Condensed Cross-Clinical Knowledge

Jun Nakaya†, Koji Sasaki††, and Hiroshi Tanaka†††

†Clinical Genome Informatics Center, School of Medicine, Kobe University, 650-0017 Japan
††AdIn Research Inc., 102-0094 Japan
†††Tokyo Medical and Dental University, 113-8510 Japan

Summary
To explore biomedical space as post-genome research, knowledge can be a good navigator and a good exploratory basis. Before combining biological and clinical knowledge, clinical knowledge itself should be integrated with a unique scale within a clinical field. This unique scale should be cross-clinical and integrated knowledge should be condensed cross-clinically. To condensing knowledge cross-clinically, we considered three points. The first is to use the unified and optimized knowledge skeleton, the second is to introduce an unified description rule, and the third is to describe knowledge with standing on verbal side. Disease knowledge is described with its normalized skeleton to describe it with a single and optimized scale. To describe each feature of disease, we set down an unified and intentionally rough description rule that covers all clinical fields. Considering that words based description is ultimate human understanding and standing point on verbal side of ontology, we described all knowledge with words and texts. Each disease knowledge is composed into partially detailed anatomically index. This paper shows its methodology and results.

Key words:
Condensed, Cross-Clinical, Knowledge, Normalization, Hierarchical expression, Logical Atomism, Anatomical Index

1. Introduction

Clinical medicine is expected to blossom its glory with receiving outcome of genomic research in post-genome era. Considering that the essence of fruitful outcome from genomic research is knowledge itself, integration of clinical knowledge from cross clinical fields as internal medicine, surgery, gynecology, etc is crucial as a basis to receive and to utilize genomic outcomes. But different arrangement criteria among clinical fields make integration of clinical knowledge itself difficult. On the other hand, information is increasing rapidly on Internet. It makes clinical knowledge complex and also makes integration more difficult. Ontology is a choice to describe cross-clinical knowledge, but too detailed and strict description rule in ontology tend to make knowledge more complex and difficult.

Manuscript revised July 5 2006.

To integrate vast exploded amount of information and make effective knowledge, condensing methodology of cross-clinical information and transforming methodology to knowledge are essential issue. Simplified but enough and pointed cross-clinical knowledge and its integration are principal and fundamental basis of post-genome research.

2. Background

The knowledge will have an important role to establish next success of public health care through post-genome research as illustrated in NIH Roadmap [1]. Knowledge is being bridged here and the essence of bridging is bi-directional migration of knowledge between genomic science and clinical medicine. But clinical knowledge itself has different arrangement standard and is still fragmented. SNOMED-CT is an integrated knowledge for the clinical terms [2], this integrated clinical ontology is coordinated good but is still have difficulties based on too detailed and too expensive contents. This work tries to condense knowledge among cross-clinical fields with single standard and a concept such as simple, easy, but enough contents.
As concerns applying methods of electronic knowledge technology to the medical field, some tool projects based on ontology like PROTÉGÉ [3] are under way. Based on ontological technology, protocol based care tools like EON [4] are developed. Recently genomic ontology dictionary is constructed as Gene Ontology (GO) [5]. But detailed and precise description rule of these ontological tool make knowledge huge and too segmentalized. Now ontology tool needs effective condensing philosophy of knowledge to integrate for practical use. Besides ontological description rule is far from end-user understanding and end-user understandable representation. To become ontology more popular among domain end-users, more user-understandable representation that has user's viewpoint is principal demand for knowledge processing in post-genome research.

We compartmentalized knowledge as logical conceptual units as reported in previous paper [6]. Logical conceptual unit has its philosophical background in logical atomism as a logical system theory [7]. Logical system theory fits to genomic science and clinical medicine, because they are also kinds of logical system based on logical unit in each respective hierarchical domain. These logical systems need logical consistency and their reason. Focusing on the verbal side of ontology, we try to represent knowledge for practical use among cross-clinical fields.

Practically this knowledge is the basis of supporting post-genome targeting translational research (TR). TR itself has been popular in cancer research, and genome-science based trial has started recently. Genome targeting translational research is a practical interdisciplinary research aiming to bridge genomic science and clinical trial. Since early time its clinical endpoints are to improve QOL (Quality Of Life), clinical safety and clinical efficiency.

