Hanfei Bao: The Theory of Biomedical Knowledge Integration(Ⅰ), Chinese Journal of Medical Treatment, Vol(2),No(21), 1-6, 2003

The Theory of Biomedical Knowledge Integration(Ⅱ) ——What is underlying the domain concepts

Hanfei Bao
Lab of Informationization and Standardization of TCM
Shanghai Univ. of Traditional Chinese Medicine
Shanghai, China, bhf2002@online.sh.cn

Abstract This paper continues to discuss some basic problems related to the biomedical knowledge integration(BMKI) and other knowledge engineerings. The differences between the concepts of information in biomedical area and that described by C.E. Shanonn have been explored. The paper has identified that the biomedical "structures" play a key role in those differences and they are the biggest challenges facing Biomedical Informatics. 27 definitions of the fundamental concepts which may influence and even determine the formation, certainty, attributes of static and dynamical states, relations between, presentations and operation, etc of those domain concepts, physically and mentally, are described afterwards.

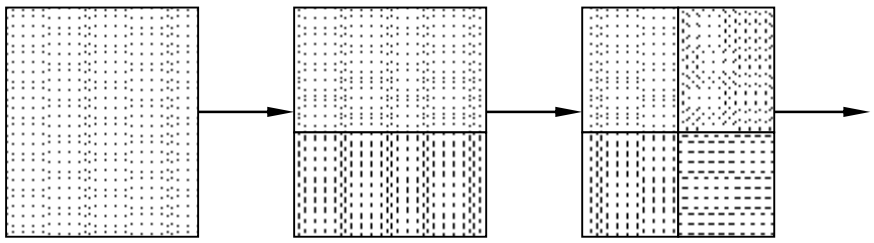
Key words Biomedical knowledge integration(BMKI) Biomedical Informatics Artificial Intelligence Cognition Science

1. A discussion on the information by Shannon and in Biomedical fields

We can't talk about biomedical information, biomedical informatics, biomedical knowledge base and biomedical knowledge integration(BMKI)[1-12] beyond the practical biomedical objects and their characters.M.S. Blois pointed out: "Medicine is complex and the ontology of medical knowledge is multilayered and multifaceted. Medical informatics deserves recognition as a specific discipline (distinct from other forms of informatics) because of the unusual intricacy of the ontologies that drive our systems. The practitioners of our craft need to understand not only the basic principles of informatics in general, but also the details of clinical practice that can make modeling the knowledge of health care such a thorny problem."[13] Similarly MA Musen said, "If there is a slogan that characterizes why informatics is different from computer science, it is 'ours is the discipline that cares about the content.'" [13]. These are the to-the-point analyses of the nature of medical informatics..

As we known, American scientist C.E. Shanonn gave the formula for the information content based on the difference of occurrence probabilities of events. Because philosophically the difference in probability of events, ie the chance-unevenness, is only one kind of unevenness, either entropy presented by German scientist R.E. Clausius and Austrian scientist L. E. Boltzmann or information described by C.E. Shanonn is no more than a kind of evenness or unevenness of a thing. Whereas the evenness or unevenness of a thing is one of the top-level abstracts from or

simplest model of “structures”. Fig. 1 shows the relation between the degree of unevenness of a thing and the complexity of its structure. From the even or non-structural world (if exists), through molecules of inorganic substances, molecules of organic substances, biological large molecules, cell structure, to organism structure, the degrees and the qualities of the evenness or unevenness, and the complexities or information content as well, are very different or, in fact, worlds apart (see Fig 2) from each other. Thus for the highly organized structures such as biomedical structures, entropy by Clausius and Boltzmann and information by Shanonn, described by the chance-unevenness, can give us only very general, most simplified or bottom-level descriptions. Information in biomedical area has new qualities which represents the much higher classes of unevenness rather than that described by chance-unevenness presented by the founders of entropy and information. Namely, it needs new explanation which is related closely with the biomedical structures.



A B C

Fig.1 People often take the unevenness of the distributions of several liquids in a container as the example to explain that in B the information (or entropy) is larger (or smaller) than that in A. Along this train of thought, it is clearly that the increase of unevenness will lead to the more and more complicated structures.

