Hanfei Bao: The Theory of Biomedical Knowledge Integration(Ⅵ), Medical Information, 19(2): 183-190，2006

The Theory of Bio-Medical Knowledge Integration(Ⅵ)

Hanfei Bao

Lab of Informationization and Standardization of TCM
Shanghai University of Traditional Chinese Medicine(TCM)
E-mail:bhflhl@yahoo.com.cn,
Web site:
http://www.miforum.net/bhf/english/index_english.htm

Abstract This paper introduced the following new concepts：the cognitive goal, the cognitive goal for the declarative data of the patient records (PRs), The basic attributes of PR's data at the sides of generation, construction and cognition, the generalized data creator (GDC), type Ⅰ to Ⅵ+ of GDC, the cognitive directions of data: forward direction and backward direction, the apparent cognitive orientation and inapparent cognitive orientation, the cognitive granularity difference principle between the natural intelligence and the artificial intelligence, the generalized variable(GVAR) and the generalized value(GVAL), the variable and value transitivity law(V-V transitivity law), the attribute-combination irreversibility between the concept abstracting and embodying, an open model of the launching engine of bio-medical cognition, etc

Key words Medical Informatics Artificial Intelligence The Theory of BioMedical Knowledge Integration(BMKI)

1. A work independent of the particular works of knowledge engineering in BioMedicine

The Theory of BioMedical Knowledge Integration(BMKI)[1-16] considers neither much upon the aspects of syntactic level of information (such as rule-, frame-, object-, ontology-, text-, image-, progress- or video-based knowledge forms), which being the hottest spots nowadays in the topic of knowledge support in the Medical Informatics, nor upon the

development of knowledge or data bases themselves. BMKI focuses very much herself on the semantic behaviors of data, information and knowledge. Supposing one day all the phases of the issues on the information storage, communication, showing, etc ie all those of information-syntax-level, had been fully or perfectly resolved, or in other words, the diverse heterogeneous or hetero-structured data and knowledge bases reciprocally and ideally linked or communicated, or, if I can say, that all the original information always presented to the users according to their wishes and ready to be used, what kind of challenges on the medical knowledge processes would have occurred to them. What sorts of the ***black holes*** would have waited for them. Those considerations make the sights of the Theory of BioMedical Knowledge Integration(BMKI) having constantly or invariably been concentrated on such a mysterious scene. Therefore, it may be said, in a certain sense, BMKI is independent of many activities of the knowledge or data operations in medical informatics today, but relevant to their future.

2. Three basic kinds of attributes of EMR data

1) The generation attributes of data

From a generalized view, we might say that any data or information is created by somewhat data generator. We may farther say that any data generator consists of two parts: generalized measurer and generalized judger(see an example in Fig. 1).

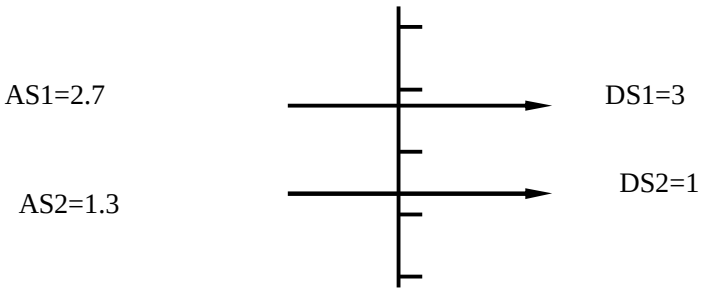


Fig. 1 The measuring step and judging step between the analog and digital signals(A/D conversion）.

Thus about the so called the ***Generalized data creator (GDC)***, we may describe: (1) being composed of two components, ie the generalized observer or measurer and the generalized reader or estimator or judger; (2)In nature, there are several different types of GDC, and the data or

information created by the different types of GDC may be different in their cognitive validity or efficiency.

The GDCs can be essentially classified into two types: the objectives and subjectives and they may generate the distinctive data and information in their certainty or accuracy. The objective GDCs refer to those laboratory devices, equipments, reagents, etc. And the subjective GDCs might be divided further into three subtypes: the ones of the sense organ level, the general psychology level and the scientific domain knowledge level. The validity or efficiency of the semantics of data or information are also diverse in their depth, certainty, granularity, etc something being relatively shallower and the others deeper. The deeper data or information generally more involve the GDC types of the psychology level and the scientific domain knowledge level.

