Knowledge Architecture based on Evidence Based Logical Atomism for Translational Research

Jun Nakaya[†], and Tetsuo Shimizu[††]

[†]Graduate School of Medicine, Kobe University, Kobe, 650-0017 Japan
[††] Institute of Medical Science, University of Tokyo, Tokyo, 108-8639 Japan

Summary
In post-genome research, bridging genomic science and clinical medicine through in silico knowledge is crucial for the translational research. Here the key bridging issue is to establish the logical integration and the bi-directional migration of knowledge based on multidisciplinary information. For this purpose, we propose the Knowledge representation architecture based on Evidence based Logical Atomism (KELA) that has the single and share-able representation as the logically extended anatomically hierarchical structure from genome to human. It makes logical integration possible. Knowledge is represented with the knowledge atom that is a set of logical atom (LA) and its supportive evidences. Evidence and its EBM (Evidence Based Medicine) based quality assessment make the background of LA and its quality clear. KELA is the share-able knowledge representation architecture and have potential to facilitate strategic post genome research or to support translational research.

Key words:
Post Genome, Logical Atomism, Translational Research, Evidence Based Medicine, Knowledge Architecture

1. Introduction

Genomic science is expected to contribute effectively to clinical medicine in post-genome era. For contribution, bridging genomic science and clinical medicine is key issue. Genomic translational research is an interdisciplinary practical research aiming to bridge genomic science and clinical medicine. Clinical objects are to improve QOL (Quality Of Life), clinical safety, and clinical efficiency. Translational research itself has been popular in cancer research; recently the genomic application starts. Here what bridged is knowledge and the essence of bridging is the bi-directional migration of knowledge between genomic science and clinical medicine. The bi-directional migration and bridging needs the integration of these two kinds of knowledge on computer. But it is not easy, because they are multi-disciplinary and have the multi-layered multi-dimensional non-linear relations of cause and effect. Too much and vast background information of knowledge needs IT power to form knowledge. The architecture of knowledge on computer is essential for the integration, the bi-directional migration, the bridging, and to handle vast information. According to apply the methodology of the electronic knowledge technology to the medical field, Stanford SMI has a project named PROTÉGÉ [1]. Based on PROTÉGÉ, SMI develops the EON system [2] for the protocol-based care. Recently the ontology dictionary for genome is constructed with a concept of Genetic Ontology (GO) [3].
Our starting clinical mission is to support the post-genome translational research. We design KELA on the focused area and try to establish the useful architecture for the genomic translational research. The logical atomism is a system theory based on the logical atom (LA) that is a unit of the logical concept [4]. We can say that the genomic science and the clinical medicine are also the system theories based on the logical unit in each hierarchical domain. Therefore the logical atomism can be a pertinent theory to integrate the knowledge. Based on the logical atomism, we build up the knowledge architecture with introducing EBM concept. This paper reports our fundamental knowledge architecture based on the evidence based logical atomism and its prototype for translational research.

2. Methods and model description

2.1 Knowledge Architecture based on Evidence Based Logical Atomism (KELA)

Fig. 1 shows outline of KELA. Logical atomism supports KELA philosophically. KELA has the Logically Extended Anatomically Hierarchical Structure (LEAHS) of knowledge. Knowledge Atoms (KA) are built into LEAHS. KA is a set of Logical Atom (LA) and supportive evidences. The quality of evidence is evaluated with

Manuscript revised January 2006.

modified EBM concept.

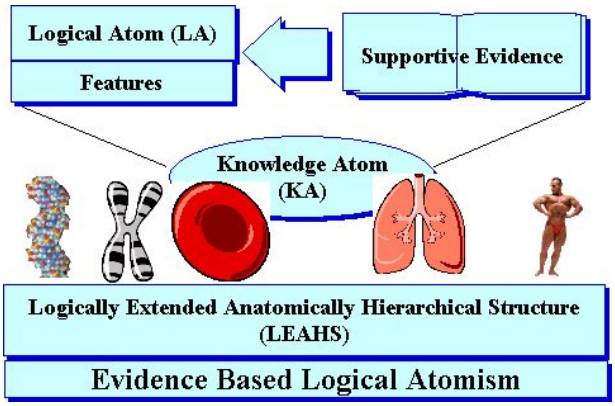

Fig. 1 Knowledge Architecture based on Evidence based Logical Atomism.

