Extended HowNet 2.0 – An Entity-Relation Common-Sense Representation Model

Wei-Yun Ma, Yueh-Yin Shih
Institute of Information Science, Academia Sinica
128 Academia Road, Section 2, Nankang, Taipei, Taiwan 115
{ma, yuehyin}@iis.sinica.edu.tw

Abstract

In this paper, we propose Extended HowNet 2.0 – an entity-relation common-sense representation model. Comparing to HowNet and Extended HowNet, E-HowNet 2.0 has the following improvements: (a) Reorganizing the hierarchical structure of primitives and basic concepts; (b) Rich lexical information: In addition to sense definition, each entry of lexical sense may also include operational expressions as well as semantic functions which facilitate future semantic composition processes. (c) Improvement of sense definitions and sense definitions for basic concepts. (d) Developing a new automatic ontology reconstruction system. (e) Developing a query system called E-HowNet Relation Database for flexibly clustering concepts.We hope Extended HowNet 2.0 can bring significant benefits to the community of lexical semantics and natural language understanding.

Keywords: E-HowNet, entity-relation, lexical semantics, ontology

1. Introduction

The purpose of designing the lexical semantic representation model E-HowNet is for natural language understanding. E-HowNet, an evolution and extension of HowNet (Dong and Dong, 2006), is a frame-based entity-relation representation model to define lexical senses and to achieve compositional semantics.

The current E-HowNet 2.0 [1] shows the following improvements: (a) Reorganizing the hierarchical structure of primitives and basic concepts; (b) Rich lexical information: In addition to sense definition, each entry of lexical sense may also include operational expressions as well as semantic functions which facilitate future semantic composition processes. Event frames are also provided. (c) Improvement of sense definitions and sense definitions for basic concepts. (d) Developing a new automatic ontology reconstruction system: In case of revisions of lexical sense expressions or nodes of conceptual hierarchy, the ontology reconstruction system may re-attach each lexical entry to appropriated ontological nodes and results a new ontology. (e) Developing a query system called E-HowNet Relation Database for flexibly clustering concepts.

The remainder of this paper is organized as follows. We describe the background of developing E-HowNet in Section 2 and elaborate the feature improvements of the current E-HowNet version 2.0 in Section 3. The online systems based on E-HowNet are introduced in Section 4. Section 5 is conclusions and future work.

2. Background

HowNet is an on-line common-sense knowledge base unveiling the inter-conceptual relations and inter-attribute relations of concepts conveyed by Chinese words and their English equivalents (Dong & Dong, 2006). Compared with WordNet, HowNet's architecture provides richer information apart from hyponymy relations. It also enriches relational links between words via encoded feature relations. The advantages of HowNet are (a) inherent properties of concepts are derived from encoded feature relations in addition to hypernymous concepts, and (b) information regarding conceptual differences between different concepts and information regarding morpho-semantic structure are encoded. HowNet's advantages make it an effective electronic dictionary for NLP applications. In recent years, HowNet has been applied to the researches of word similarity calculation (Liu & Li, 2002), machine translation (Dong 1999), and Information Retrieval (Dorr, Levow and Lin, 2000) etc.

When we say that a sentence is 'understood', we mean that the concepts and the conceptual relationships expressed by the sentence are unambiguously identified, and we can make correct inferences and/or responses. Therefore, to achieve natural language understanding, computer systems should know the sense similarity and dissimilarity of words and sentences. A representational framework which represents knowledge about lexical concepts and performs the following functions is needed. (a) Identifies synonymous concepts and measures similarity distance between two concepts (Liu and Li, 2002). (b) Knows the shared semantic features and feature differences between two concepts. (c) Provides unique indices to each concept, such that associated knowledge can be coded and accessed. (d) Language independent sense encoding. (e) Logical inferences through conceptual property inheritance system. (f) Dynamic concept decomposition and composition mechanisms. None of the currently available ontology provides all of the above functions and so far there has been little research on applying HowNet to semantic composition. We therefore extend HowNet to deal with this problem. The resulting system is called E-HowNet.