Type Ⅰ of GDC refers to the non-human data acquirement tools such as the laboratory devices, equipments, reagents, etc. The data gained by this type of GDC are more objective or, in some cases, more accurate. For example, "blood urea nitrogen 7.14mmol/L, blood creatinine 88.4 μmol/L, plasma total protein 55g/L, plasma albumin 28g/L, plasma globulin 27g/L, serum cholesterol 12.93 mmol/L";

Type Ⅱ of GDC: The measurements and estimations based on the sense-organ-levels of doctor, producing the data such as "no yellowing of the skin", "no abnormalities found in sense organs", etc

Type Ⅲ of GDC: The measurements and estimations on the psychological levels of doctor;

Type Ⅳ of GDC: The measurements and estimations on the expertise of doctor;

Type Ⅴ of GDC: The measurements and estimations on the sense-organ-level of patients or their family members;

Type Ⅵ of GDC: The measurements and estimations on the psychological levels of patients and their family members;

Type Ⅵ+ of GDC: The measurements and estimations unclearly belonging to the types mentioned above or making no differences for the cognitive goals.

2) The presenting attributes of data

Any knowledge or its unit can be reduced to a relationship or the relationships and the primary kinds of the relationships include three: the 1st one is ***is_a,*** the 2nd is ***-be-*** (generalized) , and the third is the other semantic or predicate relationship, which very likely potentially exists or appears only when meeting the partner of the relationship. And in structure, any knowledge or its unit may be partitioned into three parts, ie infoCarrier, infoVariable and infoValue, as an triple or the nested structure of several triples(for examples see Table 1), resembling the form described by RDF (Resource Description Framework)[17], ie the basic

frame of semantic Web or ontology Web[18]: the triple of resource-property-value.

3)The cognitive attributes of data

The cognitive goal: The problem space plus the method or solution space that people are looking for and trying to resolve.

The cognitive goal for the declarative data of the patient records (PRs): (1) Are there any abnormalities in the patient? (2)If exist, then what are they? (3) If exist but not certainly identified, then how to gain new information? (4) If exist and certainly or uncertainly identified, then how to treat them?

In general, the cognitive attributes of the data in patient record are related to the extents to which the cognitive concepts below can be realized:

Specification or designation: The degree for an information to be designated to an instance. Two granularity degrees may be identified in an designating course (a) a rougher assess, ie to determine the cognitive direction of clinic data;(b) a more refine assess, ie to be particularized.

Certainty: The possibility of the existence of an information, eg "A reactivation of tuberculosis was assumed".

Granularity: The degree of the detail of an information according to a given cognitive goal and it will not be further discussed in the paper, because there have been many discussions already in other publications of BMKI.

For *the apparent cognitive orientation*, the author raised two new concepts:

cognitive forward data: the data supporting apparently the existence of abnormal things, expressed by positive form or negative form.

cognitive backward data: the data supporting apparently the nonexistence of abnormal things, by positive or negative forms or literal expressions.

Inapparent cognitive orientation: the cognitive orientation of data being covert or unclear.

A method of assess of *the cognitive backward data*: Fig.2 shows an example.

3. The formalization of concepts

The work of the human brain is founded on the thinking logic and the computer on calculating logic. The work on the formalization of concepts

originated from a belief or principle bellow:

The cognitive granularity levels are different between the thinking logic(TL) of natural intelligence and the calculating logic(CL) of artificial intelligence.

That is the granularities of the conceptual quanta of TL may be rougher than those of CL, because the knowledge backgrounds space(KBS) of the former may be relatively higher dimensional and the KBS of the latter lower. That means the basic constructive elements for the KBS of TL are usually more compounded than those of CL.

This viewpoint led the work of mine on the formalization of concepts, including the efforts of the meta-dimensional analyses of the general linkages and modifiers of SNOMED[14], recognizing that TL rather than CL is the base of most medical clinic standards, such as SNOMED.

