Integration of Heterogeneous Knowledge-Bases in Medical Domain

Shusaku Tsumoto and Hiroshi Tanaka
Department of Information Medicine
Medical Research Institute, Tokyo Medical and Dental University
1-5-45 Yushima, Bunkyo-ku Tokyo 113 Japan
TEL: +81-3-3813-6111 (6159) FAX: +81-3-5684-3618
E-mail:{tsumoto, tanaka}@tmd.ac.jp

Hiromi Amano, Kimie Ohyama, and Takayuki Kuroda
Department of Orthodontics(II),
Faculty of Dentistry, Tokyo Medical and Dental University
1-5-45 Yushima, Bunkyo-ku Tokyo 113 Japan

Abstract

Medical data consist of many kinds of data from different resources, such as natural language data, sound data from physical examinations, numerical data from laboratory examinations, time-series data from monitoring systems, and medical images (for example, X-ray, Computer Tomography, and Magnetic Resonance Image). Therefore it has been pointed out that medical databases should be implemented as multidatabases. However, there have been few systems which integrate these data into multidatabases. In this paper, we report a system called COBRA (Computer-Operated Birth-defect Recognition Aid), which supports diagnosis and information retrieval of congenital malformation diseases and which also integrates natural language data, sound data, numerical data, and medical images into multidatabases on syndrome of congenital malformation.

As a result, it is easy to implement these knowledge-databases in COBRA on the object-oriented scheme, which suggests that these clinical databases should be implemented as object-oriented databases.

1 Introduction

There have been developed many medical decision support systems, such as MYCIN since the end of 1970's [2]. One of the most important problems of these decision support systems is that they cannot handle multi-data. Actually, medical data consist of many kinds of data from different resources, such as natural language data, sound data from physical examinations, numerical data from laboratory examinations, time-series data from monitoring systems, and medical images (for example, X-ray, Computer Tomography, and Magnetic Resonance Image). Therefore it has been pointed out that medical databases should be implemented as multidatabases. However, there have been few systems which integrates these data into multidatabases. In this paper, we report a system called COBRA (Computer-Operated Birth-defect Recognition Aid), which supports diagnosis and information retrieval of congenital malformation diseases and which integrates natural language data, sound data, numerical data, and medical images into multidatabases on syndrome of congenital malformation [4, 5]. Furthermore, since it also has an expert-system module and a module of case-based reasoning, COBRA can diagnose future clinical cases.

This system is implemented on several kinds of object-oriented databases or programming language, such as ONTOS system in Sun SPARC Station, SuperCard in Macintosh and Visual Basic in PC. It consists of the following four knowledge-bases, called *ontology*, which are implemented as object-oriented databases, and three modules, which are implemented by object-oriented programming language.

These knowledge-databases in COBRA are easily implemented in the object-oriented scheme, which suggests that these clinical databases should be implemented as object-oriented databases. In this paper, we present the present architecture of COBRA, and report how naturally we can implement medical databases as an object-oriented databases.

The paper is organized as follows: in Section 2, we discuss the four implemented knowledge-bases: concept ontology, task ontology, event ontology, and case-databases. Section 3 presents four kinds of integration of these heterogeneous knowledge-bases: information retrieval, expert-system module, case-based reasoning module, and case-follow-up module. In Section 4, we discuss the problems about COBRA. Finally, Section 5 concludes this paper.

2 Knowledge-Bases: Ontology

COBRA includes the following four knowledge-bases, called ontology: concept ontology, task ontology, event ontology, and case-databases. In the subsequent subsections, we discuss these knowledge-bases.

2.1 Concept Ontology

These knowledge-bases describe knowledge on diseases, clinical observations (symptoms), and other medical concepts, all of which are implemented as the object-oriented databases [3, 7].