The development of E-Hownet started in 2003 (Chen et al., 2005). We adopt the set of primitives and taxonomy of HowNet and adjusted to suit the goal of semantic composition. The major extension features are: (a) Word senses are defined by not only primitives but also any well-defined basic concepts and conceptual relations; (b) Semantic relations are explicitly expressed; (c) A Uniform representation for content words, function words, as well

[1] http://ehownet.iis.sinica.edu.tw/index.php

as phrases; (d) The capacities of semantic composition and decomposition; (e) Near canonical representations for lexical senses and phrasal senses. The above characteristics of the E-HowNet which make the E-HowNet different from other ontologies.

Rather than creating a completely new ontology, E-HowNet links different ontologies. For instance, we established the links between HowNet sememes and WordNet synsets. Thus WordNet synsets (version 1.6) are used as an alternative intermediate representational language. In order to achieve unambiguous and language-independent definitions, E-HowNet adopts WordNet synsets as an alternative vocabulary for conceptual indexing and representation. As a conceptual representation that may use WordNet synsets as its description language, E-HowNet is universal and language-independent.

Figure 1: Top-Level of E-HowNet ontology

3. Improvements of Extended-HowNet

As mentioned above, the E-HowNet ontology is a reconstruction of the HowNet ontology. The major revision was to include the hierarchy for relations to enable semantic composition and decomposition (Chen et al., 2004). Therefore, the E-HowNet ontology is formed by entity taxonomy and relation taxonomy. Each word sense is a node of the taxonomy and expressed by an E-HowNet expression. Synonyms or near synonyms should be expressed by the same expression. The top levels of E-HowNet ontology is shown in Figure 1 and a complete taxonomy can be found in http://ckip.iis.sinica.edu.tw/taxonomy/.

3.1 Reorganizing the Hierarchical Structure of Primitives and Basic Concepts

In E-HowNet 2.0 all concepts are either primitive concepts or defined by simpler concepts (either primitive concepts or basic concepts) in terms of an entity-relation model (Chen et al., 2004; Chen K.J. et al., 2005; Chen Y.J.et al., 2005; Huang, Chung and Chen, 2008). A primitive concept will have an English equivalent beside it, e.g. {read|讀}, whereas a basic concept will be expressed by a Chinese word and its English translation pair which is further defined by primitive concepts, e.g. {狗|dog} defined as {livestock|牲畜:telic={TakeCare|照料:patient={family|家庭},agent={~}}}.

E-HowNet ontology is formed by all lexical senses as well as primitive and basic concepts in a hierarchical order. Any concept inherits all the fundamental features of its hypernym and must have at least one feature that its hypernym does not own. The improvement of E-HowNet 2.0 will be elaborated in the following subsections.

3.1.1 Multi-level Sense Representation by Primitives and Basic Concepts

Conventional sense representation have used semantic primitives to define and achive canonical representation for concepts (Wierzbicka, 1972), such as Conceptual Dependency representation (Schank, 1975) and HowNet. However, using primitives only to define concepts causes information degrading as it is almost impossible to understand a definition of a complex concept merely with primitives. Furthermore, it is debatable whether there exists a limited and fixed set of so-called primitives. Therefore, we adopt 2,233 primitives from HowNet and extend 2698 basic concepts which make a deeper hierarchical structure and more precise semantic branching. It also results that lexical senses expressed based on basic concepts become more precise and readable. For example, both *dog* 狗 and *Beijing dog* 北京狗 are defined as def: {livestock|牲畜} in HowNet and the hypernym-hyponym relation of these two concepts is missing.

Figure 2: Example of a hieratical structure including primitives and basic concepts

In E-HowNet 2.0, {狗|dog} is a basic concept under {livestock|牲畜} and defined as {livestock|牲畜:telic={TakeCare|照料:patient={family|家庭},agent={~}}}. Thus, the top-level definition of *Beijing dog* 北京狗 is as (1a) and can be further extended into the ground level definition as (1b). Such a multi-level representational framework makes sense definitions more precise. It also retains the advantage of using semantic primitives to achieve canonical sense representation. The hierarchical structure of {狗|dog}, {livestock|牲畜} and other related concepts is shown in Figure 2.