The examples at Table 2 might be explanative:

Table 1 The triple form of information

InfoCarrier	infoVariable	infoValue
-be-		
patient	gender	female
patient	age	55 yr
patient	date of admission	Dec 29, 2000
lung	crackles	yes, bilateral
patient	white cell count	15,500 cells/μL
patient	shortness of brerath	yes
patient	cough	yes
patient	place of birth	rural Jiangsu province
patient	history of travel	denied
other semantic rel		
patient	symptom include	fever
patient	symptom include	abdominal pain
patient	symptom include	abdominal distention
patient	therapy include	sirolimus
patient	therapy include	tacrolimus
patient	therapy include	prednisone(20 mg a day)

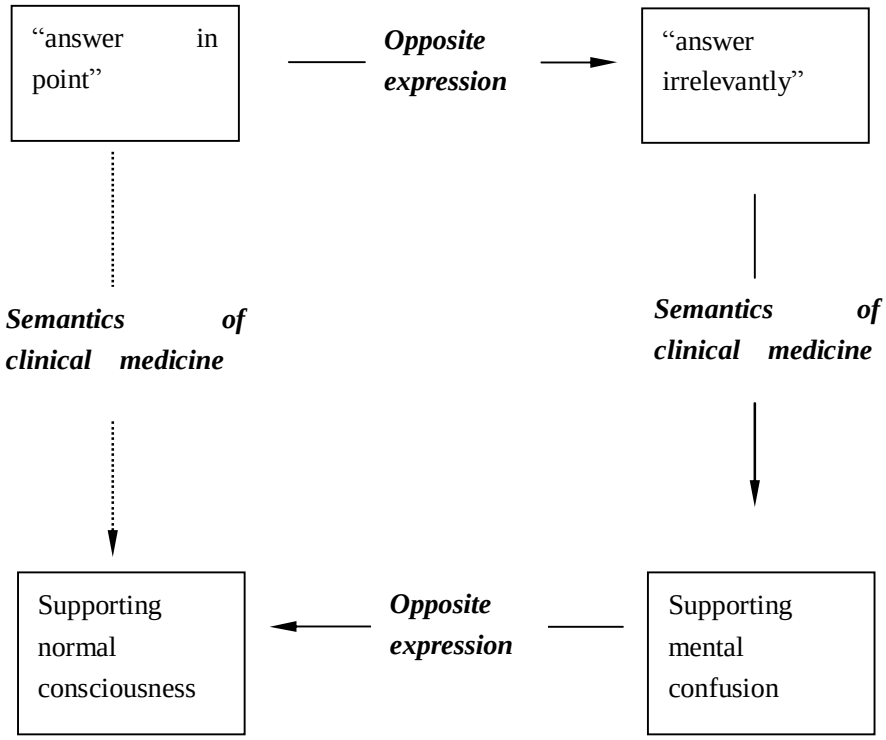


Fig. 2 The clinical semantics of *the cognitive backward data* may usually be deduced by its opposite expression or *the cognitive forward data*, because often the latter has more clarified clinical semantics than the former. Here is an example for this case.

Table 2 The example of relationships of SNOMED and their meta-dimensional combinations.

Original relationships	meta-dimensional combinations
G-A236(Advanced)	***continuity-discontinuity*degree* motility***
G-A336(Expanding), G-A337(Shrinking)	***Size*motility***
G-A127 (Afferent) , G-A128 (Efferent)	***Inside-outside*transporting***
G-A334 (Endogenous), G-A335(Exogenous)	***Inside-outside*generation***

4.The abstracting and embodying of concept

Abstraction of concept is a process to get meta-concepts or meta-class, through continuing to take some attribute or their combination away, ***freely,*** from the definitional attribute set(DAS) of the original concept or class(see Fig. 3). Contrarily, ***Embodying*** of concept means to specify the sub-concepts or the instances, by means of adding some attribute or their combination, ***not so freely,*** to the DAS of the original concept or class(see Fig. 4).

In the courses of abstraction and embodying, there are two important concepts involved:

Generalized Variable(GVAR): The aspects of observation or description of thing.

Generalized Value(GVAL): The specification of the GVAR.

In these courses, things change their roles between GVAR and GVAL. For examples, "(GVAR)which type of student? (GVAL)medical student" → "(GVAR) which type of medical student? (GVAL) medical student majoring pediatrics" →……(see Table 3).