In the <Handbook of Medical Book> edited by JH van Bemmel and MA Musen, there is a description on cognition process in medicine: “the patient or some (biological) process generates *data* that are observed by the clinician. From those data, by the process of interpretation or reasoning, *information* is derived.”, “By carefully studying many such interpretation processes in medicine or by collecting interpreted data from many patients, inductive reasoning may lead to new insights and new *knowledge*. ”[14]

We can see here in biomedicine the information is between the practical biomedical data and practical biomedical knowledge. That means the information is also practical one. As the author‘s

understanding, there is roughly an order like following, *data:* the original records—→*information*: the relations of data—→*knowledge*: the relations of information or the more deep relations of data.

The biomedical knowledge area are such a place where fully exist the serious conflicts between generality and practicality of the knowledge. Lets take the basic theories in biomedical fields as the examples. Contrasted with other sciences, the basic theories of biomedicine such as the central dogma (the theory on protein synthesis directed by NDA), enzyme-catalysis-biochemistry theory, cell-differentiation theory, neurotransmitter theory, second messenger theory, hormonal and humoral regulation, humoral and cellular immunity theory, carrier and receptor theory, ion-channel theory, etc barely have the universality in clinical applications. In other words, in many cases those theories are only something philosophical interpretations, with very limited clinical guide value or clinical validity. They can hardly replace the roles of the empirical or experimental knowledge. Philosophically speaking, it is because that those basic theories are not behaving independently, freely or evenly like X,Y,Z axes doing in Descartes space. They are acting in the joint or structural way or perhaps somewhat like the fractal way(see the following text). Biomedicine is the world filled with the structures of all forms, where the latter plays a fundamental role.

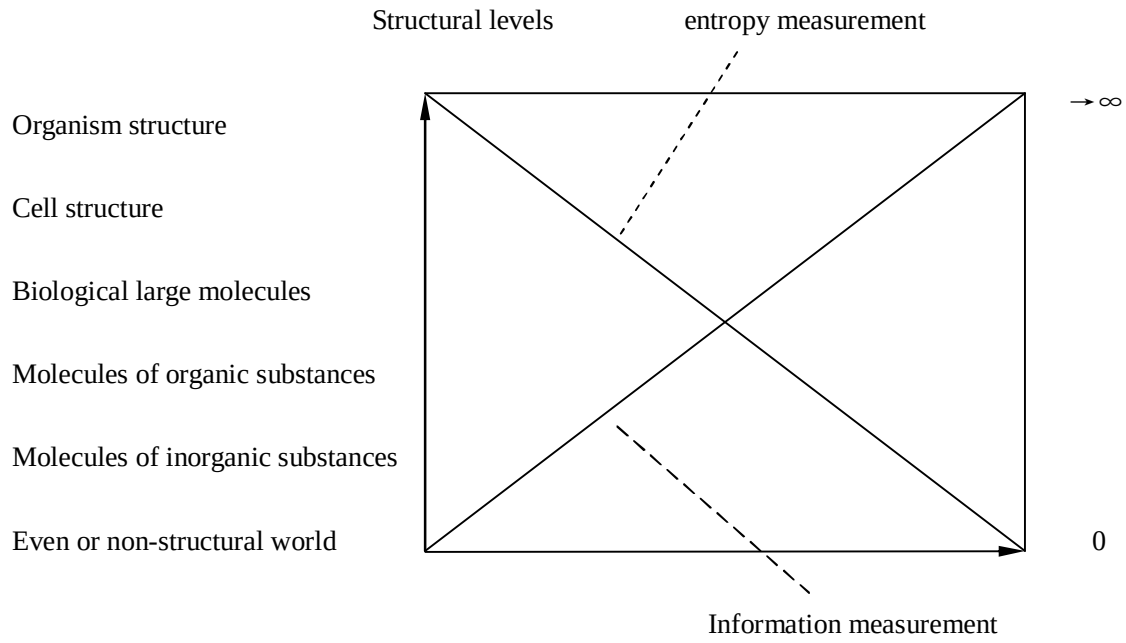


Fig.2 the diagram of the relationships among the structure, entropy measurement and information measurement.