3. Methods

3.1 Normalized skeleton for disease

We designed a skeleton to describe the disease knowledge with a unique scale. The skeleton was created through many times of build and scrap process during constructing process of integrated knowledge. The skeleton is normalized until the 3rd level normalization to optimize knowledge in whole medical domain fields. The normalization has done in all cross-clinical domain and is a result of the many times of "build and scrap" processes. Fig. 1 shows a part of normalized skeleton for the inter-clinical diseases. This skeleton has a hierarchical expression in its feature description and can cover all clinical fields. The difference of hierarchies are represented with the difference of indent level. This difference of hierarchies indicate the difference of the grading scale of knowledge. The disease name is defined as the most commonly used name among MDs who are the domain users. Alias name includes synonyms that means the same disease. We distinguished the definition and the concept. A definition is the precise concept that is already accepted among the domain users. This term is used only in case of precise definition exists. In case of definition does not exists, we described the outline of disease concept instead of precise definition. The epidemiology was classified into categories as the frequency of occurence (e.g 1.5 person in 100 thousands, 0.2 %), the sex difference (e.g ♂ : ♀＝9 : 1), the favorite age, the favorite site, the favorite side (including unilateral or bilateral), the heredity (including HLA concern and autosomal recessive inheritance). Etiology is classified into cause, sideration mechanism, and pathophysiology. A cause is a fatalistic condition or a disease that cause the disease directly. A sideration mechanism is a common sideration mechanism following to the cause. A pathophysiology is a bio-response around sideration. As above, we defined the skeleton format for disease.

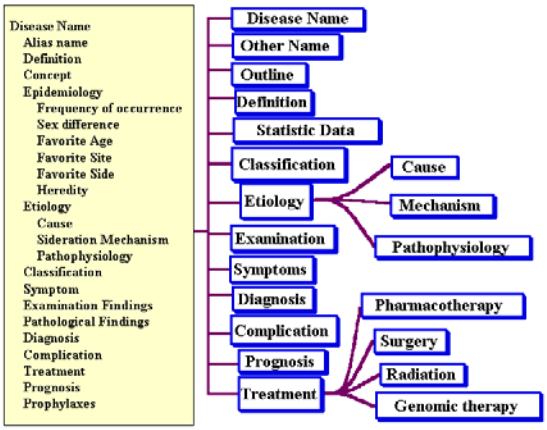

Fig. 1 Normalized skeleton for inter-clinical disease knowledge

3.2 An unified description rule of medical knowledge

An unified description rule covering all clinical fields was introduced to describe whole clinical knowledge with single standard. A part of an unified description rule is shown in Fig. 2. This unified rule is intentionally rough because detailed rule interrupts condensing knowledge. Through build and scrap of constructing integrated knowledge, we learned that the unified description rule should have moderate/optimized roughness. To condense meaning of terms essentially, we placed much value to inevitability and causation at description.

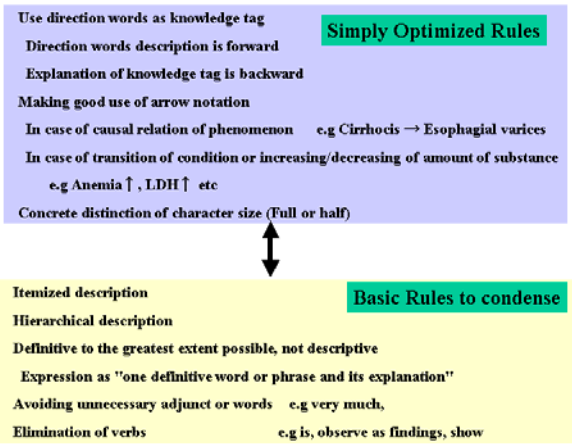

Fig. 2 An unified description rule

3.3 Combined hierarchical index and formulation process

Knowledge of each disease is integrated with indexing its location to anatomical structure. Integrated knowledge has an anatomically hierarchical structure as a single backbone of cross-clinical knowledge (Fig. 3). Disease knowledge is not classified on concept of clinical field but is indexed to anatomically hierarchical structure. At the construction beginning, disease knowledge that are gathered from all clinical fields is scraped into parts according to skeleton format and classified according to rough anatomical index. Then knowledge is gathered again and is built up by embedding into the skeleton format. Knowledge of each disease is indexed to anatomically hierarchical structure with intention to set one fact at one place.

