INNOVATION APPROACHES TO THE CONSTRUCTION OF BIOINFORMATION SYSTEMS WITH DATABASES: KEYS BASED ON GENETIC CODES

Introduction. The application of achievements of biology, biophysics in technique opens new opportunities for innovations, in particular for construction of relational databases (DB) with biomedical data, helps solve problems and obtain results at qualitatively new level.

Problem Statement. The development of information systems with biomedical information is relevant both in peacetime and in wartime. Implementation of contemporary information and computer technologies for the development of information systems with DB in biology and medicine has its own specifics. That is why innovative approaches for the construction of biomedical relational DBs with use of keys with advanced capabilities are relevant.

Purpose. The development and design of biomedical relational DBs with keys based on genetic codes of organisms in alphanumeric expression with further application as part of the novel bioinformation systems.

Materials and Methods. Methods of object-oriented system analysis for optimal construction of DBs with biomedical information, method of Entity-Relations (ER)-diagrams design, methods of DB design.

Results. By the example of relational DB with information about some fish species, the approach based on object-oriented analysis for the construction of DBs in optimal way has been suggested, applied, and the algorithm of their construction has been described. Particular attention is attracted to the solution of creating keys based on genetic codes of fishes in alphanumeric expression (especially the primary keys) that provide links


© Publisher PH “Akademperiodyka” of the NAS of Ukraine, 2024. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
between individual tables of DBs, ensure integrity of information in such system and reliable access. The high level of data individualization with using the keys based on genetic codes in DB has been analyzed and substantiated.

Conclusions. The results can be used for the creation of appropriate information systems, including bioinformation systems. They have both theoretical, for the further development of DBs construction methods, and practical value for improving some methods of data protection and can be useful to solve tasks of construction of DBs with biomedical materials for peacetime and wartime use.

Keywords: information technologies, information system, databases, primary key, coding, biomedical information, and chemical substances.

BIOMEDICAL DATABASES WITH GENETIC INFORMATION AND RESULTS OF ITS ANALYSIS AND PROCESSING

Innovations in biophysics, biotechnology, molecular biology, genetics, other sciences of biomedical areas, influence significantly economic progress and all our contemporary life. Some methods of information technologies developed previously for engineering and other technical branches can be used with some modifications for the creation of databases with biological and medical contents for bioinformation systems. Specifics of biology and medicine are the studying of complex objects, living systems, multifactorial influences on them, and etc. In process of scientific investigations it is necessary to analyze this specificity of biomedical objects, and take it into account for computer technologies development, for example, construction of databases (DB) in biology and medicine. It is also important to understand that due to this specificity of medical and biological objects, huge amount of accumulated material and data about them, the results of experiments, observations, and etc., the use of modern information and computer technologies based on DB will contribute to further progress in these areas, because they allow analyzing and processing a huge (and growing from day to day!) amount of the data.

Achievements in life sciences, being combined with modern developments in technique form the potent base for their progress today. For example, biology and medicine have such powerful modern methods as organizing of the data obtained in experiments and observations in databases (DB), which are information repositories. They are often called biomedical databases (BiomDB), or databases with biomedical materials. Biomedical materials for such databases can be obtained from different levels of hierarchy of the world of biological organisms. For example, if we speak about the databases with human biomedical materials, then these can be databases with genetic material (DNA, RNA), biological molecules of protein nature, samples of body tissues, photos of virtual sections of various tissues, organs (brain, liver, etc.).

Updated lists of such databases are regularly published. For example, the list of 180 BiomDB was published in Nucleic Acids Research, where previously described BiomDB were updated in 2018. Other numerous DBs examples can be listed too [1, 2]. These numerous BiomDB with biomedical information have already been united into the groups; examples of lists of such groups are given in [1, 2].

Such BiomDBs are important tools for practical use, because they contain data on hereditary diseases, oncology (for example, in comparison with data on pollution, including chemical and radiation ones), others. With such databases the biomedical studies at new technological level became possible — with the use of genetic information of humans that can be examined together with personal and medical data. Such novel approach has excellent prospects for medicine — for the prognosis, diagnostics and cure of diseases, as well as for industry — for the development of new products and services [1].

In contemporary biomedical practice there is an understanding that the establishment, harmo-
Innovation Approaches to the Construction of Bioinformation Systems with Databases: Keys Based on Genetic Codes

Innovation Approaches to the Construction of Bioinformation Systems with Databases: Keys Based on Genetic Codes

nization and broad use of human biobanks and genetic research databases are the key for future progress in biomedical sphere. The technologies in which the use of the data about the human genetic material and analytical information derived from it are seen as very prospective today. Human biobanks and genetic research databases permit to unite different types of information, as well as ordering and sharing of human biological data and information after their processing and analysis; and these possibilities are very important for contemporary researches [1, 2].