The structures in biomedical area are the reasons why the applicability and explanatory

ability of those basic theories are so severely challenged, doing like the tractors ploughing the lands on the mountains, carefully regulating their each step. The tribal chiefs "experiences" and "experiments" remain the "absolute authority", leaving only small room for reasoning. The kings "scientific theories" nowadays are forced to lower their arrogant heads, paying respect to the structures, complexities and variable individualities of biomedical objects. Any responsible doctors in their work always keep their eyes greatly open, repeatedly and continuously doing their diagnosing-observing-treating circle. That is the actual situation of biomedical information world to which the medical informatics has to face. Consequently we can see the qualities of information there is barely the same as that of the information described by Shannon in general sense.

About the most important philosophic and scientific concept "structure" in biomedical fields, the author would emphasize three points here.

(1) Sciences usually start with breaking up the structures of objects they deal with, trying to understand the attributes or behaviors of the objects through the knowledge on their components. But that goal can quite impossibly be attained, because it is simply the great variety of biomedical (physical, functional or mental) structures, or some kinds of "AND" relations as mathematicians say, create or keep the varied and colourful individualities of biomedical information. Suppose we break up water(H_2O) into hydrogen(H_2) and oxygen(O_2) and we have known all about hydrogen and oxygen, we can't still deduce all the attributes or behaviors of water from those of hydrogen and oxygen. The reason is that contrasted with the elements hydrogen and oxygen water is a structure. Unfortunately, between the structure and its elements there is usually a cognition black-hole or gray-hole which will be described in the ensuing paper. In other words, there is generally no cognition-bridge for understanding the processes of transformation between the elements and their structure, and people usually leap over the cognition-black-holes or get the mapping relations between them, if speaking using a mathematical term.

(2)The behaviors of an object in independent form might be completely different from those in structural or dependent form. For example, a man as an ideal concept is greatly different, as far apart as heaven and earth in fact, from him in a queue, a simplest structure connected by an order relation, to hand the pails of water one after another, in order to put out a disastrous fire. It is merely because that the former is in an independent, freely-existing, unrestricted, full-potential, or somewhat ideal state which perhaps the philosophic ontology seeks after, whereas the latter is contrastedly in a dependent, jointly-existing, integrated or restricted, partial-potential and usually a physical state, serving as an attachment of one structure.

(3) Any discussions solely based on the highly abstracted or general information, if neglecting the basic philosophic concept "structure", are scarcely significant in biomedical area. If Medical Informatics views the biomedical systems, which are such complicated systems and full of various structural semantics, as the simple "throwing dice system", it will be difficult to get into the core of biomedicine, but playing the characters of "tools" or "applications of computer" in these fields. Here we may take electrocardiogram(ECG), B-type ultrasonogram, X-ray and CT in medicine as examples. These new techniques have created much new medical semantic contents which have strongly attracted the eyes of people in medicine on themselves. Consequently nobody in medical area, in fact, minds how the "tools", ie ECG, B-type ultrasonogram, X-ray and CT work. The unparalleled informatics techniques will, undoubtedly, create unparalleled new medical semantics, which, of course, are not those data or information in sense of light, voice, electricity, rays, magnetic fields, but of knowledge, decision making, cognition, recognition and other

intelligences, ie the much-higher-level biomedical semantics. An ancient Chinese poem says “To get a thousand li sight, to go one more floor upstairs” (a li equals half kilometer). In fact, biomedical informatics perhaps needs “to get ten thousand li sight”, thus it has “to go ten more floors upstairs”. That means, for creating incomparable new biomedical semantics, biomedical informatics needs a much longer preparing period and has to pay much more paintaking efforts than other disciplines. The instances such as UMLS, Galen project, various domain ontologies, the non-semantic digitalization of body structure (virtual human body or visible humans projects) and the much more challenging project of the semantic (biologic and medical) digitalization of body structure, the Digital Anatomist[15-16], are the most representative ones of those efforts.

2. The concepts underlying the biomedical knowledge integration

We can regard knowledge as the models of both the objective substantial world and subjective mental world. Whereas at last the knowledge and mental worlds rely on the objective substantial world. The knowledge and mental worlds however can, to certain extent, act and develop independently following their own laws. As the third part, data-information-knowledge system world is the mixture of models of the substantial world and mental world, making knowledge having both physical attributes and mental attributes. It is the topological relationships(see Fig. 3) underlying our knowledge, knowledge engineering and integration.