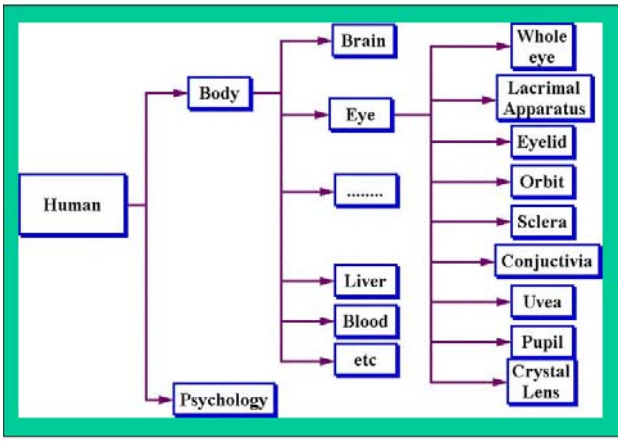

Fig. 3 A part of anatomically hierarchical structure

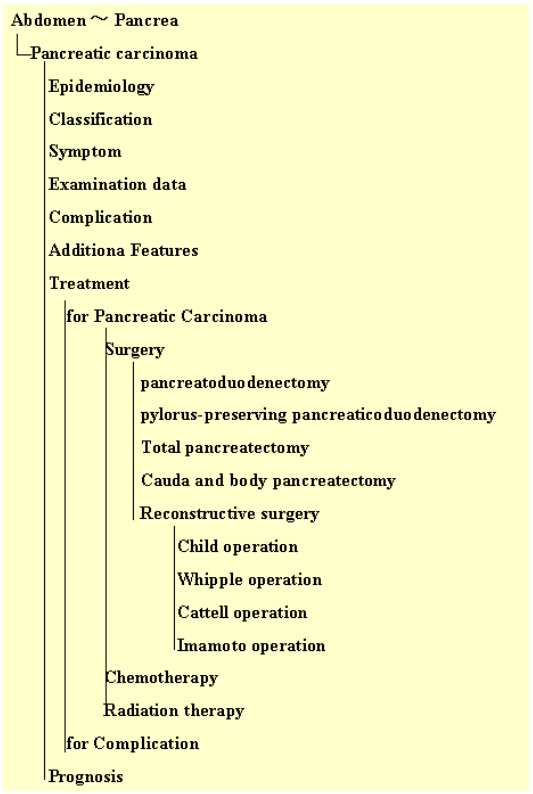

Fig. 4 A part of abdominal disease knowledge

3.4 Text base description

Considering that text base description is an ultimate human understanding, and standing point on verbal side of ontology, we described the cross-clinical knowledge with words based text description. Images or sounds are also described as its features with words (text). The other points of adoption of text description are to minimize the disk space on computer and to choose the survival format in future. We considered that the text format will survive and can be converted easily to the other formats.

4. Results

Fig. 4 shows knowledge for the pancreatic carcinoma that is an abdominal disease. According to the normalized disease skeleton, the feature of disease knowledge are embedded hierarchically. We can find that treatment of this knowledge is cross-clinically collected from internal medicine, surgery, and radiogenics. Treatment approaches for pancreatic carcinoma itself are classified into surgery, chemotherapy, and radiation therapy. The surgery is classified according to the operational methods. The treatment approach for the complication is the other choice of treatment that is a kind of palliative treatment.

Fig. 5 shows a part of anatomical structure. According to anatomical hierarchy, the diseases are classified. Eye is a part of head. Sclera is a part of eye, and scleritis is a disease of sclera. The other scleral diseases are indexed. These diseases have the integrated knowledge collected from not only ophthalmology but also neurosurgery or the other medical area. The integrated knowledge are described according to the skeleton format that is previously introduced.

