *whole* of “heart” and “CPU”. Both the “heart” and “CPU” function as the focal point of their respective whole.

Common knowledge tells that if the “heart” is damaged, the *whole* will malfunction. Descriptions of this kind will help inference.

(5) In specifying the relation between a concept and an event, the following rules should be observed:

(a) If the concept is itself an event, mark the main feature as “fact”, mark the second item with the main feature of the event. Pointers are not necessary. For example:

tug-of-war: DEF=fact|事情,exercise|锻炼,sport|体育

(b) when the concept and the event are related in terms of event role, pointers are necessary. For example:

employer: DEF=human|人, \*employ|雇用  
employee:DEF= human|人, \$employ|雇用  
iron: DEF=tool|用具, \*AlterForm|变形状, #level|平  
vacation:DEF=time|时间, @rest|休息, @WhileAway|消闲  
hotel: DEF=InstitutePlace|场所, @reside|住下,#tour|旅游  
lifeboat: DEF=ship|船,\*rescue|救助

(c) If the event role relations involved between the concept and the event is complex, more pointers are necessary and ordering between the pointers is important. For example:

washing machine: DEF= tool用具, \*wash|洗涤, #clothing|衣物

In the above example, “wash” is the function of the “tool”, or that the “tool” serves to wash. “clothing” is marked with # to indicate that it is the patient of “wash”. This order cannot be reversed or mixed up. Yet another example is:

iron: DEF=tool|用具, \*AlterForm|变形状, #level|平

In this example, “level” is the attribute belonging to the patient of “AlterForm”, that is to say, it is the resultant change in the attribute of the patient undergoing “AlterForm”.

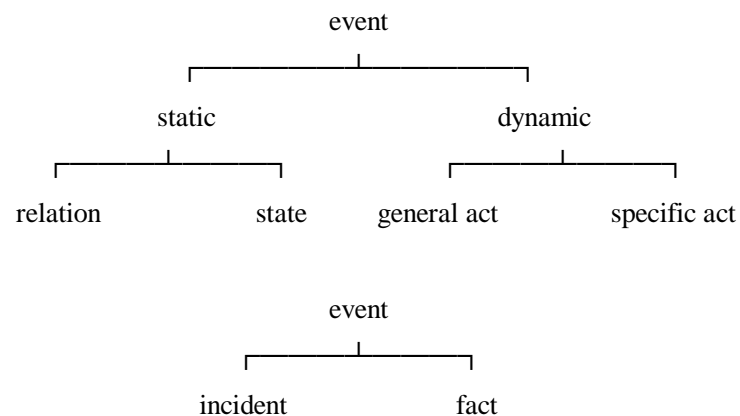
The above should give the reader a better understanding of the KDML. We believe that this language will be improved as we advance to make the grammar of the KDML more expressive and powerful.

## 7. On the Concept of “Event”

The main features of Events are shown in "HowNet Management Tool". There are more than 800 of such features, representing half of the total features as included in HowNet. This tells the importance of this class of concept as well as its status in HowNet. In the above-mentioned file, every main feature is attached with a set of necessary roles expressed within curly brackets { }. There is also a square bracket [] containing the relevant features.

### 7.1 Relation between Main Features

In HowNet, concepts under the Event class can be broadly classified as follows:



HowNet examines every concept under the Event category using a bottom-up approach and concluded that there are four types of relationship between the main features:

- (1) hypernym versus hyponym relation
- (2) static versus dynamic relation
- (3) relatedness of events
- (4) role-shifting

We have dealt with the hypernym-hyponym relation above.

Here, we like to first touch upon the static versus dynamic relation. Under Static, there are two categories, Relation and State of Event. Under Dynamic, there are “General action” and “Specific action” serving as the motivation in creating the Relation and State of Event. This forms the structure of a corresponding static and dynamic relation in HowNet. To put it simply, the relation or state of event always correspond to the relevant action. For instance, possession expresses the relation between things such that the sentence “I have a book” states the relation between "I" and "the book". Corresponding to this relationship or that which can change the possessive relation are actions such as *take* or *give*.

HowNet has identified 9 types of relation. Under state of event, there are two main categories, the physical state and the spiritual state. The physical state includes *Existence-Appearance*, *Be Normal*, *BeGood*, *BeBad*, *Disappear* (e.g. the living, aging, ill and death of living things). Spiritual state includes *Emotion*, *Attitude*, *Volition* and *Recognition*. HowNet held that all actions under the Event class correspond to the above mentioned relations and states. In the final analysis, all serve to show some “change”, be it a change in relation or a change in state. There are two categories which we like to draw your attention to: first, actions that changes specific attributes, such as *Make higher*, *Make lower*, *beautify*, *warm up* among others. Secondly, actions that changes a *Make Act* or *cause not to do*, such as *cause to do*, *request*, *order* and *prohibit*. Broadly speaking, these two categories of actions do not correspond to specific relation or state but are themselves a change in relation or state. For any physical entity, a change in attribute, for instance from cold to warm (under *warm up* action), is undergoing an internal change of state. Any physical entity, when develop other action or stop some specific action because of *the Make Act* or *prohibit* act, represents a change of its relation with the outside world. To better illustrate the picture, we lay out the structure of main features under Event as follows:

V	event   事件		
V1	static   静态	V2	act   行动
V1.0	relation   关系	V2.0	AlterRelation   变关系
V1.01	isa   是非关系	V2.01	AlterIsa   变是非
V1.02	possession   领属关系	V2.02	AlterPossession   变领属
V1.03	comparison   相比关系	V2.03	AlterComparison   变相比
V1.04	suit   相适关系	V2.04	AlterFitness   变相适
V1.05	inclusive   蕴涵关系	V2.05	AlterInclusion   变包含
V1.06	connective   关联关系	V2.06	AlterConnection   变关联
V1.07	CauseResult   因果关系	V2.07	AlterCauseResult   变因果
V1.08	TimeOrSpace   时空关系	V2.080	AlterLocation   变空间位置
		V2.081	AlterTimePosition   变时间位置
V1.09	arithmetic   数量关系		
V1.1	state   状态	V2.1	AlterState   变状态
		V2.11	AlterPhysical   变本体
V1.11	StatePhysical   物理状态		
V1.111	ExistAppear   存现	V2.111	CauseToExist   使存现
V1.112	begin   起始		
V1.113	BeNormal   常态	V2.113	AlterStateNormal   变常态
V1.114	BeGood   良态	V2.114	AlterStateGood   变良态

V1.115	BeRecovered 复原	V2.115	resume 恢复
V1.116	change 变		
V1.1161	AppearanceChange 外观变		
V1.1162	QuantityChange 量变	V2.1162	AlterQuantity 变数量
V1.1163	BeBad 衰变	V2.1163	AlterStateBad 变莠态
V1.1164	end 终结	V2.1164	kill 杀害
V1.1165	disappear 消失	V2.1165	CauseToBeHidden 使消失
V1.1166	WeatherChange 天变		
V1.117	ChangeNot 不变	V2.117	stabilize 使不变
V1.12	StateMental 精神状态	V2.12	AlterMental 变精神
V1.121	feeling 情绪	V2.1210	AlterEmotion 变情感
		V2.1211	ShowEmotion 表示情感
V1.122	Attitude 态度		
V1.123	volition 意向		
V1.124	recognition 感知状态	V2.124	AlterKnowledge 变感知
V1.1241	HaveKnowledge 有知	V2.12410	MakeOwnKnowledge 使自我感知
		V2.12411	MakeOthersKnowledge 使他人感知
V1.1242	NoKnowledge 无知	V2.1242	MakeNoKnowledge 使不知
V1.1243	misunderstand 误信	V2.1243	MakeMisunderstand 使误知
V1.1244	BeUnable 无能		
		V2.2	AlterAttribute 变属性
		V2.3	MakeAct 使之动

Relatedness of events involves interaction between dynamic states. The interaction can occur within the same category (within Static or within Dynamic) as well as across categories. For instance, *own* and *lose* are under one category. The relation between the two is such that the former is the necessary condition for the later. That is, there cannot be a *lost* for *OwnNot*. In another instance, *buy* and *own* though belonging to different categories, are related in the sense that the former is the necessary condition for the later. Also, between *regret* and *apologize*, the former is a static state while the later, in dynamic emotional state, is an action expressing sentiment. The internal relation between them is that the later is the logical result of the former. To illustrate further, *BeRecovered*, *cure* and *SufferFrom* all come under different categories. Both *SufferFrom* and *BeRecovered* belong to the static category while “cure” is dynamic. The link between them is that “cure” turned *SufferFrom* from the state of *BeBad* to the *BeRecovered* state.

Role-shifting refers to the case where the event role of an Event naturally performs another role in the cause of action, or that it is concurrently an event role of another event. For instance, the *agent* of *buy* will turn to become the *relevant* of *own*. Another example is the *experiencer* of *SufferFrom* is rightly the *patient* of *cure*, and the *patient* of *cure* will turn into becoming the *experiencer* of *BeRecovered*.

## 7.2 Necessary Roles

In HowNet, all 800 main features of the Event class are attached with a set of necessary roles. These stipulated event roles are described in the file “event roles and attributes”(ERF). Listed in the set are the must have roles of the feature concerned. This means, missing any of the roles listed cannot constitute the named event. We wish to highlight that what we are referring to is where the event does happen, it will necessarily involves all the listed roles, and this, however, may not be the case in actual speech, for which is not our concern. For instance, when the event “buy” takes place, it must involve the questions of who buys (the agent), buy what (the possession), from where (the source), how much to pay (the cost), and for whom (the beneficiary). When the event "pity" takes place, the roles of who pities (experiencer), pity whom (target) and for what (cause) naturally follows. Hence in Main Features of Concepts (1) (MFC-1), both "buy" and "pity" are attached with the following frame respectively:

buy 买	{agent, possession, source, cost, ~beneficiary}
pity 怜悯	{experiencer, target, cause}

Nevertheless, in actual speech, not all the above roles need to be mentioned in a sentence and what is not mentioned in actual speech does not signifies the absence. For reason that any event would take place at a specific time and space, it is not necessary to include time and space in the set.