2.1.1 Knowledge on Disease

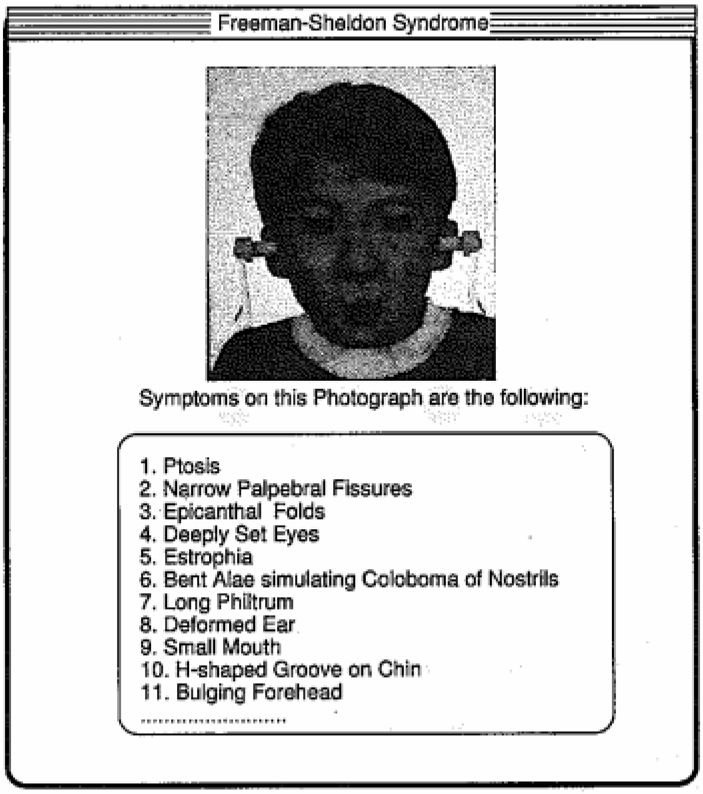
This knowledge-base represents information on diseases, such as the definition, typical clinical cases, examinations needed for diagnosis. If needed, sound data, such as voice or heart sound, are attached to this knowledge-base. For example, the description of Freeman-Sheldon syndrome consists of the following six items:

(1) Definition of this Syndrome: this syndrome is symptomatologically defined as a set of the following 19 symptoms: ptosis, narrow palpebral fissures, epicanthal folds, deeply set eyes, estropia, exotropia, small nose, bent alae simulating coloboma of nostrils, long philtrum, deformed ear, small mouth, H-shaped groove on chin, bulging forehead, short stature, poor weight, scoliosis, ulnar deviation of fingers, whistling face, and mask-like face.

Each symptom is linked to its knowledge in concept ontology. For example, ptosis is linked to its definition, how to measure the severity of ptosis and its therapy.

(2) Typical Photograph of this Syndrome: this item is linked to case-databases. Fig. 1 shows a typical case of Freeman-Sheldon syndrome linked to this item, where its symptomatological description is attached. Since each symptom is also linked to its knowledge in concept ontology, users can refer to precise information on each symptom.

(3) Laboratory examinations: this item is linked to *task ontology*, which is discussed later.



Symptoms on this Photograph are the following:

1. Ptosis
2. Narrow Palpebral Fissures
3. Epicanthal Folds
4. Deeply Set Eyes
5. Estrophia
6. Bent Alae simulating Coloboma of Nostrils
7. Long Philtrum
8. Deformed Ear
9. Small Mouth
10. H-shaped Groove on Chin
11. Bulging Forehead

....

Figure 1: A typical Case-Database of Freeman-Sheldon Syndrome

(4) Medical Images: this item is linked to *task ontology*, which is discussed later.

(5) Mechanisms of this Syndrome: although we have few knowledge on the causes of most of the congenital malformation, some diseases are said to be of genetic origin. For example, in the case of down syndrome, the trisomy of the chromosome XVIII is pointed out to be its cause. However, in the case of Freeman-Sheldon syndrome, there is no evidence on its cause.

(6) Genetics: there have been reported several famous family, some of whose members suffered from the same syndrome. These genetic knowledge makes us understand the characteristics of this disease from the viewpoint of genetics, such as autosomal dominant. This item is also linked to *event ontology*, and case-databases. For example, in the case of Freeman-Sheldon syndrome, this item is linked to family trees of typical cases, as shown in Fig. 2.

This figure shows a family tree of a typical case shown in Fig. 1. The arrow shows where this case is located in this tree. Square and circle denote male and female respectively, and black square and circle denotes a patient suffering from the same syndrome. Therefore this figure shows that his mother suffers from Freeman-Sheldon syndrome solitarily, since there are no relative who has the same disease, and also shows that her younger sisters suffers from this syndrome.

This family is also linked to each case-database, which can be retrieved from this screen.