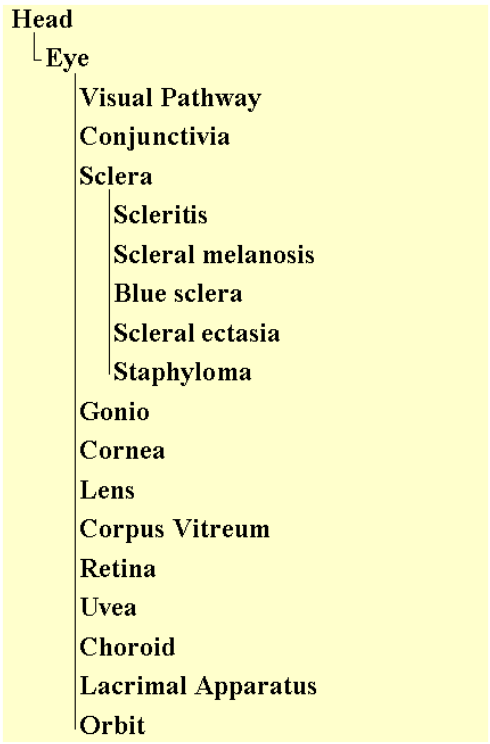

Fig. 5 A part of anatomical structure

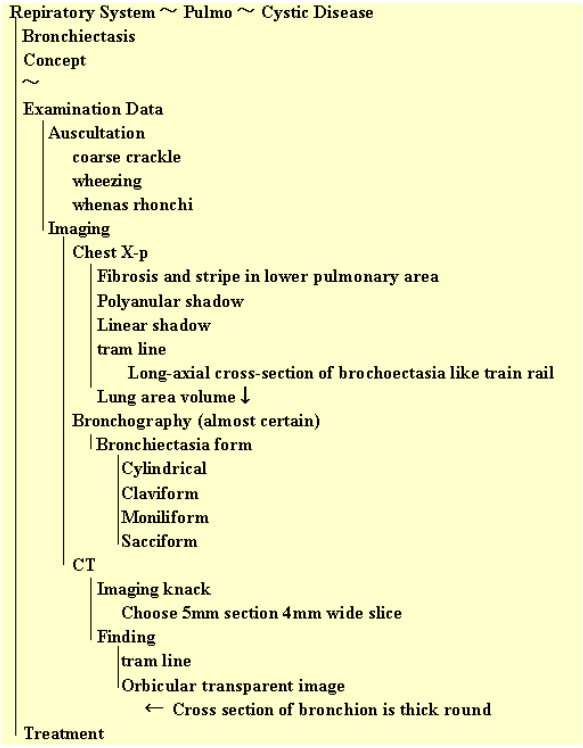

Fig. 6 A part of pulmonary knowledge

Fig. 6 shows a part of bronchiectasia as pulmonary knowledge. We can see the concept, examination data, treatment. The examination data has the terms of auscultation and imaging. The imaging term has the terms of chest X-p, bronchography, and CT. In the examination data, we can see that the features of sounds and images are described with only words (by text only). The reason of "orbicular transparent image" on CT is described with the rule of "Arrow notation" and "hierarchical description". The decrease of "Lung area volume" in "Chest X-p" is also represented with "arrow notation".

Total file size of whole clinical knowledge is 33.6MB in Japanese. In English, it means almost half of this.

5. Discussion

In post-genome research, establishing bi-directional migration and bridging of knowledge between genomic science and clinical medicine are essential. The integration of huge knowledge into an electronic representation on a computer is necessary to establish bi-directional migration and bridging at first. But their integration is not easy because they are multi-disciplinary, fragmented and have multi-layered multi-dimensional non-linear relation of cause and effect. Too much and vast background information of knowledge in this Internet society needs IT power to assemble knowledge. Simplified electronic representation and its condensing on computer are important for practical integration, bi-directional migration, bridging and handling cross-clinical information.

Condensing knowledge among cross-clinical field will be a kind of optimized integration over all clinical fields. For optimized integration we should think a great deal of keeping the meaning of original concepts. Arrangement with keeping the meaning or the taste of original concepts should be done with standing on user's viewpoints and should have user-oriented description. An unified

description with introducing skeleton format and setting down an unique scale are the good helps for that, because this is a kind of shared description format among users in all cross-clinical fields. For domain users, easy comprehensivility means that it is similar to domain user's thinking pathway. We took pain to establishing this easy comprehensibility with skeleton format and description rules through try and error for many years.

In other hands, knowledge described with unified description rule and skeleton format make comparison of knowledge easy. Unifiying description methodology leads to improve objectivity of information. Through normalization we can minimize multiplication of knowledge. Minimizing multiplication of knowledge minimizes computer memory and makes finding the location of knowledge easily for both human and computers.

