

## A Double Semantic Network for Intelligent Search

Guogang Li 1st

College of Science, Hebei University of Science and Technology, Shijiazhuang 050018, China

Guoqiang Li 3rd

Science and Technology Information Institute of Shijiazhuang

Tao Feng 2nd

College of Science, Hebei University of Science and Technology, Shijiazhuang 050018, China

Hua Wang 4th

North University Park

**Abstract**—The traditional semantic network retrieval has the following two urgent problems to be solved . The first problem is that the two concepts on the surface look very relevant, but vocabulary used to describe the two concepts has no contact, and in fact ,the two concepts are very poorly correlated. The second problem is that the two concepts on the surface do not seem related, but many words used to describe the two concepts have the same or related aspects, and the two concepts actually have great correlation. We call the two kinds of phenomena “misjudge of the relationship” between the two concepts. The retrieval mechanism studied in this paper aims to synthesize direct relations and internal relations between concepts, and to establish a higher intelligent double semantic network retrieval of concept network and sememe network.

**Keywords**- Double Semantic; network; Concept Network

### I. INTRODUCTION (HEADING 1)

The traditional intelligent search based on semantic network only stays in the “concept” level, which knows that some concepts have direct correlation only through the classification tree, but does not know the specifically related degree, and does not understand true inner meaning of the concept [ 1 ]. Therefore, system can’t backtrack to the result orderly according to true inner correlation of the concept, and the intelligent degree is not high either.

We also know that “concept” is described by a “knowledge representation language”, and this kind of “knowledge representation language” is combined by a series of “vocabulary”. To describe two totally different concepts may also use some of the same vocabulary or related words, and this is the intrinsic correlation of two concepts [2].

Therefore, the following two questions become the urgent problems. The first problem is that the two concepts on the surface look very relevant, but vocabulary used to describe the two concepts has no contact, and in fact ,the two concepts are very poorly correlated. The second problem is that the two concepts on the surface does not seem related, but many words used to describe the two concepts have the same or related aspects, and the two concepts actually have great correlation. We call the two kinds of phenomena “misjudge of the relationship” between the two concepts.

The retrieval mechanism studied in this paper aims to synthesize direct relations and internal relations between concepts, and to establish a higher intelligent double semantic network retrieval of concept network and sememe network.

### II. SEMANTIC NETWORK

#### A. The Representation Method of Semantic Network

A semantic network method is a knowledge representation of artificial intelligence [ 3 ].

The representation method of semantic network shows that human knowledge system becomes network, and makes the associative reasoning play an important role and lay a solid foundation for complex reasoning [ 4 ]. The relationship between the various concepts of the semantic network is mainly represented by the following predicates : IS-A (representing “specific – abstract” relationship ), PART-OF (representing “overall –part ” relationship), IS ( representing that a node is the property of another section point), HAVE (“possession”relationship) ,BEFORE/AFTER/ AT (representing the order of things) and LOCATED-ON (representing the location of things)

#### (1)The Binary Semantic Network Representation

A simple fact can be expressed with two nodes and one arc. For the semantic network of possession relations, a permit node can express not only an object or a group of objects ,also the case and the action. Each case node may have a plurality of outward arc (examples arc) to illustrate the various variables related to the examples.

For instance, LIPING is a man which be expressed as ISA with predicate logic . As is shown in the following figure with semantic network [ 5 ].

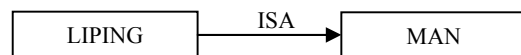


Figure 1. Figure The Binary Semantic Network Representation

A semantic web is a network structure and nodes are connected by chains. In essence [ 6 ].

the semantic network can only represent a binary relation. If the fact represented has multiple relationships, they can be changed into a combination of a set of binary relations or

conjunctive binary relation[7]. Specifically, multiple relationships  $R(X_1, X_2, \dots, X_n)$  can be changed into  $R_1(X_{11}, X_{12}) \wedge R_2(X_{21}, X_{22}) \wedge \dots \wedge R_n(X_{n1}, X_{n2})$ .

Because a semantic network can capture the true proposition, it is often regarded as assertion. It can be inferred that "LIPING IS A MAN" is a true proposition through the semantic network.

In a semantic network, the relationship provides the basic structure of the organizational knowledge, so it is very important but has no relationship [8]. Knowledge is only a collection of unrelated facts [9]. If there is a relationship, knowledge is a cohesive structure which can launch other knowledge [10]. Therefore, a semantic network can represent the relationship of property inheritance, supplement, variation and refinement between things.

**B. Concept Network**

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. If you are using A4-sized paper, please close this template and download the file for A4 paper format called "CPS\_A4\_format".