(1a) def:{狗|dog:source={北京|Beijing}}
(1b)def:{livestock|牲畜:
telic={TakeCare|照料:
patient={family|家庭},
agent={~}}},
source={北京|Beijing}}}

3.1.2 Hierarchy Structures for both Entities and Relations

We also adjust the ontology structure into two parts. The first part is hierarchy for entities and the second part is hierarchy for relations, i.e. semantic roles. The entity subtree is formed by event subtree and object subtree. The relations include attribute and function.

Entities indicate concepts that have substantial content. By contrast, relations play the role of linking semantic relations between entities. (Chen et al., 2004; Chen K.J. et al., 2005; Chen Y.J. et al., 2005; Huang, Chung and Chen, 2008). Semantic roles also form a hierarchical structure from coarse-grained semantic roles to fine-grained semantic roles. There are 409 relations in E-HowNet 2.0.

Function is a special kind of relation, i.e. a one-to-one relation, in which a concept is mapped onto another concept of the same domain. Rather than establishing the thematic relation or property attribute between two parameters, functions transform a concept to a new concept. Function has compositional property. New functions can be constructed by combining functions of the same domain. For instances, the kinship function of {YoungerBrother({father({x})} denotes 'younger brother of x's father (叔父)' and the direction function of {north({east({place|地方})})} denotes 'the direction of north-east (東北方)' Both are compositions of basic functions.

3.1.3 Uniform Representation for content words and function words

The sense of a natural-language sentence is the result of the composition of the senses of constituents/words and their relations. Conventional linguistic theories classify words into content words and function words. Content words denote entities and function words mainly mark grammatical functions. Actually, there is no clear-cut distinction between the two classes. Therefore, by adding the hierarchy for relations, E-HowNet provides a uniform representation for both function words and content words and enable the capabilities of semantic composition and decomposition. An example is given below to demonstrate the semantic composition process under the framework of E-HowNet.

(2) Because of raining, clothes are all wet. 因為下雨，衣服都濕了

In the above sentence, '*wet* 濕', '*clothes* 衣服' and '*rain* 下雨' are content words while '*all* 都', '*le* 了' and '*because* 因為' are function words. Their E-HowNet sense representations are shown in Table 1. The difference of their representation is that function words start with a relation but content words have under-specified relations.

Word	POS	Definition
because 因為	Cb(conjunction)	cause ={ }
rain 下雨	VA(intransitive verb)	{rain|下雨}
clothes 衣服	Na(common noun)	{clothing|衣物}
all 都	Da(adverb)	quantity={complete|整}
wet 濕	VH(state verb))	{wet|濕}
le 了	Ta(particle)	aspect={Vachieve|達成}

Table1: Sense definitions for each constitute in sentence(2)

If a content word plays a dependency daughter of a head concept, the relation between the head concept and this content word will be established after parsing process. Suppose that the following dependency structure are derived after parsing the sentence (2).

(3) S(reason:VP(Head:Cb: 因為 |dummy:VA: 下雨)|theme:NP(Head:Na:衣服) | quantity: Da:都 | Head:Vh: 濕|particle:Ta:了)。

After unification process, the following semantic composition result (4) is derived. The representations of dependency daughters became the feature attributes of the sentential head 'wet|濕'.

(4) def:{wet|濕: theme={clothing|衣物},
aspect={Vachieve|達成},
quantity={complete|整},
cause={rain|下雨}}.

In (4), function word '*because* 因為' links the relation of 'cause' between head concept '*wet* 濕' and '*rain* 下雨'. The result of composition is expressed as cause(wet|濕)={rain|下雨}.For the sake of notational convenience, the head argument of a relation is omitted. Therefore cause(wet|濕)={rain|下雨} is expressed as cause={rain|下雨}; theme(wet|濕)={clothing|衣物} is expressed as theme={clothing|衣物} and so on.