Then we come to:

Variable and value transitivity law(V-V transitivity law): Let $Y_{i+1} = f(Y_i)$ is a hierarchical tree by means of ***is_a*** relationships, Y_{i+1} is a subclass or instance of Y_i, i = 1,2,3,…. then when Y_i is applied as a generalized variable(GVAR), Y_{i+1} would be its generalized value(GVAL). And the same for Y_{i+1} and Y_{i+2}……

More formalized, let A or B be a concept or GVAR or GVAL and ***"A → B"*** means ***(A is a (meta-)GVAR of B)∨(B is a (sub-)GVAL of A)***,then ***(meta-)^{n}concept is a (meta-)GVAR of (meta-)$^{n-1}$concept)∨((meta-)$^{n-1}$concept is a (sub-)GVAL of (meta-)^{n}concept)***.

From Fig. 3 and 4, we may conclude a ***Freedom-Irreversibility*** for working attribute set selection or combination between the courses abstracting and embodying.

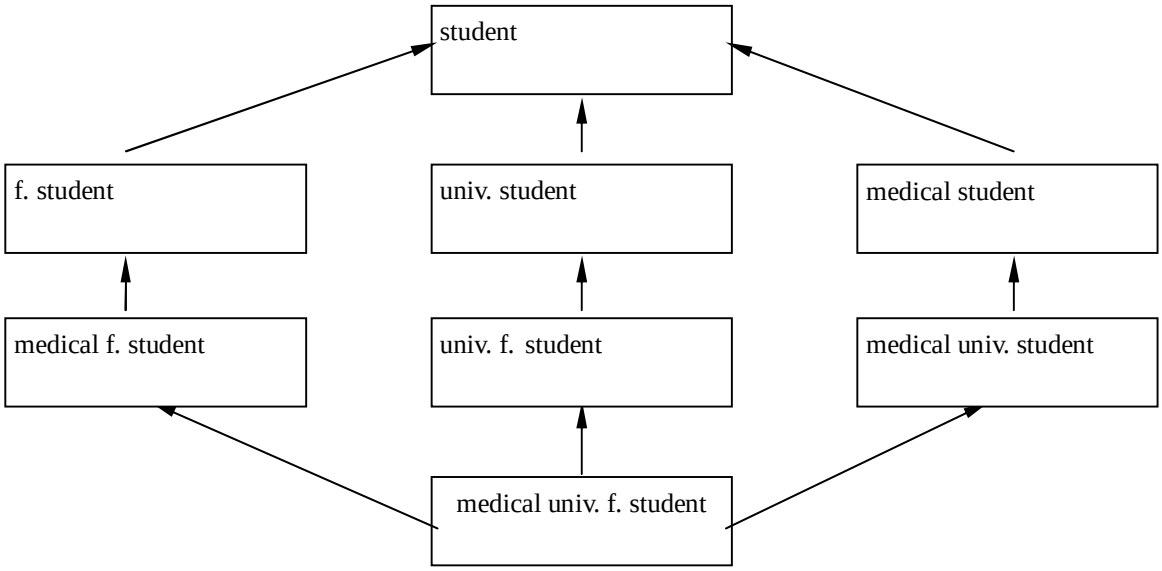


Fig. 3 An abstracting process through "***is_a***" relationship. The attributes "medical", "university (univ.)", "female (f.)" and "student " are freely separated, combined and ordered.

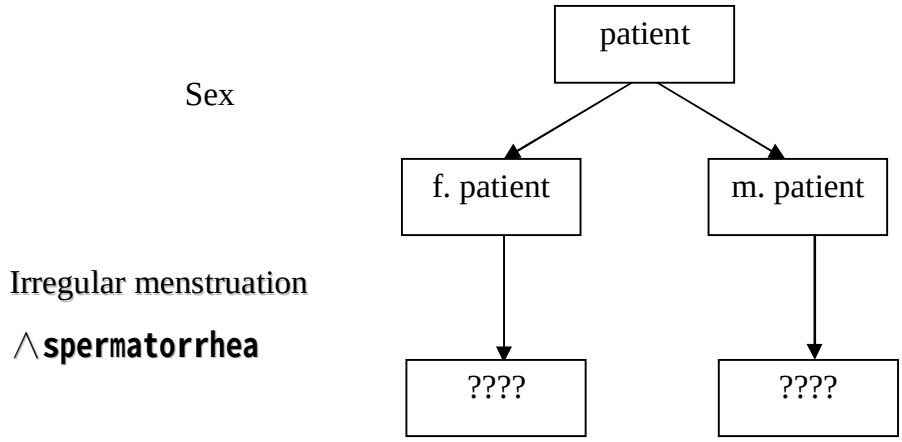


Fig 4 embodying process through adding attribute set, and the latter can't be composed freely.