2.2 Knowledge structure

KELA has logically enhanced anatomically hierarchical structure (LEAHS). LEAHS is the continuous, seamless, and hierarchical structure that contains both microanatomy and macro anatomy. It means that it bridges the world of DNA, RNA, proteins, micro organella, cell, sub organs as assemblies of cells, organs, individual as assemblies of organs, environment as assemblies of humans. The anatomically hierarchical structure has the easy understandability and the stability as for suffering structural change in the future. This structure indexes knowledge by pointing the KA. We design LEAHS based on almost stable and certain knowledge that is described commonly in plural books.

2.3 Knowledge Representation

Knowledge is described with knowledge atom (KA). A KA is defined as a set of a logical atom (LA) features and its supportive evidences. LA is a minimized conceptual unit in each hierarchy and is described with its features. The features are diversified attributes of LA. The features are normalized within the same hierarchical level and the same domain to avoid repetitions. The evidences that support backgrounds of the features are represented as the references to electrical publications. Classification for quality assessment of evidences is summarized in a table.

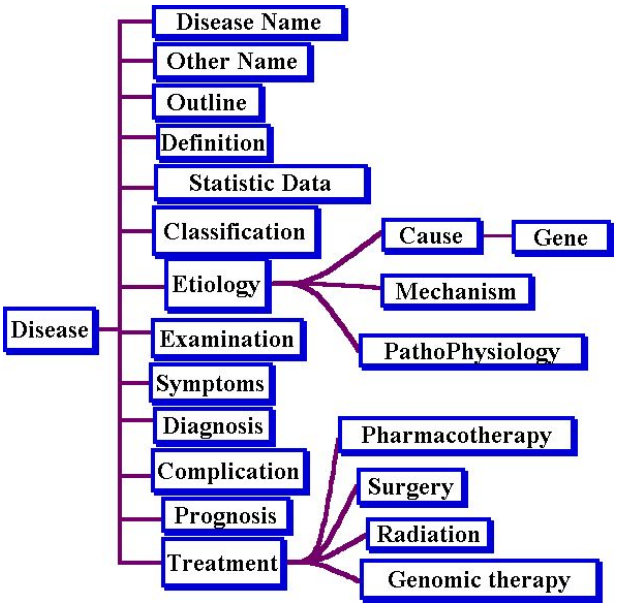

Fig. 2 A part of skeleton for disease

Fig. 2 shows a part of skeleton for disease. We describe knowledge atom of disease according to this skeleton format. Here each disease is KA. LA is a concept of disease. LA is described with features. Evidences support features. As an example, genomic knowledge of entity is inserted in clinical knowledge of disease-etiology-cause-(gene, mechanism). This skeleton structure is normalized until the third level for optimization.

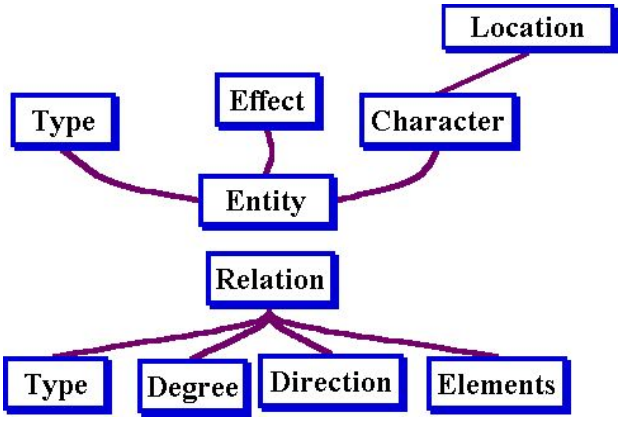

Fig. 3 A part of skeleton for molecular findings

Fig. 3 shows a part of skeleton for molecular findings. Knowledge atoms of molecular findings are described with entity and relation. At this molecular layer each entity and relation is described as a KA. A KA has its LA, and LA has features and evidences. To locate an entity in LEAHS, we describe species, birthplace, and existing place as features in an entity. A relation has features of type, degree,