3.1.4 Correspondence between Attribute Types and Value Types

Some attributes may have specific range of values. For instance, values of color are red, blue, and yellow etc. In E-HowNet, attributes and their respective values are constructed in parallel. Such information is very useful in identifying semantic relations between two constituents while doing semantic composition.

3.2 Rich Lexical Information for Automatic Semantic Composition

A lexical word may play different syntactic and semantic functions and ambiguously denote many lexical concepts. Therefore, in E-HowNet, each lexical concept of a word is identified and provided with its sense definition, English translation, part-of-speech. To facilitate automatic semantic composition and language understanding, E-HowNet 2.0 ontology provides additional lexical information other than conceptual definitions and part-of-speeches. Operational expression, event frames, semantic functions etc. are provided for lexical entries to facilitate semantic composition processing.

3.2.1 Operational Expression

A lexical word may play different syntactic and semantic functions and ambiguously denote many lexical concepts. Therefore, in E-HowNet, each lexical concept of a word is identified and provided with its sense definition, English translation, part-of-speech, and major semantic functions. To facilitate automatic semantic composition and language understanding, E-HowNet ontology provides additional lexical information other than conceptual definitions and part-of-speeches.

For instance, *orange* 橙色 is a "ColorValue|顏色值" but may play different grammatical functions such as subject/object, predicate, modifier. If *orange* 橙色 plays the role of object such as in (5), the sense definition should be applied in the composition process. However, in (6), *orange* 橙色 plays the role of modifier so operational expression should be applied. Possible lexical features for *orange* 橙色 are shown in Table 2.

(5) I like the color orange. 我喜歡橙色

def:{FondOf|喜歡:experiencer={speaker|說話者}

content={color({柳橙|orange})}}

(6) orange flowers 橙色的花

def :{flower|花:color={color({柳橙|orange})}}

<table>
<tr><th colspan="2">詞彙訊息</th></tr>
<tr><td>詞彙:</td><td>橙色</td></tr>
<tr><td>詞性:</td><td>Nad</td></tr>
<tr><td>英文意涵:</td><td>orange color</td></tr>
<tr><td>Event Frame:</td><td></td></tr>
<tr><td>Primitive relations:</td><td>implication={and({yellow|黃},{red|紅})}</td></tr>
<tr><td>Conflation of events:</td><td></td></tr>
<tr><td>定義式: [編輯]</td><td>{color({柳橙|orange})}</td></tr>
<tr><td>操作式:</td><td>color={color({柳橙|orange})}</td></tr>
<tr><td>語義功能:</td><td>ColorValue|顏色值</td></tr>
<tr><td>語義特徵:</td><td>+color</td></tr>
<tr><td>展開式:</td><td></td></tr>
<tr><td>WordNet 自動連結:</td><td>{orange.n.02, color.n.01, color.n.03, color.n.08, color.n.02}</td></tr>
</table>

Table 2: The encoded information for *orange* 橙色 in E-HowNet

3.2.2 Event Frame

Other than basic semantic expression, we like to know its event frame (i.e. arguments) while it plays the predicate role. Arguments of each event type are provided. Take {buy|買} as an example shown in Table 3. Whenever the event "buy" occurs, agent (buyer), theme (commodity), and source (seller) indicated in the event frame of {buy|買} must participated in it. They are crucial to establish relations between constituents of a phrase/sentence and are necessary elements for doing semantic composition.

<table>
<tr><th colspan="2">義原訊息</th></tr>
<tr><td>Primitive:</td><td>buy|買</td></tr>
<tr><td>定義式:</td><td>{take|取:means={pay|付}}</td></tr>
<tr><td>Event frame</td><td>ACTOR:buyer{agent},THEME: commodity{theme},SOURCE:seller{source}</td></tr>
<tr><td>Primitive relations:</td><td>implication={pay|付}; same event={sell|賣}</td></tr>
<tr><td>Conflation of events:</td><td>agent({buy|買})=target({sell|賣}); theme({buy|買})=theme({sell|賣}); source({buy|買})=agent({sell|賣})</td></tr>
<tr><td>語義功能:</td><td></td></tr>
<tr><td>語義特徵:</td><td></td></tr>
<tr><td>WordNet 連結:</td><td>buy.v.1,(同義)</td></tr>
<tr><td>WordNet 自動連結:</td><td>{get.v.01, buy.v.01, entity.n.01, abstraction.n.06, communication.n.02}</td></tr>
<tr><td colspan="2">[編輯]</td></tr>
<tr><td colspan="2">Event Frame</td></tr>
<tr><td>EHowNet Frame:</td><td>agent,possession</td></tr>
<tr><td>TreeBank Frame:</td><td>{agent,theme}{agent,goal}</td></tr>
<tr><td colspan="2">[編輯]</td></tr>
</table>