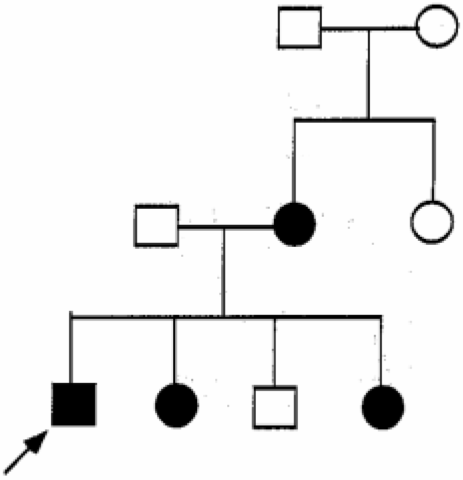


Figure 2: A Family Tree of Freeman-Sheldon Syndrome

(7) Diagnostic Rule: it is closely related with the definition of a syndrome. In the case of Freeman-Sheldon syndrome, the rule is defined as follows:

```
If a patient has the following 19 symptoms:
ptosis, narrow palpebral fissures,  epicanthal folds,
deeply set eyes, estropia, exotropia, small nose, bent alae
simulating coloboma of nostrils, long philtrum,
deformed ear, small mouth, H-shaped groove on chin, bulging
forehead, short stature, poor weight, scoliosis, ulnar deviation
of fingers, whistling face, and mask-like face,
then Freeman-Sheldon syndrome is suspected with probability 1.0.
```

However, there are many cases which do not completely match the above condition. Therefore we should calculate certainty factors for those partially matching cases. For this purpose, this rule is linked to concept ontology on each symptom, where the value of statistical measure for each disease is described. As discussed later, Expert-system module searches for each value of statistical measure when this rule is evoked [2].

2.1.2 Knowledge on Clinical Observations

This knowledge describe the meaning of clinical observations, how to observe this manifestations, and statistical measures with which we can suspect a syndrome when we observe this symptom.

For example, ptosis is a symptom of the eyelids. In the normal individual, the eyelids on each side are at the same level with respect to the limbus of the cornea [1]. However, in the cases of ptosis, the upper eyelids are retracting and going down below the limbus of the cornea. Although this symptom is a characteristic feature of muscular dystrophies, myasthenia gravis and third nerve lesions, it can also be observed in congenital malformations. However, this symptom is not specific to any congenital malformations, since almost the all cases may have it. Therefore the statistical measure for each disease to be suspected is very low.

In this knowledge-databases, the meaning of clinical observations is linked to corresponding anatomical knowledge, and physiological knowledge. For the above example, since ptosis is closely related to the third nerve, anatomical knowledge on the third nerve is referred. Furthermore, the item which shows the way to observe manifestations is linked to corresponding task ontology. For the above example, since ptosis can be observed by an ordinary neurological examination, it is linked to knowledge on neurological examinations.

2.1.3 Knowledge on Other Medical Concepts

It is not sufficient to implement only the concepts of diseases and symptoms. We also need anatomical knowledge, physiological knowledge, and knowledge on examinations, such as X-ray and laboratory examinations.

For the above example, we need to understand anatomical knowledge on the third nerve in order to understand ptosis. The third nerve cell is located in the midbrain and controlled by neurons in the upper region, such as encephaly. The third nerve passes near the chiasma of optic neurons, and cartoid artery, and connects with optical muscles. From this knowledge, the malfunction of the third nerve may cause that of optical muscles, which results in estrophia or extrophia. These symptoms are also included in a typical case of Freeman-Sheldon syndrome (Fig. 1), and in the diagnostic rule of this syndrome.

Actually, these kinds of knowledge-bases are indispensable to education of residents or students, since it is difficult for them to associate some disorders with anatomical knowledge. For this purpose, we implement such kind of knowledge as much as possible.

2.2 Task Ontology

This knowledge-bases describe the knowledge on laboratory examinations and medical images.

2.2.1 Laboratory Examinations

There are several syndromes caused by endocrinological disorders. For example, hypothyroidism, which means that a patient cannot generate enough thyroid hormone, causes several disorders, such as mental retardation. Therefore we need several kinds of laboratory examinations in order to make a differential diagnosis.

Each knowledge-base describes the way to make laboratory examinations, how to interpret the results, and statistical measures with which we can suspect a syndrome when we observe the specific results. For the above example, when concentration of thyroid hormone is very low and when concentration of thyroid stimulating hormone is very high, hypothyroidism is strongly suspected with probability 0.99.