and direction. These features are based on theory of elementary process described in Feynmann Physics [5].
We also have the program type KA. This type KA can call the external program as the computer simulator or the protocol management tool. The biomedical conditions for the computer simulations are described in the features of the LA. The features of this type LA are the entity, the relation, the motion, and its time course. On the surface, computer simulation seems to be a kind of dynamic LA.
All of the LA is described with text. Considering that the features or the attributes are ultimate essence of human understanding, the diverse representations like image or sound are also converted to text with their features. The knowledge described with words offers the survival format for its succession. The literal representation will be one of the survived IT formats in future.

2.4 Evidences and EBM

The evidences support the features of the LA. Evidences are practically only the links or the references to the evidences, considering that the standard of the quality evaluation of the evidence will change in the course of years. The substantial evidences are on electrical media like Internet. The standards of quality evaluation of evidences are summarized in a table (Table 1: quoted from [6]). This table is based on the EBM classification of evidence [7]. By adding standard of the books evaluation, the classification by the impact factor (IF) and the citation index (CI), this table can give the numerical order of evidences.

3. Knowledge collection

According to LEAHS, we performed the data mining to collect the information on Internet. The gathered information is transformed to the knowledge. The transformation has two stages as the stage of diversifying information and the stage of conversion to KA. The knowledge collection is achieved with locating KA on LEAHS. The resources of knowledge are public databases (Books, PubMed, PDBJ, PIR, Swiss Prot, PCB, Japana Centra Revuo Medicina Database, etc). This time we used some data-mining product (Cell Space, Pathway Assist) [8] in part as trial.

4. Prototype and evaluation

The prototype is concerning about the integrated hematological knowledge. We evaluated the KELA philosophy and its representation through this hematological prototype. This prototype has the integrated hematological knowledge and the handling programs. The deepness of clinical knowledge is almost internship level and the deepness of biological knowledge is almost master degree level. The prototyped computer programs are for searching, processing and input knowledge.

Table 1: Quality assessment standards

Quality	Major class	Sub-major class	Minor class	Sub minor class
High	Book	Many-sided universality	Number of supports	IF, CI
	Paper	Randomized Controlled Trial (RCT)	Multiple support	IF, CI
	Paper	RCT	Meta analysis over multiple RCT's (conclusions may differ)	IF, CI
	Paper	RCT	Single RCT conclusion	IF, CI
	Paper	Non Randomized Controlled Trial (non RCT)	Comparative studies, correlation studies etc.	IF, CI
	Paper	Quasi experimental studies	Cohort studies, case controlled studies, etc	IF, CI
	Paper	Case series, case report, etc	Number of supports	IF, CI
Low	WEB site	Opinions from experienced experts	Comments of committee, etc	IF, CI

5. Results

Fig. 4 shows sample window of KELA prototype. This is concerning about Fanconi anemia and GVHD. Anatomically path is shown in left. Anemia is a Red Blood Cell disease, and Fanconi anemia is a kind of aplastic anemia. In this case Fanconi anemia is defined at DNA level hierarchy. Clicking DNA level Fanconi Anemia pops up the features of Fanconi anemia as the center. The features of Fanconi anemia are described like the outline processor. Double clicking of the Etiology make widening and opening its explanation. Clicking GVHD complication in the Stem cell transplantation of the treatment opens its detailed explanation and the cell relation model. The evidences for clicked entities are shown in right below. The color of the evidence responds to the clicked term.