Such biodata are also important for criminology to identify criminals, terrorists, and other persons linked with the crimes. During wars, the importance of development of databases with biomedical materials increases even more. With their help, it is possible to identify persons, who are guilty of war crimes, to identify dead humans or wounded persons evacuated from the battlefield in an unconscious state, and etc. [1—4].

Over the past few years, our scientific group has been working on the development of a number of innovative methods linked with the application in technology of results of researches in biology and medicine. At the same time, there was a need to take into account the specificity of huge volumes of biomedical data, and therefore the creation of databases with biomedical material [5—12]. A number of our works were devoted to the construction of databases with genetic material [9—12], with subsequent publication of the obtained materials, both in scholarly research articles and in patents [5—12]. The specificity of this field is the organization and processing of great volumes of biomedical data and the amount of these data is increasing constantly. Therefore, it is important to develop all methods that allow improving such work with biomedical data, making such work more reliable, with higher quality and speed of the data processing, etc. The work done was aimed on improving of all these characteristics, as well on increasing of the quality of fulfilled works in general. Few patents of Ukraine were obtained for our results that we publish in present article [9—12].

Purpose of present work was development and design of biomedical relational DBs with keys based on genetic codes of organisms expressed in alphanumeric terms, with information on biological organisms and chemical substances — organic pollutants with further use in novel bioinformation systems. This idea we explain below on the example of the construction of the database with information on biological organisms (fishes) and chemical substances — organic pollutants of environment. Such tasks are relevant in our reality indeed, especially in industrial regions, areas polluted due to ecological disasters or armed conflicts.

For the solution of the tasks in this article, we used such techniques as object-oriented system analysis, Entity-Relations (ER)-diagram design, and the methods of computer relation databases construction, based on careful previous analysis, processing of biological material obtained as the result of ecological monitoring of environment. In general, during the work, a review and study of the most used mathematical methods in biology and medicine during the last decade was made, a number of them were applied in the process of this work [13, 14].

EXPERIENCE IN CREATING BIOMEDICAL DATABASES AND NEW APPROACHES TO SOME PROBLEMS SOLUTIONS

Numerous contemporary publications are devoted to the problems of creating databases with biomedical information, as well as to construction of samples of technical and information systems (ISs) in different countries of the world, including bioinformation systems [1—8, 10—30]. Numerous examples of patents, other scientific and technical sources were examined carefully in process of this work. In our previous publications, we have also described our experience in developing of bioinformation ISs based on databases of fishes [6], and insects (Noctuidae, Lepidoptera) [7, 8]. Further these ordered data from DB can be used for the purposes of ISs construction. We had already developed, patented and described in articles some
bioinformation systems with databases for the purposes of ecological monitoring of fauna [5—8].

It should be noted that the work on ordering of biomedical information in databases, including those combined with information systems are quite numerous in the world [1—8, 14—39]; including such systems for fishes [6, 31—43]. The application of achievements in biology and biophysics in technique opens new opportunities for innovations. Technology of creating of relational databases with biomedical data is one of such directions, and implementation of such innovations permits to solve problems and obtain results at qualitatively new level. That is why the author’s innovative approaches to the design of relational databases with biomedical data using the keys with advanced capabilities are really relevant.

Biomedical material — the data for such databases — we usually take from the publications of our colleagues [40—43], or from our own publications with results of experiments or observations [6—8, 43]. For relative databases construction we used the standard algorithms developed for such cases [44, 45]. For classification of biological organisms, fishes as well, classic approaches were used [46—52].

Such area as databases construction was considered as classic one with its already developed and established methods [44, 45]. However, our experience demonstrated that modern scientific and technological progress opens the opportunities for the perfections in this sphere. For example, when the developer would like to increase such characteristics of information in database as data integrity, the level of data individualization as well as data protection in whole — it is necessary to invent new methods. To solve this problem Dr. Klyuchko O.M. invented a potent method of keys coding in biomedical databases basing on the principles of genetic codes, and three patents of Ukraine were obtained exactly for this work [9—11]. These findings were described below on the example of the database with information about fishes. According to it, the keys in databases can be formed on the base of fragments of genetic codes expressed in alphanumeric terms. This is extremely important for databases with information about biological species, individual persons or tissues of different types. After the implementation of this method the integrity of information, its protection for individual objects in databases, other important characteristics could increase significantly.

The material below will be given in following logical sequence: 1 — the identification of the main necessary objects and the analysis of the objects to be included into relational database; 2 — on the base of this analysis — making of the “core” of relation DB; 3 — the analysis of the problem of relations formation on the base of traditionally used keys; 4 — the description of invented method of keys formation on the base of organism genetic codes (natural key); 5 — the substantiation of using this method for DB with the information of organisms, other biomedical information. Wishing to concretize the problem, we have chosen the two following main objects for analysis: fish species as biological objects (for our database three different species of fishes were taken); and organic “Chemical compounds” that, together with wastewater, flow to the media of water basins in industrial regions and affect negatively on the fish fauna. The third object in Fig. 1 includes taxonomic characteristics for main object — fish species.