When we talk about the modeling of biomedical concepts , knowledge engineering and integration, we perhaps can’t forget those more underlying concepts which influence or determine, physically or mentally, the formation, certainty, attributes of static and dynamical states, presentation and operation of and relations among those domain concepts. In the coming text, the author tries to describe those conscious or subconscious mother concepts.

Definition 1 Integral dimension: the factor which can influence a thing independently.

Definition 2 Space: A generalized or imagined circumstance where a thing is controlled by one dimension simply or several dimensions jointly. In 2-D space, both the dimensions are not free individually any longer, but the complex consisting of the two dimensions remains free.

Definition 3 The background space of knowledge(BSK): The total name of the things which are outside a presentation form of knowledge but influential in the meaning of the kno

wledge presentation. In other words, a background space of knowledge is the total of those things on which the exact meaning of a knowledge presentation relies. They can be the generalized dimension space or parameter-space, by which the knowledge has been measured or found out. An example of BSK is the knowledge domain which we are familiar with. Another clearly and rigorously built BSK is the 5-axiom logic space of Euclidean geometry. In the BSK of Newtonean mechanics, the nature of space and time are even, and the substance mass is unchangeable. Whereas the BSK of Einstein’s Theory of Relativity， where the uneven nature of space and time, conservation of light velocity, mutually transform between mass and velocity exist beyond imagination. BSK also includes the view-point or view-field, mental-direction or mental-field of the knowledge user. Let’s take a fact of “a cat caught a mouse” as an example, the meaning of this fact is “a delicious meal”, “the end of life” and “the species balance of ecology ”

for the cat, mouse and the ecologist in the different BSKs, respectively. The default space of general logic(DSL) of human being, depending on which so called the standard or universally accepted knowledge (the various ontologies, for examples) is validated, is the most fundamental one of BSK. DSL is usually something subconscious and is waiting for the experts of AI to explore. Where in DSL the sense organ systems of human body and their working principles maybe play a primary role.

It is perhaps because of the inherent economic feature of our brain, those background spaces usually exist subconsciously, making, to different extent, the experts in the fields of logics or AI to misunderstand that knowledge and its behavior are doing "nudely" (without backgrounds). But when the artificial intelligent systems developed are, at last, applied in the complex situations, especially in area of biology and medicine, they become "miss Lin Daiyu", who is a girl character of famous Chinese ancient novel "A Dream of Red Mansions", being "too weak to withstand a gust of wind".

In a domain, the background spaces of knowledge should be limited in nature, implying that they might be processable.

Definition 4 Generalized experimental background space of knowledge(GEBSK): One of the BSK where the corresponding knowledge was created, ie the integration of all the methods and dimensions by which the knowledge is created or found out. It involves the ways by which people observe, measure and treat the object being explored. It means also that what kind of generalized or particular tools one uses to observe the object. Naked eye, colored glasses, telescope, microscope or other apparatuses?

Another example of GEBSK is that the passengers in a train are moving at the viewpoint of the railway station but moveless for the train itself.