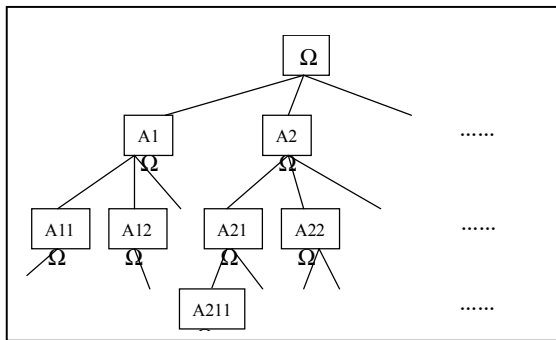


Figure 2. diagram of semantic network

In figure 1,  $A_i$ ,  $A_{ij}$  represent respectively certain concept (node) with a collection  $\Phi$ . For instance, the concept of "computer" points to the following collection:

$\Phi$  ("Computer") = {computer, computer, micro computer .....}.

It is visible that  $\Phi$  has established a set of description elements for each concept, and each description element is described between synonymous relations, which can conveniently support synonymy extend retrieval. There is no limit to the number at the same time, allowing add, delete, modify, and other operations at any time according to the need. The connection between each node represents the relationship between concepts, and for convenience,  $\Omega$  is used to represent the uppermost layer of the virtualization layer. This semantic network establishes the upper and lower layers of the relationship between the concepts by the tree method. The upper concept is a generalization of the common attributes of all of its sub-concepts and sub-

concepts are the specialization of their parent concept from different angles. All the sub-concepts of the same parent concept form the equal brotherhood relationship, and then relationship is added on the basis of the classification tree to link independent tree together. Therefore, this network is called conceptual network.

**C. Sememe Network**

The "concept" is described by "knowledge representation language" whose vocabulary is defined as "sememe", which is "the smallest unit of meaning" to describe a "concept".

Sememe has a lot of classifications, such as events, entities, attributes, attribute values, the number, the number of values, the secondary features, all of which are basic sememe. In addition, there is grammar, dynamic roles, and dynamic attributes, etc.

Grammar refers to "grammar sememe" used to describe grammatical features of the words. The relationship sememe of dynamic role and dynamic properties is used to describe the relationship between different concepts.

On the one hand, sememe is the most basic unit to describe the concept. On the other hand, there is a complex relationship within it. Relationship between sememe includes hyponymy, synonymy, antonymy, attribute of host, part of the overall, material to finished-product and event-role relationship

Actually, it is a complex network structure rather than a simple tree structure, which is known as sememe network.

Hyponymy is the most important relationship in sememe, based on which all the basic meaning compose the sememe level system. As shown in Figure 2:

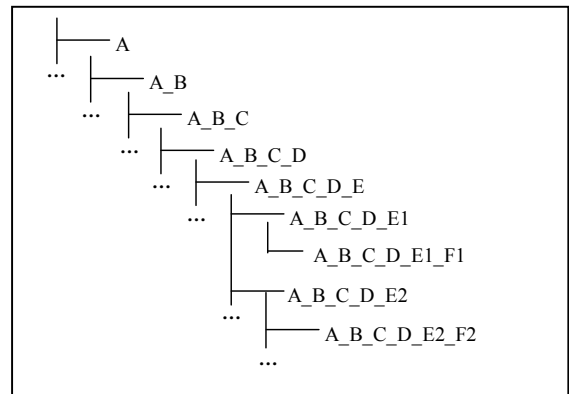


Figure 3. Sememe hierarchy diagram

For a specific database, it is necessary to select the specific sememe to build the specific meaning of the original network, in which we use a series of "sememes" to describe a concept.

Thus the sememe network has also become sub-network of the concept network. The former is used to describe the intrinsic relationship between concepts, while the latter is used to describe a direct relationship between the concepts. Based on both of the two networks, a double semantic network for intelligent search is established.

### III. THE CALCULATION OF THE CORRELATION DEGREE

#### A. The Degree of Correlation in the Concept Network

The calculation the correlation degree of two words in the tree method depends on the length of the path between the two nodes, and here we use a simple way to calculate the distance. Suppose that the distance between two words is  $dt$ , and the correlation of the words in the classification tree can be calculated by the following formula:

$$Sim(W_1, W_2) = \frac{\alpha}{dt + \alpha} \quad (1)$$

In the formula,  $Sim(W_1, W_2)$  is the correlation degree between the two words, and  $W_1, W_2$  represent two words.  $dt$  is the distance between  $W_1$  and  $W_2$  in the classification tree, which is a positive integer, and  $\alpha$  is an adjustable parameter.

In practical applications, in addition to the node distance, we should also take into account the depth of the node in the classification tree. Two nodes within the same distance depart from each other if they are nearer to the classification tree roots.