2.2.2 Medical Images

Medical images consist of X-ray, Computer Tomography (CT), and Magnetic Resonance Image (MRI). Since these medical images have their own characteristics, medical experts use those images for their purpose. For example, when they would like to characterize whether a target region is rich in water or not, they will use MRI technique to focus on that region. This ontology describes these characteristics as the following two items.

(1) Parts: information on what parts and components of the body can be focused on by this image. For example, X-ray is suitable to observe characteristics of bones, but not to observe characteristics of air-rich regions, such as lung. This item is linked to *concept ontology* on anatomical knowledge.

(2) Resolution: information on what size can be analyzed. For example, as to Computer Tomography, 1.0 cm is the minimum size to recognize substances. This item is also linked to *concept ontology* on anatomical knowledge.

2.3 Event Ontology

This knowledge describes two kinds of temporal knowledge on diseases. One is knowledge on genetics, or family history, and the other is a clinical course of a patient, or present history.

The former part is represented as a family tree, that is, tree structure as shown in Fig. 2. Each node is implemented as an object, which is linked to ancestor or successor of this node, and to the *case-database* which describe the case of this node. For example, the arrowed black circle is linked to a typical case of this syndrome shown in Fig. 1.

The latter part is linked to each case-database. Since most of the patients treat their malformations by surgery operations, it is important to follow the courses after the operation. In order to realize temporal relations, each case-database has a time-stamp and has descriptions on the links to the past and the future case-database. Therefore we can retrieve the past or the future courses from a given databases.

2.4 Case-Databases

This databases consist of 267 clinical cases on congenital malformations, whose patients came to outpatient clinic in Tokyo Medical and Dental University. Each case is composed of family history, description of symptoms, photographs of this patient, results of laboratory examinations, and medical images, all of which are stored as sub-databases.

These knowledge-databases in COBRA are implemented in the object-oriented scheme, which suggests that these clinical databases should be implemented as object-oriented databases.

3 Integration of Knowledge-Bases

The present version of COBRA supports the following four kinds of main procedures: information retrieval, expert system, case-based reasoning, and case-follow-up, all of which COBRA uses the four knowledge-bases mentioned above.

3.1 Information Retrieval Module

Basically, COBRA provides *event-driven* interfaces. That is, users can refer to all the information related to each item in the databases. For example, when users are retrieving a case shown in Fig. 1, they can refer to any items in this case-record, such as the characteristics of the syndrome and those of symptoms.

3.2 Expert-System Module

This Module is developed in order to diagnose future clinical cases, and also used as an intelligent tutoring system for medical residents. It consists of three submodules: Input-interface, Inference engine, and Diagnostic outputs, which connect with concept ontology, task ontology, and case-databases. Fig. 3 shows these connections and how diagnostic procedures work. In the subsequent paragraphs, we discuss these diagnostic processes in detail.

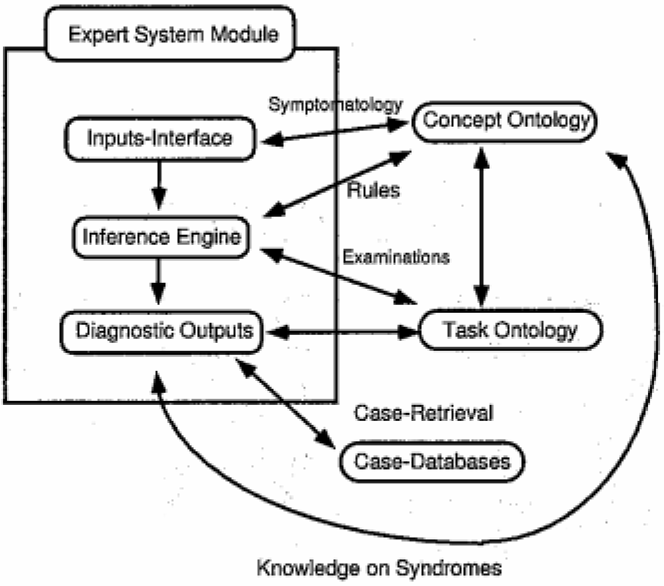


Figure 3: Integration in an Expert-system Module

First, users input observations from menus given by Input-interface submodule. COBRA provides two kinds of interfaces, one of which is a list of descriptions on symptoms, and the other of which is a list of photographs. As to symptoms better to visualize, COBRA supports the latter interface. For example, Fig. 4 shows a list of photographs which describes typical patterns of symptoms of eyes. Users can not only select one of those photographs, but also see their information. For example, if users select the right bottom picture and click the help button, then COBRA shows the help window as shown in Fig. 5, which gives information on what kind of symptoms can be observed.