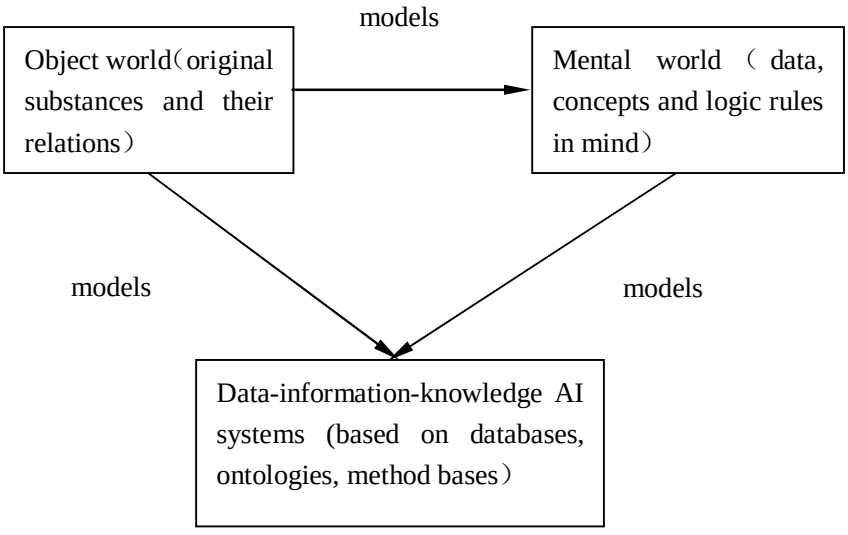


Fig.3 The relationships among original, mental and AI systems worlds. Mental world is the models of original world and AI systems world is the models of both original and mental worlds.

Definition 5 Significant background space of knowledge(SBSK): The subpart or subspace of BSK which is directly related with the goal of one's effort using the knowledge. All other spaces or subspaces rather than significant ones are not SBSK.

Boxing may be an explanatory situation, where in the boxing ring the difference of the points the two boxers have got and the case in which one boxer hit the opponent down, together with body weight rank, rules for the actions, the method for gain in point, etc form the generalized significant space. Whereas the other aspects of the boxers, such as higher or lower education, good or bad looking, having or not married, etc form the nonsignificant spaces.

The significant background spaces of biomedical knowledge are fully diverse, ie they are not unified, resulting in the extreme complexities of biomedical knowledge. That is the diverse-background-space is one of the sources of the complexities and the knowledges where are the knotty challenges to the efforts of biomedical knowledge integration.

Definition 6 Concept: "General notion or idea defining a class of objects", as defined in Handbook of Medical Informatics[14]. More detailed speaking, a concept is the corresponding unit in our mind of a set of objects which are linked together through some relationship(s) and, according to their attributes and behaviors in certain BSK, could be considered as a whole. Thus a concept may be also dependent on BSK, not a "naked" notion.

It seems that nowadays only the brain of human being rather than computer can create concepts.

Definition 7 Structure: The significant and relatively fixed combination of a set of (physical or mental) relations in certain BSK. Thus a structure reflects a kind of physical or mental non-discreteness among a set of relations.

Any two-element-or-over things, if in relatively fixed form, are all structures. The famous three-body problem, most of knowledge integration and knowledge presentation based on framework, for examples, are problems of (static or dynamic) structure.

Definition 8 Mental structure: The structures which exist in mind. They include logical modes, mathematical concepts or algorithms, knowledge presentation forms, etc.

Definition 9 Physical structure: The structures which exist outside mind or in real world.

Definition 10 Natural physical structure: The physical structures formed by non-man-power.

Definition 11 Artificial physical structure: The physical structures formed by (physical or mental) man-power.

Definition 12 The structure of natural physical carrier of human mind: The brain of human being.

Definition 13 The structure of artificial physical carrier of human mind: This concept includes computer, Internet, programs or applications and other artificial intelligent products, excluding human brain.

Definition 14 Mechanism of structure formation: The particular or generalized dynamics of the formation of a structure, including self-organization mechanism, non-self-organization mechanism, self-plus-non-self organization mechanism, etc.

Definition 15 Self-organization mechanism: The internal mechanism of structure formation or information accumulation of the open systems. For examples, the origin and evolution of life or cell, the differentiation and development of biological species.

Generally speaking, for a physical self-organization mechanism, the human's will can only make its influences outside the mechanism, eg on the input or output, etc. For instances, we only can apply fertilizer and pull weeds away for the growth of crops., we can't "try to help the shoots grow by pulling them upward".

Definition 16 Non-self-organization mechanism: The mechanism of structure formation where the external forces (mental, physical, artificial, natural) are the leading factors. The examples of this kind of mechanisms are the manufacture and assembling of various artificial products, such as cars, watches, etc and the structure formation by mind power such as the logic structures of computer and Euclid Geometry and so on.

Definition 17 Mental self-organization mechanism: The mechanism means here the congenital formation of the mental structures and their behaviors.

Definition 18 Mental non-self-organization mechanism: The non-self-organization mechanism where the external force is mental in nature. It is related with those mental structures non-congenitally acquired, learned and formed. Thus it is a special one of the non-self-organization mechanisms. The mechanism forms many ideal structures such as the mathematical structures and algorithms: parallelogram, isosceles triangle, equilateral triangle, circle, etc and various knowledge presentations and their reasoning modes, software systems, abstract problem-resolving methods, etc ie all the mental structure produced by acquired mind power.