Suppose the depth of the root is 0,  $t_1$  and  $t_2$  respectively represents the depth in the classification tree, then the formula mentioned above can be improved as follows:

$$Sim(W_1, W_2) = \frac{\alpha}{dt - (t_1 + t_2) / 2 + \alpha} \quad (2)$$

#### B. Semantic Degree of Correlation

For two words like  $W_1$  and  $W_2$ , if  $W_1$  has  $n$  kinds of meanings (concept):  $S_{11}, S_{12}, \dots, S_{1n}$ ;  $W_2$  has  $m$  kinds of meanings (concepts):  $S_{21}, S_{22}, \dots, S_{2m}$ , then we prescribe that the semantic relevance between  $W_1$  and  $W_2$  depends on the maximum of each degree of correlation, that is:

$$Sim(W_1, W_2) = \max_{\substack{i=1, \dots, n \\ j=1, \dots, m}} Sim(S_{1i}, S_{2j}) \quad (3)$$

In this way, we put the problem of relevance degree of semantic correlation of two words down to the problem of the correlation between the two concepts.

#### C. Degree of Correlation in Sememe Network

Suppose  $dp$  be the distance of the original path between two sememes in this hierarchy, we can get the semantic distance between the two sememe:

$$Sim(P_1, P_2) = \frac{\beta}{dp + \beta} \quad (4)$$

Letters  $P_1$  and  $P_2$  represent two sememe; letter  $dp$ , a positive integer, refers to path length between  $P_1$  and  $P_2$  in the sememe network;  $\beta$  is an adjustable parameter.

When we use this way to calculate the degree of sememe correlation, we only use the hyponymy relation of sememe. Actually, there exist many other kinds of relationship in this network except hyponymy. If all of these would be taken into account in the calculation, the results may be finer and more accurate.

In addition, in a large repository of knowledge description language, specific word (concept) may appear in the position of certain sememe. So we should also take into consideration such elements as similarity calculation between specific words and specific words, specific words and sememe in the calculation of the correlation. We require:

- The degree of correlation of specific words and sememe shall be treated as a small constant ( $\gamma$ );
- The degree of correlation of the specific words and specific words, if the two words are the same, is 1, or 0 otherwise.

#### D. The Degree of Correlation of the Words

After discussion of the calculation of correlation between words and sememes, semantic relevancy between words can be got .

$$Sim(W_1, W_2) = \frac{1}{2} \left[ \frac{\alpha}{dt - (t_1 + t_2) / 2 + \alpha} + \frac{\beta}{dp + \beta} \right] \quad (5)$$

Letters  $t_1$  and  $t_2$  represent the depth of the word  $W_1$  and word  $W_2$  in the classification tree respectively;  $dp$  refers to path length between  $P_1$  and  $P_2$  in sememe network;  $\alpha$  and  $\beta$  are adjustable parameters.

The above formula treats the average correlation between words in the classification tree and semantic correlation equally. According to the different resource libraries, we can also take two parts in different weights, and then we get:

$$Sim(W_1, W_2) = \frac{\alpha}{dt - (t_1 + t_2) / 2 + \alpha} \times \gamma_1 + \frac{\beta}{dp + \beta} \times \gamma_2 \quad (6)$$

Among which

$$\gamma_1 + \gamma_2 = 1 \quad (7)$$

After reasonable construction of concept network and sememe network, through the calculation of the correlation between the degree of the word, you can achieve a higher intelligent concept of associative retrieval functions, namely the higher intelligent semantic containing extended retrieval, semantic outward expansion retrieval and semantically related extended retrieval.

The nature of retrieval process is the search of the degree of correlation with the concept network and sememe network as well as calculation between that of two words. Effective control can be achieved by retrieving the result set returned by comparison with the established correlation.

#### IV. CONCLUSION

In this paper, the problems of "miscarriage of justice" and low intelligence of traditional semantic network based on intelligent search concept relations are fully considered, and a new way to build the specific sememe network is put forward by describing a "concept" with a series of "sememe". Then a higher intelligence search in double semantic network is established, among which the concept network describes a direct relationship between the concept while the sememe network is used to describe the intrinsic relationship between concepts.

After double networks have been constructed, a specific formula to calculate the degree of correlation between two words is given in this paper so as to achieve higher intelligent retrieval functions of related concepts. Besides, the system can effectively control the result set returned by retrieval according to related orderly feedback set.

#### ACKNOWLEDGMENT

This work is supported by science and technology support program of Hebei Province(12213520D).

- [1] Strogatz SH Exploring complex networks 2001(6825).
- [2] .Chowdhury A;Frieder O;Grossman D;McCabe MC Collection statistics form fast duplicate document detection 2002(02).
- [3] Manku GS;Jain A;Sarma AD Detecting near duplicates for Web crawling 2007.
- [4] Watts DJ;Strogatz SH Collective dynamics of small-world networks 1998(6684).
- [5] Dean J;Henzinger MR Finding related pages in the World Wide Web 1999.
- [6] Manber U Finding similar files in a large file system 1994.
- [7] Simon H A The sciences of the artificial 1996.
- [8] ZHANG Xiangjun;GUI Cahanglin Gene models in intelligent computer aided design 2001(01).
- [9] John F Sowa Semantic Networks 2004.
- [10] Alex Borgida.John Mylopoulos Semantic Networks 2005.