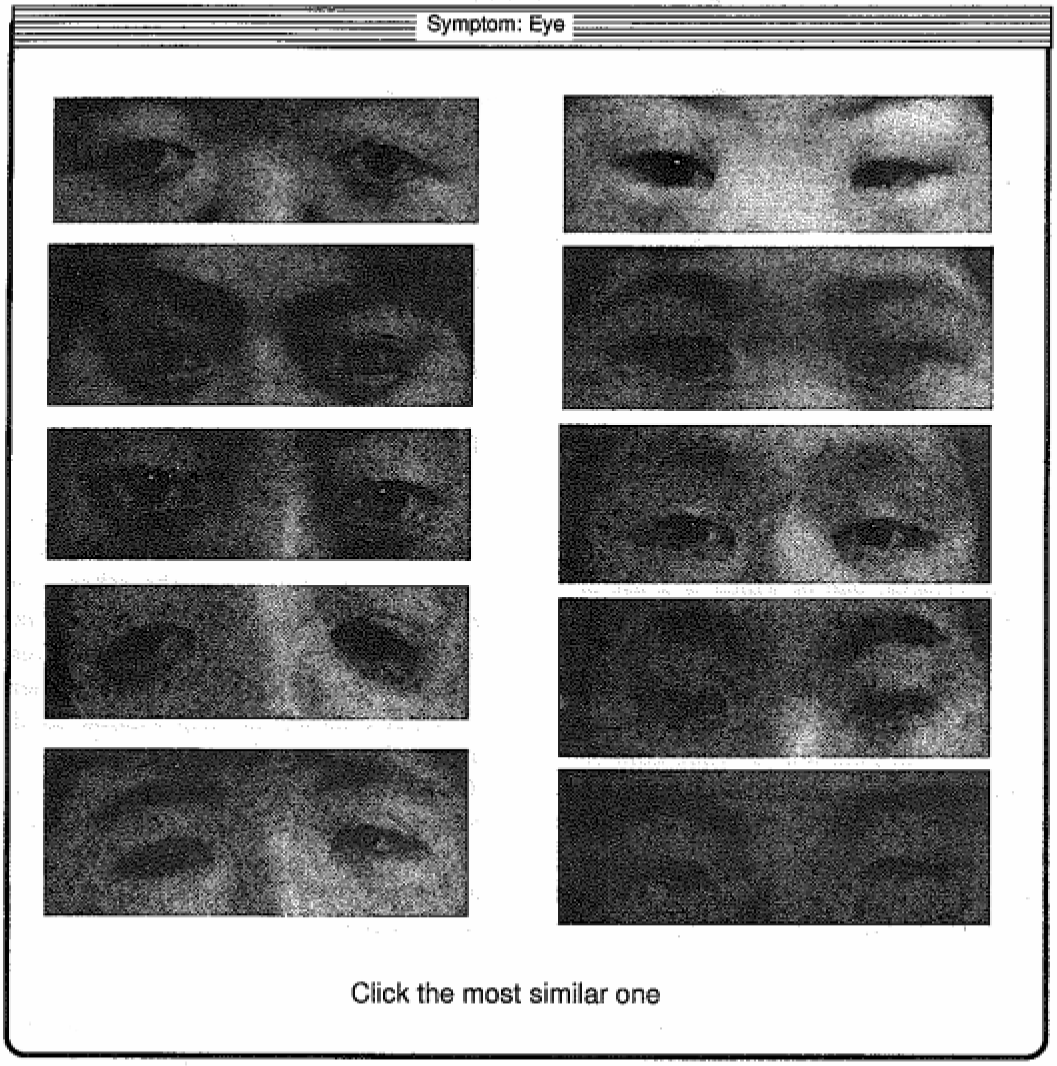
Furthermore, if users click one of the symptoms, then COBRA gives us where the selected symptom is observed. These symptoms are also connected with concept ontology, where the concepts of these symptoms are described.

After users select a photograph, COBRA interprets the input and transforms it into verbal information. For example, if users select the right bottom picture, COBRA takes the following five symptoms as inputs: ptosis, narrow palpebral fissures, epicanthal folds, deeply set eyes, and estrophia, all of which are shown in the help window (Fig. 5).