In the knowledge presentation based on semantic network, when we express a very common giving-accepting action by saying "Mr Zhang San gives Mr Li Si a book", its structures are like in Fig. 4 and Fig. 5.[17-19] Here we can see that, at bottom, the basic or particular mental (or logic) structures are based on their corresponding physical or practical structure.

Definition 19 Physical self-organization structure: The life systems, the special ones of natural physical structures.

Definition 20 Physical non-self-organization structure: The physical structures formed through non-self-organization mechanisms. Here they mean any physical (natural or artificial) non-life systems, including such as mountains, livers, sea, clock and watch, airplane, train, computer and other products.

Definition 21 Self-plus-non-self-organization mechanism: The mechanism where both internal and external forces jointly drive the formation of structure, ie the man-made-plus-self-organization (or half self- and half non-self-organization) mechanism. Usually the mental force as an external force has been imposed on self-organization mechanism. The mechanisms produce the artificial-plus-natural physical structure. Such structures include tissue-engineering organs, transgenic species, organ or tissue transplantation, plant-grafting, etc.

Definition 22 Controllability of human's willing: The extent to which the human's willing interferes a thing. It can be subdivided into three degrees: the full-, partial- and zero-controllability. The ways of the control of human being over the things may be through the input- and output-interferences and partial or all element-replacements, etc. The degrees to which human being takes part in the structure formation is one kind of reference value of this controllability. The things which are produced by mental non-self-organization or mind-guided non-self-organization mechanisms, ie artificial physical and mental structures, are fully or, at least, partially controlled by human.

Definition 23 Structural dimension: The nature of the dimension is a structure rather a

"mass-point" in generalized sense. It may be viewed as the generalized inertial movement of a structure formation mechanism, such as fractal mechanism described by B. Mandelbrot. As the author's understanding, the continuation of human being, other animals and plants, ie the biogenetic law presented by E.H. Haeckel, is a type of inertial movement characterized by fractal dimension.

Definition 24 Objective thing: Every thing in the objective world.

Definition 25 State: The total of the attributes, relationships and movement forms of a thing. We have two kinds of state, ie the static state and dynamic state.

Definition 26 Static state: The constant state or moveless state of a thing in a significant (or effective, operating) domain of a BSK.

Definition 27 Dynamic state: The variable or move state of a thing in a significant (or effective, operating) domain of a BSK. It could be viewed as the sequences of causality of space, time or other relations. The stability of each step of the sequence is so small that it can't be observed or measured at the roughest level of granularity for the effective BSK, as we often see in films and animated cartons. The aspects of the knowledge, knowledge presentation and knowledge behaviors of the things in dynamic state are much different from those in static state. Imaginably the difficulties of letting a Robot to hold a table and to catch a rabbit are absolutely different, and perhaps in our knowledge-based systems the same problems will wait for us.