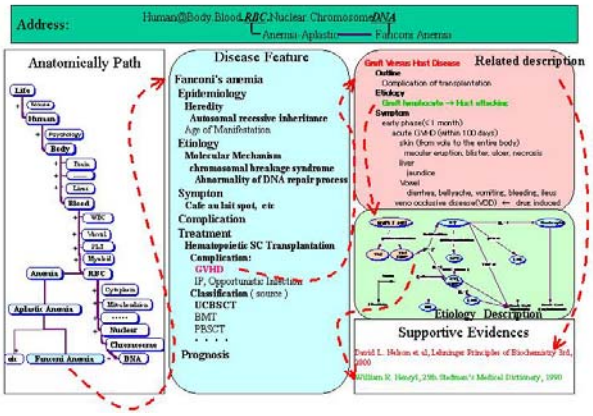

Fig. 4 Sample Window (GVHD)

6. Discussion

The logical integration of the multi-disciplinary discrepant knowledge on computer fills gap between genomic science and clinical science. This makes the knowledge continuous and seamless. The basis of the integration is to describe both kind of knowledge with the mono-scoped view. The logical atomism is the pertinent philosophy to describe both kind of knowledge with the single standard. Defining knowledge as the logical atom according to the LEAHS is the single scoped logical simplification of knowledge. The logically simplified knowledge has the easy understandability and the fast accessibility to the location of the necessary knowledge. This easy understandability and fast accessibility are the advantage points to perform the knowledge simulation and the data mining from the vast information on Internet. Through performing the knowledge simulation according to the LEAHS, we can always check the meaning, the significance or the next step of each simulation from the view of anatomical understanding. The relevant range of knowledge will be cut out at each simulation.

The data mining and the text mining are the methodology for the knowledge collection. Some companies like Adriane Genomics Inc., Cellomics Inc. release the tools for these purposes [8]. The data mining is the strong method, but the accuracy of the interpretation is contrary to the quantity of the extracted information. Presetting structure of the knowledge enables the deep formation of knowledge with holding the accuracy of interpretation. LEAHS enables the arrangement from molecule to human with single view. This singularity has the easy understandability and the location. As for the knowledge collection with KELA, we can deepen and increase knowledge with keeping accuracy. Performing data mining according to LEAHS, deep and quick knowledge collection can be achieved without losing accuracy of interpretation. A view on anatomically hierarchical structure from micro to macro scope is a good choice as a unique scale that can integrate genomic science and clinical medicine.

Storing evidences in the knowledge atom makes EBM practical in both genomic science and clinical medicine. The evidence based knowledge enables the quality assessment and the judgment of the reliability. In big buzz for EBM [9] as the scientific reasoning for the diagnosis or the treatment, the evidence based knowledge will be an answer to a question how we can establish EBM practically in genomic medicine and its translational research. In future our EBM classification table needs more modification to classify common-sense knowledge. In prototype, the knowledge written in books was referred to construct the common-sense knowledge. The common-sense knowledge is classified into single class of "book" that we added to the EBM classification table. More the classification grades are needed to classify knowledge in many books that are the common-sense knowledge.

We can achieve predicting the situation through performing the knowledge simulation. Predicting the immunological reaction and avoiding the critical clinical situations can improve the clinical safety and efficiency. Especially the human experimental research phase like translational research needs avoiding the critical clinical situation. Integrated knowledge will be the basis of prediction through this kind of knowledge simulation. The evidence based knowledge representation makes quality of knowledge clear. The qualified knowledge makes the accuracy of prediction clear. Clearing on the accuracy of prediction for the occurrence of the clinical events leads to optimize the research path and the avoiding the critical clinical situation. Optimization of research path minimizes the wasted resources of research.

The prediction based on knowledge features is a kind of analogy that can propose the possible situational hypothesis from the existing knowledge. The way of prediction is similar to the human thinking way. It is understandable, checkable, and reliable. The expert system, fuzzy inference, and neural network model are kinds of method to process knowledge and to infer situation [10]. The molecular modeling simulation, pathway model, GON [11], and virtual reality are kinds of method to estimate from the point of molecular modeling. Clinical biostatistics is a method to analyze clinical data and to estimate situation with its interpolation [12]. Fig. 6 shows future image of Umbilical cord blood stem cell transplantation. Clinically its goal is to improve Quality Of Life, clinical efficiency, and clinical safety. Ultimately it means avoiding or preventing the critical clinical situation like GVHD, or Graft Failure. To prevent the critical situation, we need the prediction of the future event. The prediction enables the optimized control of the WET devices. It leads to the optimized ex vivo expansion of

stem cells. This enables application of the umbilical cord blood stem cell transplantation to the heavy weight patients. Predicting situation needs the scientifically certified knowledge that is the evidence based knowledge. This certification is based on the scientifically certified publications like papers. The integrated genomic technology like DNA chip, Protein chip are important for the personalized medicine as the integrated scientific input. All of these future therapies are based on the logically integrated multi-disciplinary knowledge.