As for significance of each description rule, itemized and hierarchical description abbreviates sentences. The representation rule such as "one definitive word or phrase and its explanation" makes a sentence simple and patterned for users. Avoiding unnecessary adjunct words makes sentence rational. The elimination of autoptic verbs abbreviates sentence, and the amount of abbreviation in cross-clinical fields is large beyond expectation. Using the direction words as the knowledge tag makes knowledge simple and makes searching easy. Good use of arrow notation (↑、↓、→、←) abbreviates sentences and makes phenomena simply understandable. Concrete distinction of character size (Full or half) is the preparation for conversion to the other format like XML, HTML.

Easy understandable structured knowledge based on anatomical structure makes aquisition of knowledge quick and accurate. For knowledge collection, data mining and text mining are the methodologies. Some companies like Adriane Genomics Inc., and Cellomics Inc. distribute tools for these purposes [8]. Data mining is a powerful method, but accuracy of interpretation is contrary to quantity of extracted information. Presetting structure of knowledge enables deep knowledge formation without losing accuracy of interpretation. Anatomical structure can extend its covering range from human to micro organ without losing logicality. This logically extended structure enables arrangement from molecular to human with single standard. This extension means that hierarchical structure based on anatomy can be a basic backbone of bridged knowledge between genome science and clinical medicine. We have been starting to joint this cross-clinical knowledge with genetic knowledge as a concept knowledge platform [9]. Considering that we can convert this knowledge to onotological knowledge through converting program easily, we can say that this kind of knowledge representation is preprocessed knowledge of ontology.

Technically enriching knowledge concerning about image or sound and updating methodology of established knowledge are the next issues.

6. Conclusion

A normalized skeleton that covers all clinical fields is constructed through many times of "build and scrap" process. An unified description rule that has intentional roughness is set down for condensing knowledge. This rule are also found from the "build and scrap" process. An anatomical structure is a reasonable single standard to integrate all cross-clinical knowledge. The words based text description can be the ultimate format for condensing knowledge.

Acknowledgment

We wish to thank Kohsuke Kudo, Ukihide Tateishi and Rikiya Onimaru to help us in collection of Japanese knowledge. We also wish to thank Prof. Edward Feigenbaum who gave us significant advises.

References

[1] Elias Zerhouni. The NIH Roadmap. Science. 2003 Oct, Vol. 302: 3: 63-72.

[2] Stearns MQ, Price C, Spackman KA, Wang AY. SNOMED clinical terms: overview of the development process and project status. Proc AMIA Symp. 2001;:662-6.

[3] 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. 1994; 1065, PMID: 7949900

[4] Musen MA. Domain ontologies in software engineering: use of Protege with the EON architecture. Methods Inf Med. 1998 Nov; 37 (4-5): 540-50.

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

[6]Jun Nakaya, Tetsuo Shimizu: Knowledge Architecture based on Evidence Based Logical Atomism for Translational Research, International Journal of Computer Science and Network Security, vol. 6-2, ISSN: 1738-7906, 175-179, 2006

[7] Russell, Bertrand. Philosophy of Logical atomism. Monist, 1918; Vol. 28.

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

[9] Jun Nakaya, Tetsuo Shimizu, Hiroshi Tanaka, Knowledge Platform to use BioMedical Semantic WEBs individually:

Bridging Gene Ontologies and Clinical Ontologies at individual site, International Journal of Computer Science and Network Security, vol. 6-2, ISSN:1738-7906, 1-9, 2006

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.

Koji Sasaki received the B.S. degree from Faculty of Science, Univ. of Keio in 1966. He received M.S. from Stanford University and received the M.E. degree from MIT, in 1968 and 1970 respectively. After staying Hitachi Ltd as AI engineer, he founded AdIn Research Inc. in 1986. Now he is a president of AdIn Research Inc., a Professor of Chubu University, and a Chairman of Alliance group Inc.. He is an ex-president of MIT-J, Stanford Association of Japan, and MIT-EF.

Hiroshi Tanaka received the B.E. and M.E. degrees from Tokyo Univ. in 1974 and 1976, respectively. After staying as lecturer at Tokyo Univ., he received Dr.Med. and Ph.D. degrees from Tokyo Univ. in 1981 and 1983, respectively. He stayed in M.I.T., Uppsala University and Linkoping University, Hamamatsu Univ. School of Med.. Now he is a Director of Center for Information Medicine and a Professor of school of biomedical science, Tokyo Medical and Dental Univ.. He is a President of JAMI and is a member of ISO, MEJ, ICOT, CBIJ, IEEE, and VRJ.