Table 3: Information table of { buy|買}

3.2.3 Semantic Function

A lexical sense may have different meaning facets. For example, {老師|teacher} in E-HowNet is a subcategory of {專業人士|professional} therefore a hyponym of {human|人}. However, 'teacher', also denotes a kind of occupation and should be regarded as an 'occupation value' as well. Thus, we mark the semantic function of 'teacher' as {OccupationValue|職業值}to include both meaning facets. The same phonamenon occurs for most subnodes of {專業人士|professional}, so we simply mark the semantic function of {專業人士|professional} as {OccupationValue|職業值} and subnodes of {專業人士|professional} will inherit the feature automatically.

3.2.4 Other Semantic Links

E-HowNet ontology is constructed by is-a relation which has the inherent property. Hyponym concepts inherit the properties of hypernym concepts. There are also many other important relations other than is-a relation among concepts. We can look back to Table 3 to see what the relations might be. The primitive relations of {buy|買} are "implication={pay|付}" and "same event={sell|賣} ". That means whenever a event "buy" occurs, that imply the event "pay (money)" happened as well. The event "buy" and event "sell" are actually the same event but just mentioned from different participators' points of view. Since they are the same event, we can also derive the conflation of events of {buy|買} are: agent({buy|買})=target({sell|賣}); theme({buy|買})=theme({sell|賣});

source({buy|買})=agent({sell|賣}). Those conflations are indicated with lexical entries in E-HowNet 2.0.

4. On-line Systems of E-HowNet 2.0

The current E-HowNet ontology shown on the web is the result of automatic constructed by a computer program according to the pre-defined hierarchical structure of primitive and basic concepts as well as E-HowNet expressions, which contain more than 88,000 lexical senses. Based on this system, the E-HowNet Relation Database is also constructed to provide a new direction of clustering concepts.

4.1 Automatic Ontology Reconstruction

To construct a complete lexical taxonomy, we use a strategy that categorizes concepts automatically (Chen et al, 2010).

Step 1. Attach lexical senses. Words and associated sense expressions are first attached to the top level ontology nodes according to their head concepts. For instance, the head concept of the expression '{choose|選擇:manner={cautious|慎}}' is 'choose|選擇'.

Step 2. Sub-categorization by attribute-values. Lexical concepts with the same semantic head are further sub-categorized according to their attribute values. Lexicons that have the same attribute values share specific characteristics; therefore further sub-categorization is performed based on the distinct attribute-values of the lexicons.

Step 3. Repeat step 2 if there are too many lexical concepts in one category. Although the lexicons are classified after step (2), some sub-categories might still contain too many lexicons. In this situation, we further classify the lexicons in the sub-category with other attribute-values until all sub-categories contain fewer members than a predefined threshold, or all members of a category are synonyms.

In case of revisions of lexical sense expressions or nodes of conceptual hierarchy, the ontology reconstruction system may re-attach each lexical entry to appropriated ontological nodes and results a new ontology. For instance, 貓頭鷹 'owl' is defined as {bird|禽 :predication={SelfMove|自移 :duration={night|夜},theme={~}}} and we can find several similar words defineds as the same way. Therefore, {貓頭鷹|owl} is chosen as basic concepts under {bird|鳥} and lexical entries with the above definition are all redefined as {貓頭鷹|owl} and placed in the same subcategory of {bird|鳥}.