Table 3 The ***v-v transitivity*** for the instance of abstraction hierarchy in Fig. 3.

GVAR	GVAL
Which kind of student	undergraduate student
Which kind of undergraduate student	male undergraduate student
Which kind of undergraduate student	male undergraduate student
Which kind of male undergraduate student	male undergraduate student of medicine

5. An open model of the launching engine of medical cognition

The author is trying to, in this section, explore the root where our cognitive actions originate from. That is these efforts attempt to explore or build an open model of so called *the launching engine of bio-medical cognitive actions*, which any of our cognitive actions started up from. The idea was triggered by my work on the exploration of meta-dimensions of the general relationships of SNOMED[14]. Word "open" here imply "it is only a proembryo and ready to be modified in public". And it is perhaps not limited to "bio-medical" cognitive actions.

The parts of the results are shown in Fig. 4,5,6.

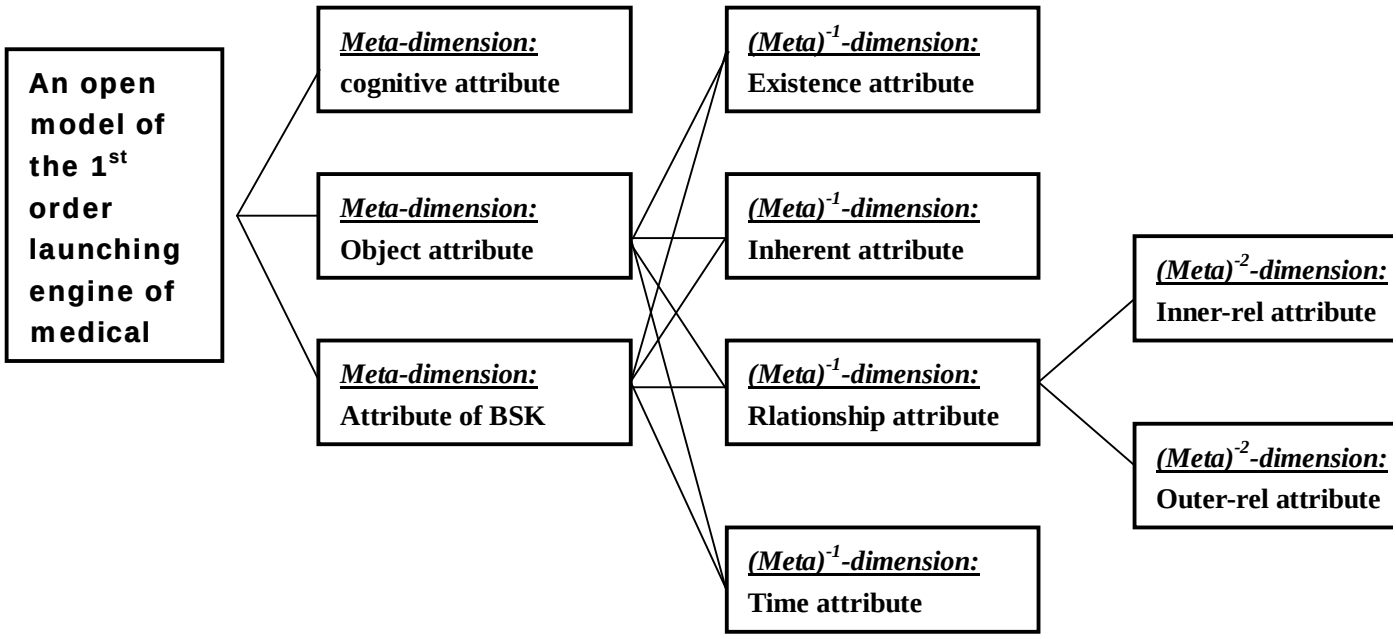
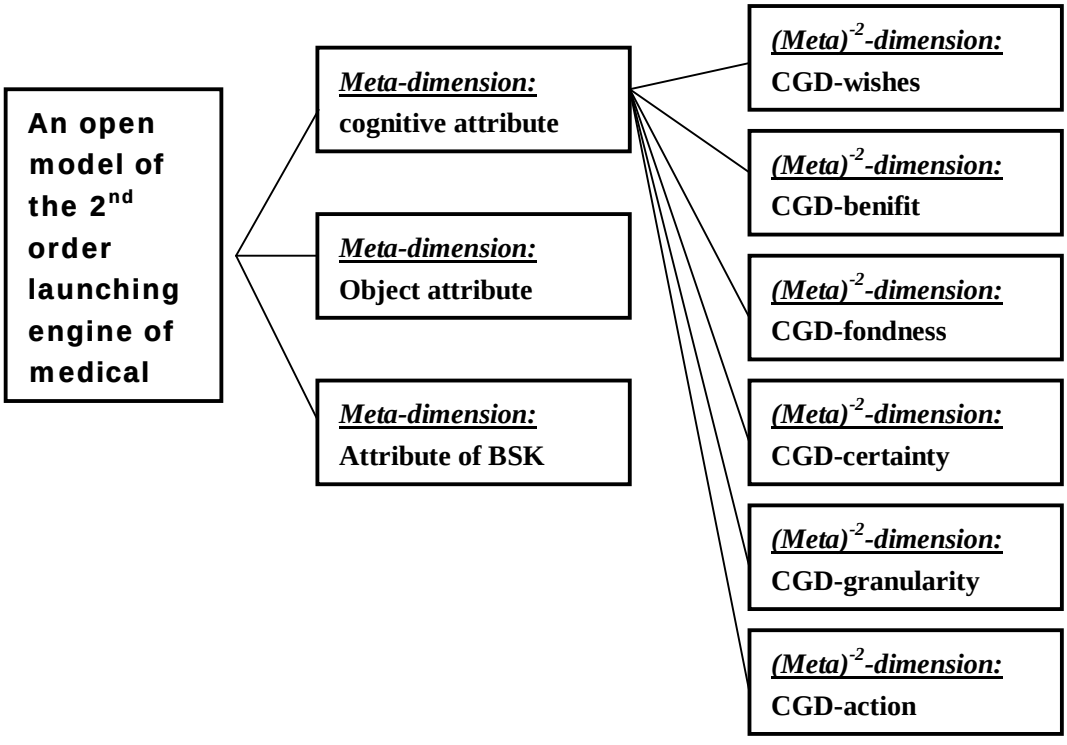