Second, COBRA applies inference rules to all the inputs, calculates total certainty factors from each statistical measure attached to each symptom in concept ontology, and orders diagnostic candidates by the values of total certainty factors.

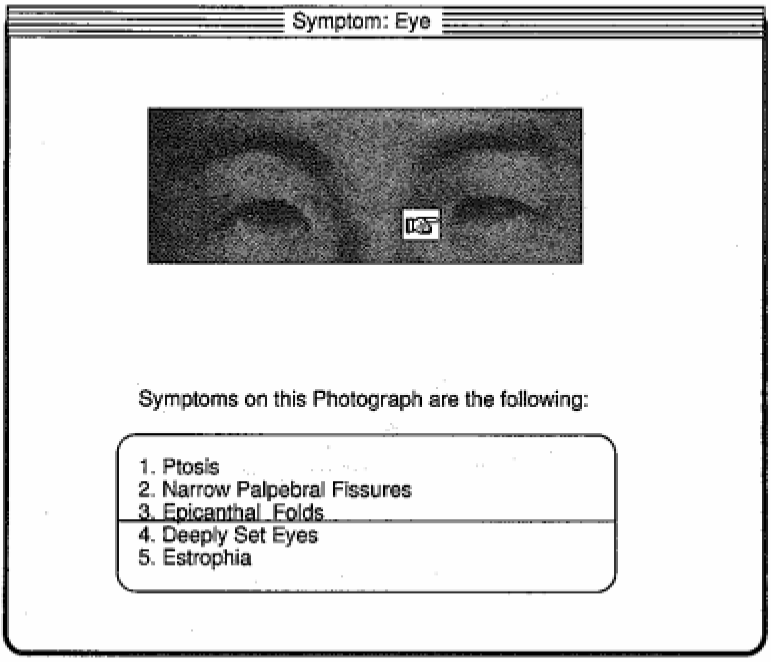
When other examinations, such as X-ray, are needed to make a differential diagnosis, inference engine looks for its knowledge from task ontology, and retrieves what kind of manifestations can be got from examinations. The inference engine also searches for statistical measures and other important knowledge on the above manifestations from concept concept ontology. Then it calculates certainty factors for each case when a specifc manifestation is derived or not.

Finally, Diagnostic-outputs submodule outputs the final candidates, and retrieves typical cases for each candidates. Furthermore, this submodule looks for the concept of syndromes from concept ontology.



Click the most similar one

Figure 4: An Example of Input-Interface



Symptoms on this Photograph are the following:

1. Ptosis
2. Narrow Palpebral Fissures
3. Epicanthal Folds
4. Deeply Set Eyes
5. Estrophia

Figure 5: A Help Window

3.3 Case-Based Reasoning Module

This Module is developed in order to diagnose future clinical cases and to retrieve similar cases, and it is also used as an intelligent tutoring system for medical residents.

Although the interfaces for inputs are same as the Expert-System module, reasoning strategy is different: when all inputs are completed, similarity measures are calculated, and the module outputs a list of best-fit cases.

A Case-Based Reasoning System module (CBR module) consists of three submodules: Input-interface, Indexing and Retrieval, and Diagnostic outputs, which connect with concept ontology, task ontology, and case-databases. Fig. 6 shows these connections and how case-retrieval procedures work. In the subsequent paragraphs, we discuss these diagnostic processes in detail.

First, users input observations from menus given by Input-interface submodule. This module is almost the same as that in Expert-system module. That is, COBRA provides two kinds of interfaces, one of which is a list of descriptions on symptoms, and the other of which is a list of photographs. These symptoms and photographs are connected with concept ontology where the precise knowledge on those symptoms is stored. After users select a photograph, COBRA interprets the input and transforms it into verbal information.

Second, COBRA calculates a similarity measure from all the inputs. Each case in Case-Databases has a list of symptoms as shown in Fig. 1 and we attach weights acquired by interview with domain experts to each symptom.