However, some concepts do not have natural head (hypernymous) concepts and it is problematic for classification. For example, relations such as kinship relations (e.g. father's younger brother 叔父) and directions (e.g. northeast 東北方) are not suitable to be defined by their hypernyms but the compositions of basic functions. In E-HowNet 2.0, we set rules to classify words of this type according to their first function. Therefore, "father's younger brother 叔父" and "northeast 東北方" are attached to {YoungerBrother|弟} and {north|北} respectively.

Some attribute-type and value-type words are not distinguishable due to having the same sematic head and need to be differentiated by marking the semantic function. For example, "price 價位" is defined as def:{price|價格} and "mid-priced 中價位" is defined as def:{price|價格 :value={intermediate|中等}}. We need to mark semantic function of "price 價位" as ATTRIBUTE and "mid-priced 中價位" as PriceValue|價格值 then they can be attached to appropriate position in the ontology.

4.2 The E-HowNet Relation Database

With the rapid development of semantic networks, related search tools have progressively emerged. Users can set query criteria to find words that match the condition. In Chinese WordNet[2], the interface allows users to enter a keyword and the result shows both lexical meanings and semantic relations of that word. In addition to word senses and relations, Extended-HowNet also clearly presents the position of the word in the ontology. Take *bird* 鳥 for an example, the search results of Chinese WordNet and Extended-HowNet are shown in figure 4 and figure 5 respectively.

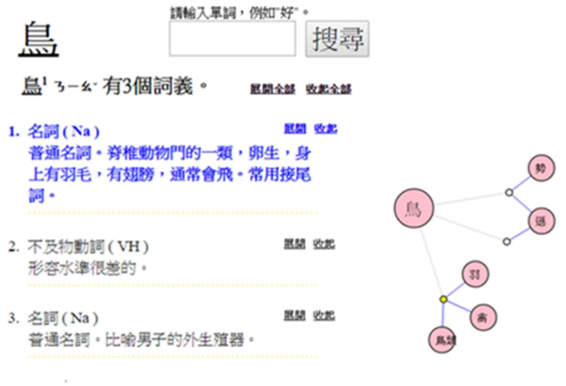

Figure 4: 鳥(bird) in Chinese WordNet

Figure 5: 鳥(bird) in E-HowNet

[2] http://lope.linguistics.ntu.edu.tw/cwn/query/

However, this keyword-search method cannot succeed in finding semantic relation among entities. We advocate that a system should be more flexible in searching specific semantic relations and words should be able to further classify into categories according to their semantic relations. For example, if a user want to find “all entities that contain the function of protection”, or “all entities that denote some kind of protectors”, 墨鏡“sunglasses” (tool to protect eyes) and 專利“patent” (rights to protect intellectual properties) are the possible answers for the former, and 護花使者“lady’s escort”(human to protect the female) and 保 鏢 “guard for goods/persons in transit”(human to protect goods/persons) are for the latter.

EHowNet Relation Database

Host Attribute Value
顯示欄位：
限制條件： protect * *
提交

共108筆 2頁 上一頁 第 1 頁 下一頁

<table>
<tr><th>NO.</th><th>Host</th><th>Attribute</th><th>Value</th><th>Frequency</th></tr>
<tr><td>1</td><td>protect|保護</td><td>instrument</td><td>tool|用具</td><td>24</td></tr>
<tr><td>2</td><td>protect|保護</td><td>agent</td><td>human|人</td><td>19</td></tr>
<tr><td>3</td><td>protect|保護</td><td>patient</td><td>human|人</td><td>9</td></tr>
<tr><td>4</td><td>protect|保護</td><td>manner</td><td>biased|不公正</td><td>8</td></tr>
<tr><td>5</td><td>protect|保護</td><td>patient</td><td>country|國家</td><td>7</td></tr>
</table>

Figure 6: Lexicon categories with host of ‘protect

From the word similarity point of view, the degree of similarity for 墨鏡“sunglasses” and 專利“patent” should not be high no matter which ontology is applied, for the former is a concrete object but the latter is an abstract one. Therefore, their distance in an ontology is also far from each other. However, they could be dynamically clustered to a category while certain semantic constrain is applied.