7. Conclusion

Integration of genomic science and clinical medicine with single knowledge architecture on IT system is the principal demand in genome translational research. The logical atomism is the pertinent theory to describe both kinds of knowledge with the single standard. The logical atomism based KELA architecture enables the logical integration and the bi-directional migration of the multidisciplinary knowledge. LEAHS makes knowledge continuous and seamless from genome to human. Including evidences in KA makes the quality of LA clear. Introduction of evidence concept and its assessment method make the EBM concept effective in translational research.

Acknowledgment

The authors would like to express their cordial thanks to Dr. Mitsuo Ohta for his valuable advice.

References

[1] Musen MA, Eriksson H, Gennari JH, Tu SW, Puerta AR., "PROTEGE-II: a suite of tools for development of intelligent systems from reusable components.," Proc Annu Symp Comput Appl Med Care. 1065, PMID: 7949900, 1994.

[2] Musen MA, "Domain ontologies in software engineering: use of Protege with the EON architecture.," Methods Inf Med., 37(4-5), 540-550., 1998.

[3] Iwei Yeh, Peter D. Karp, Natalya F. Noy, Russ B. Altman, "Knowledge Acquisition, Consistency Checking and Concurrency Contol in Gene Ontology," Bioinformatics, vol. 19, 241-248., 2003.

[4] Russell, B., "Philosophy of Logical atomism. Monist," vol. 28, 495-527, 1918

[5] Richard P. Feynmann, "The Character of Physical Law," MIT, CambMass, 1967.

[6] Nakaya, J., Ohtsuka, Y., Sakamoto, N., Agishi, Y., "Knowledge Framework aiming at Disease Prevention through Logical Translation of the Balneology to the Modern Medicine," J. J.A. Phys. M. Baln. Clim., vol. 68(4), 209-222, 2005.

[7] Pamela J. White, "Evidence-based medicine for consumers: a role for the Cochrane Collaboration," J Med Libr Assoc., vol. 90 (2), 218–222, 2002.

[8] KA Giuliano, R Kapur, J Wang, and ES Woo., "Cellular screening of lead compounds in the post-genome era: An integrated approach to knowledge building from the cellome.," Integrated drug discovery technologies. Marcel Dekker, 2001.

[9] Friedland, Daniel J., "Evidence-based Medicine: A Framework for Clinical Practice," Stamford, CT: Appleton & Lange, 1998.

[10] Fukumura A, Shimura M, Nagao S, Tsuji S, et al, "AI handbook," Japan AI Society, Ohm, 21-66, 1990.

[11] Matsuno H, Doi A, Hirata Y, Miyano S., "XML documentation of biopathways and their simulations in Genomic Object Net," Genome Inform Ser Workshop Genome Inform. , vol. 12, 54-62, 2001.

[12] Spiessens B, Lesaffre E, Verbeke G, Kim K, DeMets DL, "An overview of group sequential methods in longitudinal clinical trials," Stat Methods Med Res., vol. 9(5), 497-515, 2000.

Jun Nakaya received the B.S. and M.S. degrees in Mech. Eng. from Hokkaido Univ. in 1985 and 1987, respectively. After staying in IBM, he received M.D. and Ph.D. degrees from Hokkaido Univ. in 1995 and 1999, respectively. He stayed in M.I.T., Institute of Medical Science, Univ. of Tokyo, and Tokyo Medical and Dental Univ.. Now he is an Associate Professor of graduate school of medicine, Kobe Univ.. He is a member of ISO, HL7, AMIA, ISMH, MIT-J, SSJ, and MSJ.

Tetsuo Shimizu received the B.S. degree from Faculty of Science, Univ. of Tokyo in 1971. After working as a researcher in the National Institute of Radiological Sciences, he worked as a scientific manager in the Fujitsu Limited. His research interest is in translational research and ontology. Now he is a Professor of Institute of Medical Science, Univ. of Tokyo.