The set of necessary roles serves to illustrate the general property of events. Therefore, they are essentially the basis for judging concepts in the construction of HowNet. For instance, to determine if the word “please” should go under a mental state category of static event “joyful|喜悦” or under a mental change category of dynamic event “please|取悦”, the judgement will be based on the respective set of necessary roles. For the event “please”, *target* must be one of the necessary roles, for example, in "he tried to please her", "her" is the *target*.

## 8. On the Concepts of “thing”

The main features for “thing” are shown in "HowNet Management Tool". These features are organized in a hierarchy to present the hypernym- hyponym relation relationship. The hierarchy in the “thing” class does not run as deep as in the "events" class and the descriptions are targeted at demonstrating both the general characteristics as well as the particular features. The general



characteristics of each concept are listed in square brackets [] while the particular features are coded in the respective DEF. Take the noun "teacher" for example.

DEF: human|人,\*teach|教,education|教育

As mentioned above, for the noun "teacher", "teach|教" and "education|教育" are the specific characteristics. Since the main feature of "teacher" is "human|人", the following constitutes its general characteristics: "!name|姓名", "!wisdom|智慧", "!ability|能力", "!occupation|职位", "\*act|行动". In addition, it naturally inherits all the general characteristics of its hypernyms "animate|生物", "physical|物质", and "thing|万物", that is, "!sex|性别", "\*AlterLocation|变空间位置", "\*StateMental|精神状态", "\*alive|活着", "!age|年龄", "\*die|死", "\*metabolize|代谢", "!appearance|外观", "#time|时间", "#space|空间".

Structuring the features in this approach make HowNet economical and effective. However, if a user does need to have all the features coded in a concept he/she has the flexibility to tailor to the specific needs using self-devised software.

## 9. Conclusion

The research and construction of HowNet span over a more than 10-year period. The author felt that it is most difficult to handle the following:

- (1) determine the main and secondary features as well as their organization;
- (2) determine the description method and establishing the KDML;
- (3) defining each and every of the concept that amounts to more than 50,000 entries.

The research and construction of HowNet is a piece of engineering work and basically an exploration in approach. We are certain that as a source of knowledge, it has wide application.

Future development of HowNet rests in four areas:

- (1) expand on the number of concepts within the existing language types;
- (2) expand to cover other language types;
- (3) refine KDML to make it more powerful;
- (4) identify a specific domain knowledge with reasonable scope and experiment on establishing specific domain knowledge base.

What is mentioned above centered on the development of HowNet. What is apparently more important is its application. It is on this account that it is released in the Internet.

## **Acknowledgements**

We are most grateful to all institutions and individuals that have supported and assisted us in one way or another. Not in the least, we like to acknowledge the following institutions: Chinese Information Processing Society of China, Research Center of Computer and Microelectronics Industrial Development, former Institute of System Sciences of National University of Singapore, Research Center of Computer & Language Information Engineering of the Academy of Sciences. We like to thank Project 97@YY001 funded by State Language Commission of China and the Project HKUST 6149/98E funded by the Hong Kong Research Grant Council for their investment in the further development of HowNet. We record special appreciation to Beijing Creative Next Technology Ltd. who has rendered us great support for many years and provides this web site. We also like to thank Dr. Tham Wai Mun, Nanyang Technological University, Singapore, for translating the introduction from Chinese into English and Dr. Gan Kok Wee, Department of Computer Science, HKUST, for his careful proof-reading this translation and his valuable suggestions on the revision of HowNet.

## **References**

(We only list those references indispensable for the construction of HowNet. We apologize for neglecting all the references we have used in the 10 years of research.)

- [1] General Character Dictionary of Contemporary Chinese, Institute of Language and Character Research, Chinese People University, Foreign Language Teaching and Research Press, 1987
- [2] Dictionary of Contemporary Chinese (Revised Edition), Dictionary Compilation of Institute of Language Research, Chinese Academy of Social Sciences, Commercial Press, 1996
- [3] Chinese-English Dictionary (Revised Edition), Dictionary Compilation of English Department, Beijing University of Foreign Languages, Foreign Language Teaching and Research Press, 1995
- [4] WordNet 1.6 Princeton University, 1999
- [5] SenseWeb, Institute of System Sciences, National University of Singapore, 1996
- [6] Oxford-Duden Pictorial English-Chinese Dictionary, Translated by Chunying Pu, Light Industry Press, 1988
- [7] LONGMAN English-Chinese Dictionary Of Contemporary English, Longman Group UK Limited, 1988
- [8] Grammatical Knowledge-base of Contemporary Chinese, Shiwen Yu, Qinghua University Press, 1998
- [9] English-Chinese Dictionary, Gusun Lu, Shanghai Translation Press, 1995
- [10] Tongyi Cilin (A Chinese Thesaurus), Jiaju Mei & Yunqi Gao, Shanghai Dictionary Press, 1983