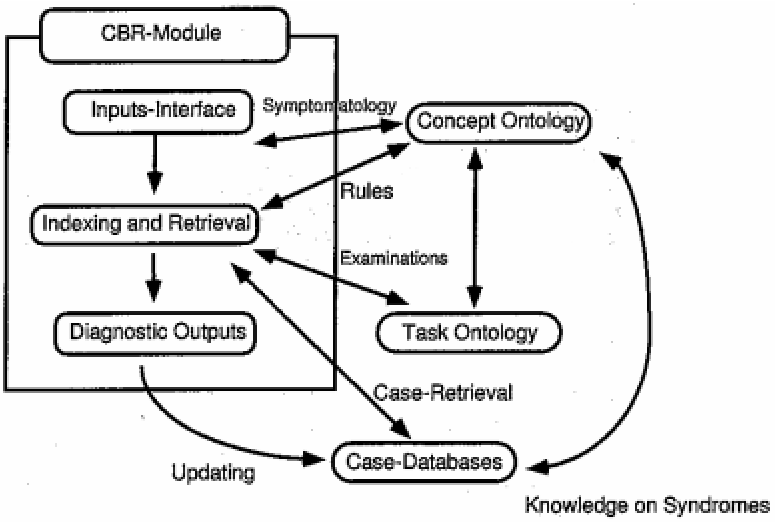


Figure 6: Integration in a CBR Module

Based on the weights, the similarity measure is defined as:

$$\frac{\sum_{i=1}^{n} w_i \chi(i)}{\sum_{i=1}^{n} w_i},$$

where n and w_i denotes the number of all symptoms and a weight for symptom i respectively and where $\chi(i)$ denotes a characteristic function of each symptom. If symptom i appears in a given case, then $\chi(i) = 1$. Otherwise, $\chi(i)$ is equal to 0. For example, when a case has only five symptoms shown in Fig. 5, the similarity measure between this case and the above case shown in Fig. 1 is equal to:

$$\frac{0.1 + 1.0 + 0.8 + 1.0 + 0.8}{11.1} = 0.333,$$

where 0.1, 1.0, 0.8, 1.0, and 0.8 denote weights for ptosis, narrow palpebral fissures, epicanthal folds, deeply set eyes and estrophia, respectively, and where $\sum w_i = 0.1 + 1.0 + 0.8 + 1.0 + 0.8 + 0.1 + 1.0 + 1.0 + 0.1 + 1.0 + 0.1 + 1.0 + 0.1 + 0.1 + 0.1 + 1.0 + 1.0 + 0.8 = 11.1$.

If small mouth and mask-like face are also observed, then the similarity measure is equal to:

$$\frac{0.1 + 1.0 + 0.8 + 1.0 + 0.8 + 1.0 + 1.0}{11.1} = 0.514,$$

since weights for both manifestations are equal to 1.0.

And if a patient have all the symptoms, then its measure is equal to:

$$\frac{11.1}{11.1} = 1.00,$$

which is the maximum value of this similarity measure. After similarity measures are calculated for all the cases, the submodule orders candidates by the values of similarity measures.

Finally, Diagnostic-outputs submodule outputs the five most similar candidates and updates case-databases with a given new case.

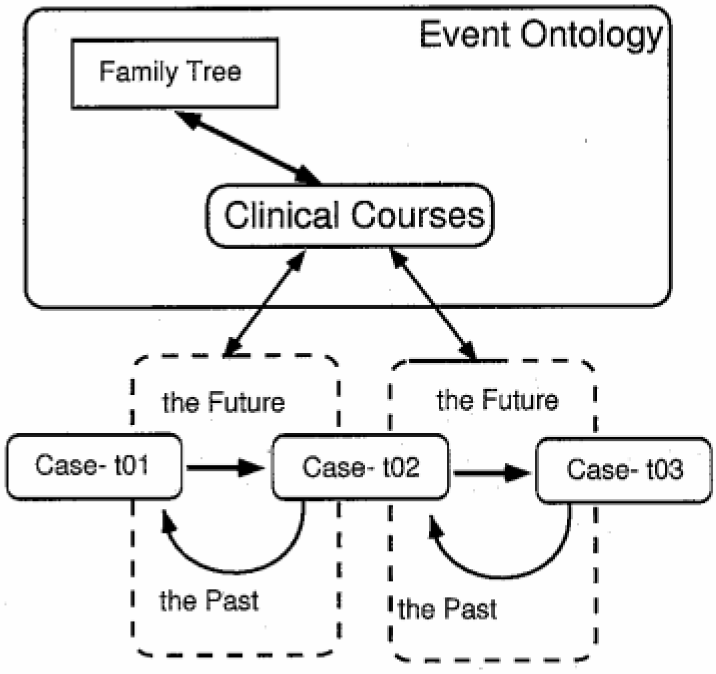


Figure 7: Case-Follow-Up Module

3.4 Case-Follow-Up Module

As discussed in the Subsection 2.3 on event ontology, COBRA supports case-follow-up study based on the scheme of object-oriented databases as shown in Fig. 7. In this module, each case-database is linked to a clinical-courses database in event ontology, which includes an abstract clinical courses of the case-databases and which is connected with a family tree. This database has a simple tree structure whose node corresponds to each case.