EHowNet Relation Database

<table>
<tr><th>NO.</th><th>WORD</th><th>POS</th><th>EHOWNET</th></tr>
<tr><td>1</td><td>介胄</td><td>Nab</td><td>{tool|用具:telic={protect|保護:instrument={~},while={fight|爭鬥}}}</td></tr>
<tr><td>2</td><td>甲胄</td><td>Nab</td><td>{tool|用具:telic={protect|保護:instrument={~},TimePoint={fight|爭鬥}}}</td></tr>
<tr><td>3</td><td>耳栓</td><td>Nab</td><td>{tool|用具:telic={protect|保護:instrument={~},patient={耳朵|ear}}}</td></tr>
<tr><td>4</td><td>耳塞</td><td>Nab</td><td>{tool|用具:telic={protect|保護:patient={耳朵|ear},instrument={~}}}</td></tr>
<tr><td>5</td><td>帙</td><td>Nab</td><td>{tool|用具:telic={protect|保護:instrument={~},patient={publications|書刊}}}</td></tr>
<tr><td>6</td><td>面罩</td><td>Nab</td><td>{tool|用具:telic={protect|保護:instrument={~},patient={臉|face}}}</td></tr>
<tr><td>7</td><td>書套</td><td>Nab</td><td>{tool|用具:telic={protect|保護:instrument={~},patient={publications|書刊}}}</td></tr>
<tr><td>8</td><td>針箍</td><td>Nab</td><td>{tool|用具:telic={protect|保護:instrument={~}}}</td></tr>
<tr><td>9</td><td>蛇籠</td><td>Nab</td><td>{tool|用具:telic={protect|保護:instrument={~},patient={地基|ground}}}</td></tr>
<tr><td>10</td><td>頂針兒</td><td>Nab</td><td>{tool|用具:telic={protect|保護:instrument={~}}}</td></tr>
</table>

Figure 7: Lexicons with host of ‘protect’, attribute of ‘instrument’ and value of ‘tool’

Such dynamic semantic clustering search can achieve a comprehensive hierarchical overview for words with the same sematic relation and provide a practically useful new query tool for lexical semantic studies. Therefore, we also developed the E-HowNet Relation Database[3] to achieve this pupose. Taxonomically unrelated but conceptually related concepts can also be computably associated through their E-HowNet definitions. Words with the same semantic relation should be able to group together no matter how far the distance is from the ontology point of view. An example of the E-HowNet Relation Database is given as figure (6) and (7) below.

Figure (6) shows lexicon categories with host of ‘protect’. Once clicking the first category, the system will list all lexicons with host of ‘protect’, attribute of ‘instrument’ and value of ‘tool’, as shown in Figure (7).

5. Conclusions and Future Work

HowNet proposed a new model to represent lexical knowledge, inspiring us to expand this framework to achieve the task of mechanical natural language understanding. E-HowNet confines each concept to a semantic type and defines the relation between these types. E-HowNet has a uniform representation system for both function words and content words to achieve semantic composition, such that meaning representations for morphemes, words, phrases, and sentences can be uniformly represented under the same framework. New concepts can be defined by previously known concepts and definitions can be dynamically decomposed into lower level representations until the ground-level definition is reached. In E-HowNet 2.0 we reorganized the hierarchical structure of primitives and basic Concepts. Near-canonical representation thus can be achieved at a suitable level of representation for synonyms or paraphrases. We also suggested compositional functions to extend the expression of new concepts and make word and phrase definitions more detailed and accurate.

To facilitate automatic semantic composition and language understanding, E-HowNet 2.0 provides additional lexical information other than conceptual definitions and part-of-speeches. Operational expression, event frames, semantic functions etc. are provided for lexical entries to facilitate semantic composition processing.

The E-HowNet 2.0 ontology online is able to demonstrate the taxonomy, sub-categories, and lexicons in a hierarchical tree structure. In addition, we provide a new direction for clustering concepts. Taxonomically unrelated but conceptually related concepts can also be computably associated through their lexical definitions.