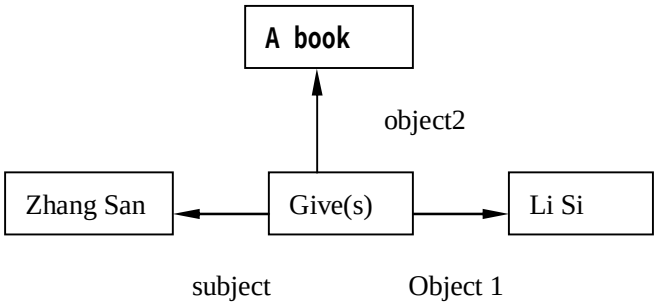


Fig.4 The semantic network structure with movement node

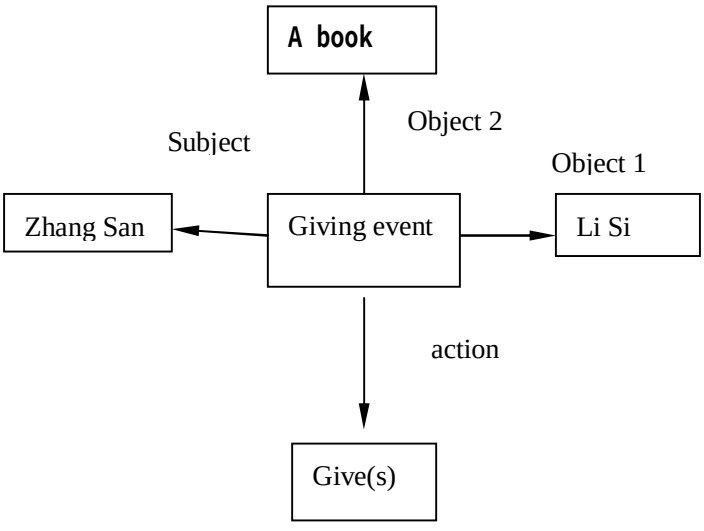


Fig.5 The semantic network structure with event node

(to be continued)

References

1.Bao H.F.: The quantitative integration of biomedical information by computer--New research QMSOC(II), The Proceeding of the First Conference of the Frontier for Life Sciences in Central-south China, 1989, 216-222

2.Bao H.F., Geng J.H. and Su Z.F.: Pansystems Methodology(PM) and a new research on large-scale integration of biomedicine—An introduction of QMSOC and its recent progresses , Acta of Jiansu Industrial College, Journal of Jiangsu Institute of Technology, 1991, 4 (2): 69-75

3.Bao H.F.,Ni X.W.,Lou S.: Integratable Relationized Medical Electronic Book (IRMEB)--- An Exploration of An New Type of Intellegent Knowledge Medium Under the Influence of Pansystems Theory. Advances in Systems Science and Applications (Inauguration Issue) 1995,p304-309

4.Bao H.F.: HCSL:A Human-Computer Commonly Understandable and Communicatable Medical Language, 《Proceedings of The First China-Japan-Korea Joint Symposium on Medical Informatics(CJKMI'99)》, p177-181,1999

5.包含飞：创造一种人机共解互通的医学语言**(**西医学部分**)**，第八届全国医药信息学大会论文集，**CMIA'99(**电子工程师增刊**)**，**1999**，**98**：**18-24**

6.刘沁，包含飞：人机共享语言 **HCSL**（医学）的第二版的计算机实现，中外名医杂志，（**1**）：**78-80**，**2001**

7.包含飞：生物医学知识整合论**(**一**)**，《医学信息》杂志，**16**（**6**）：174-279，**2003**

8.包含飞：生物医学知识整合论**(**二**)**，《医学信息》杂志，**16**（**8**）：**410-415**，**2003**

9.包含飞：生物医学知识整合论**(**三**)**，《医学信息》杂志，**16**（**11**）**:602-605,2003**

10.包含飞：生物医学知识整合论**(**四**)**，《医学信息》杂志，**16**（**12**）：（待发表）

11. Bao H.F.: The Theory of Biomedical Knowledge Integration(Ⅰ), Chinese Journal of Medical Treatment, 2(13): 1-6, 2003

12．包含飞：生物医学知识整合科学思维的成分异源性（I），计算机应用与软件，20（增刊）：171-173，2003

13. Mark A. Musen:Medical Informatics: Searching for Underlying Components, http://www.smi.stanford.edu/pubs/SMI_Reports/SMI-2002-0916.pdf

14．JH van Bemmel, MA Musen:《Hnadbook of Medical Informatics》, Houten:Bohn Stafleu Loghum, 1997

15. C. Rosse and J.L.V. Mejino. A reference ontology for bioinformatics:the Foundational Model of anatomy, http://sigpubs.biostr.washington.edu/archive/00000135/01/Review.pdf

16. P.J. Neal, L.G. Shapiro and C.Rosse: The Digital Anatomist Structural Abstraction: a Scheme for the Spatial Description of Anatomical Entities, http://sigpubs.biostr.washington.edu/archive/00000135/01/Review.pdf

17．王万森：《工智能原理和应用》，北京：电子工业出版社，2000

18．林尧瑞，马少平编著：《人工智能导论》，北京：清华大学出版社，1989

19. 陈世福，陈兆乾等编著：《人工智能与知识工程》，南京：南京大学出版社，1997，109-116