Fig. 4 The 1st order launching engine or ontology of medical cognition，BSK：Background Space of Knowledge.



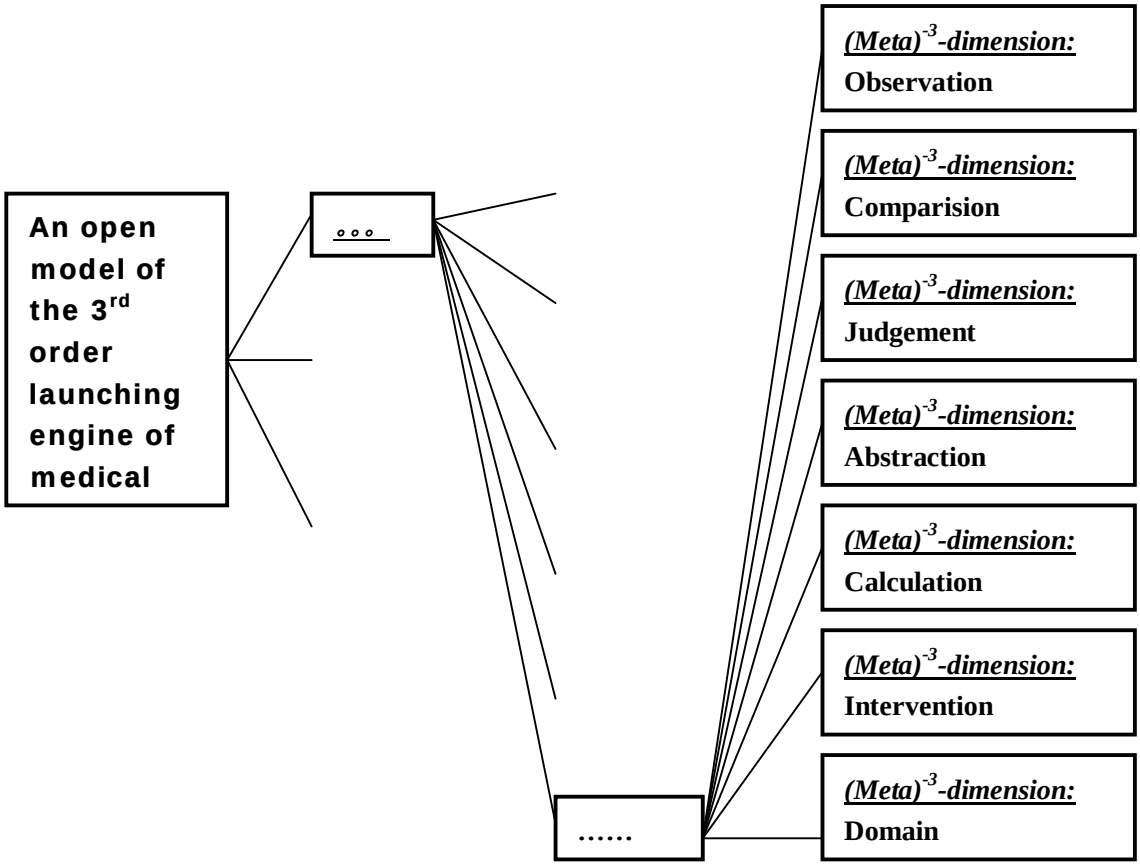


Fig. 6 The 3rd order launching engine or ontology of medical cognition.

References

1.包含飞：生物医学知识整合论(Ⅰ)，《医学信息》杂志，16（6）：174-279，2003
2.包含飞：生物医学知识整合论(Ⅱ)，《医学信息》杂志，16（8）：410-415，2003
3.包含飞：生物医学知识整合论(Ⅲ)，《医学信息》杂志，16（11）:602-605,2003
4.包含飞：生物医学知识整合论(Ⅳ)，《医学信息》杂志，16（12）：666-670,2003
5.包含飞：生物医学知识整合论(Ⅴ)，《医学信息》杂志，17（5）：244-250,2004

6.包含飞：生物医学知识整合论(Ⅵ)，《医学信息》杂志，17（8）：452-457,2004
7.包含飞：生物医学知识整合论(Ⅶ)，《医学信息》杂志，17（11）：685-692,2004
8.包含飞：生物医学知识整合论(Ⅷ)，《医学信息》杂志，18（1）：11-15,2005
9.包含飞：生物医学知识整合论(Ⅸ)，《医学信息》杂志，18（5）：420-424,2005
10. Bao HF: The Theory of Biomedical Knowledge Integration(Ⅰ), Chinese Journal of MedicalTreatment, 2(13): 1-6, 2003
11. Bao HF:The Theory of Biomedical Knowledge Integration(Ⅱ) ——What is underlying the
domain concepts, Chinese Journal of Medical Treatment, 2(21): 1-6, 2003
12. Bao HF: The Theory of Biomedical Knowledge Integration(Ⅲ), Chinese Journal of Medical
Treatment, Vol(3),No(7),2004
13.Bao HF: The Theory of Biomedical Knowledge Integration(Ⅳ), Medical Information, 18(6), 545-550, 2005
14.HF Bao: The Theory of Biomedical Knowledge Integration(Ⅴ)——The meta-dimensions and their combination of the General Linkage/Modifiers of SNOMED, Medical Information, 18(10): 1227-1232，2005
15. Bao HF: Conceptual, Experimental and Compass-beacon Ontologies——The Principles, Concepts and Methodology of BMKI(I), Medical Information, 18(8), 2005
16. Bao HF: The Concepts and Definitions Related to BMKI——The Principles, Concepts and Methodology of BMKI(Ⅱ), the Proceedings of the 7th China-Japan-Korea Symposium on Medical Informatics, pp135-137,Shenzhen 2005, in Chinese Hospitals, Vol.9 supplements, 2005
17.RDF Primer,http://www.w3.org/TR/rdf-primer
18. Fensel D. and Mark A. Musen MA: The Semantic Web:A Brain for Humankind, *http://www.cs.umbc.edu/771/papers/ieeeIntelligentSystems/introduction.pdf*