This whole structure can be viewed as a temporal hierarchy, since a family tree is temporal description where time interval between nodes are from 20 to 40 years, clinical courses is one where time interval between nodes are from 1 to 10 years, and each case describe the status of each patient at one instant.

From any point, users can retrieve any temporal information on the selected patient. For example, users can searches for his or her family history, for his or her past history before the selected point, and for clinical courses after that point. Since each case is linked to concept ontology, and task ontology, users also retrieve enough information to understand the status of the patient.

This structure is very important for students or residents to learn clinical courses, and therapy dynamically, since this architecture shed light on dynamic aspects of reasoning of domain experts.

In the present version of COBRA, this architecture is only applied to case-follow-up module, since we do not have so many follow-up data. However, we incorporate this architecture into other module in the near future, because this temporal architecture is also very important for clinical support.

4 Discussion – Related Work

This system is implemented on several kinds of object-oriented databases or programming language, such as ONTOS system in Sun SPARC Station, SuperCard in Macintosh and Visual Basic in PC.

The hardest problem is that a huge space is needed to include photographs and medical images of high quality. In general, COBRA needs 1.5 to 2.0 MB to describe one clinical case-data. Therefore it spends about 560MB only to store case-databases, which makes the seeking speed very slow. However, there are more than 1000 syndromes reported in the literature, although we are now using only 260 clinical cases and 66 syndromes. This means that we need much space to extend our system.

Actually, since this size is beyond the power and capacity of personal computers, we cannot extend the PC version of COBRA into a more general version. Although the only solution seems to develop our system only in Sparc station, the searching speed is too slow even in the present version. Therefore we feel that we need much computational power to implement a full system.

As to relate work, there is a diagnostic system, called *POSSUM* which supports diagnosis of congenital malformation [6]. This system is implemented on PC-AT compatibles with a lot of audio-visual interfaces. First, users select symptoms from a list, then POSSUM calculates similarity measures, then searches for the nearest record from Magnetic-Optical desk. This system is very fast, and retrieves the suitable case correctly.

POSSUM can also be viewed as multidatabases. However, the problem of this system is its interface. If user do not know exactly about symptoms, then they cannot use POSSUM. Moreover, users cannot retrieve information on what users want to know. They may want to understand the meaning of symptoms when they are selecting symptoms from a list.

On the other hand, COBRA, based on object-oriented scheme, can respond any requirement of users. This architecture is the most advantage of the object-oriented scheme, and it enables COBRA to be useful.

However, as discussed above, our system is very slow because of spending too much space for image data. To solve this problem is our future work.

5 Conclusions

In this paper, we report a system called COBRA (Computer-Operated Birth-defect Recognition Aid), which supports diagnosis and information retrieval of congenital malformation diseases and which also integrates natural language data, sound data, numerical data, and medical images into multidatabases on syndrome of congenital malformation.

As a result, it is easy to implement these knowledge-databases in COBRA on the object-oriented scheme, which suggests that these clinical databases should be implemented as object-oriented databases. It is our future work to evaluate the performance of our system, and to incorporate learning strategy into this system.

References

[1] Adams, R. D. and Victor, M. *Principles of Neurology*, 5th edition, McGraw-Hill, NY, 1993.

[2] Buchnan, B. G. and Shortliffe, E. H.(eds.) *Rule-Based Expert Systems*, Addison-Wesley, MA, 1984.

[3] Code, P. and Yourdon, E. *Object-Oriented Analysis. 2nd Edition*, Prentice Hall, NJ, 1991.

[4] Goodman, R. M., Golin, R. J. *Atlas of the face in genetic disorders 2nd edition*, C.V. Mosby. Saint Louis, 1977.

[5] Jones, K. L. *Smith's Recognizable Patterns of Human Malformation 4th edition*, W. B. Saunders Philadelphia, 1988.

[6] POSSUM: Picture of standard syndromes and undiagnoses malformations. Computer Power Pty. Applied Research and Development. Melbourne.

[7] Taylor, D. A. *Object-Oriented Technology: A Manager's Guide*, Addison-Wesley, 1990.