There are still many obstacles to achieving the goal of automatically extracting knowledge from language. Apart from sense disambiguation, discord between syntactic structures and their associated semantic representations is another critical problem. To reveal all fine-grained semantic relations for constituents at different levels of syntactic structure, we had just start the project of E-HowNet SemBank annotation. Gap filling processes, as discussed, need to be an integral part of the mechanism. Normalization of sense representation to achieve real canonical sense representation and fine-grained semantic representations are also indispensable. Our future research will continue to address these issues.

[3] http://ckip.iis.sinica.edu.tw/~mhbai/relation/

6. Bibliographical References

Dorr, Bonnie J., Gina-Anne Levow, and Dekang Lin, 2000, Construction of Chinese-English Semantic Hierarchy for Information Retrieval, Workshop on English-Chinese Cross Language Information Retrieval, International Conference on Chinese Language Computing, Chicago, IL, 187-194.

Chen, Keh-Jiann, Shu-Ling Huang, Yueh-Yin Shih, and Yi-Jun Chen, 2005, Extended-HowNet- A Representational Framework for Concepts, OntoLex 2005- Ontologies and Lexical Resources IJCNLP-05 Workshop, Jeju Island, South Korea

Chen, Keh-Jiann, Shu-Ling Huang, Yueh-Yin Shih and Yi-Jun Chen 陳克健, 黃淑齡, 施悅音, 陳怡君. 2004. Multi-level Definitions and Complex Relations in Extended-HowNet 多層次概念定義與複雜關係表達—繁體字知網的新增架構 Workshop on Chinese Lexical Semantics, Beijing University.

Chen, Wei-Te, Su-Chu Lin, Shu-Ling Huang, You-Shan Chung, and Keh-Jiann Chen, 2010, E-HowNet and Automatic Construction of a Lexical Ontology, the 23rd International Conference on Computational Linguistics, Beijng, China.

Chen, Yi-Jun, Shu-Ling Huang, Yueh-Yin Shih and Keh-Jiann Chen 陳怡君, 黃淑齡, 施悅音, 陳克健. 2005. Semantic Representation and Definitions for Function Words in Extended-HowNet, 繁體字知網架構下之功能詞表達初探 Workshop on Chinese Lexical Semantics, Xiamen University.

Chung, You-Shan, Shu-Ling Huang and Keh-Jiann Chen. 2007. Modality and Modal Sense Representation in E-HowNet. Proceedings of Paclic 21:136-145.

CKIP. Semantic Roles and Semantic Role Labeling. Technical Report no.15-01

Dong, Zendong. 1999. Bigger Context and Better Understanding -- Expectation on Future MT Technology, Proc. of the International Conference on Machine Translation & Computer Language Information Processing 17-25.

Dong Zendong, and Qiang Dong. 2006, HowNet and the Computation of Meaning. World Scientific Publishing Co. Pte. Ltd.

Dowty, David R. 1989. On the semantic content of the notion ‘thematic role'. Properties, types and meanings, vol. II, ed. by Barbara Partee, Gennaro Chierchia, and Ray Turner, 69-130. Dordrecht: Kluwer.

Huang, Shu-Ling, You-Shan Chung and Keh-Jiann Chen. 2008. E-HowNet: the Expansion of HowNet, the First National HowNet Workshop. Beijing, China.

Liu, Qun, and Sujian Li 劉群, 李素建. 2002. Word Similarity Computing Based on How-net 基於《知網》的辭彙語義相似度計算. In *International Journal of Computational Linguistics and Chinese Language Processing* 7(2):59-76.

Pustejovsky, James. 1995. The Generative Lexicon. Cambridge: MIT Press.

Schank, R., 1975, *Conceptual Information Processing*, Elsevier Science, New York.

Wierzbicka, A., 1972, *Semantic Primitives*, Athenaum, Frankfurt.

7. Language Resource References

Fellbaum, Christiane. WordNet. http://wordnet.princeton.edu/

Fillmore, Charles. FrameNet. http://www.icsi.berkeley.edu/~framenet/