**OBJECTS ANALYSIS — THE FIRST STAGE IN PROCESS OF DATABASES CONSTRUCTION**

The algorithm of designing of relational database with the data about fish species influenced by chemical substances we described previously in general for different species of living organisms [6, 7]. Below we will describe briefly the first steps of such algorithm because this is necessary for the explanation of the essence of the invention. Then we suggest more complete explanation of the core of traditional and invented techniques. First, we suggest the analysis of objects with their charac-
Innovation Approaches to the Construction of Bioinformation Systems with Databases: Keys Based on Genetic Codes

Firstly, characteristics and relations at the first stage of relational databases construction as necessary and recommended procedure. In our case the majority of such objects are biomedical ones. Secondly, their characteristics and links have to be ordered in relational databases, and ER-diagram has to be designed for this. The approach of Object Oriented Analysis (OOA) is used for this purpose. The aim of OOA is to enable the optimal DB construction, and we'll demonstrate this method briefly on the example of the construction of a DB with information about fish species and some chemical pollutants dangerous for them. So called “ER-diagram” have to be developed as the result of such analysis. Thirdly, we will demonstrate our idea to use fragments of genetic codes of biological organisms (on the example of fishes) as “natural keys” during these databases construction (Fig. 1).

Fig. 1. Core fragment of ER-diagram “Fishes’ Species and Organic Chemical Compounds Affected Them” for different fish species

<table>
<thead>
<tr>
<th>Key</th>
<th>Class</th>
<th>Series</th>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belongs</td>
<td>Fishes’ species</td>
<td>influenced</td>
<td>Chemical compound</td>
<td>Key</td>
<td></td>
</tr>
<tr>
<td>Primary key</td>
<td>Name of the fish (Latin)</td>
<td>Name of the fish (Ukrainian)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code of gene sequence</td>
<td>Area of inhabitance</td>
<td>Biotopes of inhabitance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spread</td>
<td>Objects of nutrition</td>
<td>Numerosity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>References to literary sources</td>
<td>Comments</td>
<td>Others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lethal dose LD_{100}</td>
<td>Semi-lethal dose LD_{50}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name of chemical compound (Latin)</td>
<td>Name of chemical compound (Ukrainian)</td>
<td>Molecular weight</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solubility in water</td>
<td>Minimal concentration of harmful action</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
If to summarize, ER-diagram is an abstraction that has to represent in whole all objects from the database with all their relations and with all characteristics of the objects — attributes; all of them are important from the point of view of DB construction. For such DB with biomedical information other data (attributes) are important as well: genetic sequences, specie, family, genus, and etc. ER-diagram helps to analyze better all entities (objects) that we would like to include into the database and observe them in their entirety. Construction of high quality DB is possible only with such analysis.

Figure 1 demonstrates that one of selected and designed objects — fish species — has some characteristics and we have identified them. These characteristics are called “attributes” in terms of databases. The same is true for two other objects: “Chemical compound” and “Taxonomic category.” Further, for the demonstration and in the framework of this task, we select some characteristics (attributes) for the objects fish species and chemical compound.
“Chemical compound”; we listed them in Fig. 2. For representation of the main notions of ER-diagram the special geometric symbols described in details in [44, 45] are used. In Fig. 1, we also use the symbol “ellipsis” alternately with many dot that mean that other characteristics as well can be selected for such schemes in framework of this task (as well as for other tasks). Symbol “many dots” is used in other figures below too.

Characteristics (attributes) in Figs. 1, 2 were selected on the base of the “Red Book of Ukraine” for few and endangered fish species of Ukraine [48] as well as from some monographs about organic substances and their derivatives — environmental pollutants identified in wastewaters from industrial regions of Ukraine [50]. For further development of the database we have to select some fish species and organic chemical compounds that are typical pollutants in industrial regions of Ukraine. For our example we have selected:

a) *Oncorhynchus mykiss* Walbaum, *Cyprinus carpio* L., *Barbus borysthenicus* Dybowski;

b) hydrocarbons and their oxidation products, petroleum, petroleum products, benzene, phenol and its numerous derivatives, indole and its numerous derivatives, resins, and etc. (only few of them were depicted in Fig. 2).

**Objects** there are the nouns — main entities; information about them has to be included into the DB (rectangles): fish species, and chemical compounds. **Attributes** there are object characteristics, necessary for object description (ovals): Latin name, native name, locations, and etc. **Relations** (rhombus) here mean links that unite the objects if they are logically linked (for example, “Fish species” is affected by “Chemical compounds”). Important characteristics of the relations between objects are the “Power of relation” (Fig. 3). If one fish species is characterized by Latin name and this name is linked with unique national name of the fish the “power of relation” is 1 : 1 (Fig. 3). From other side, if one species is affected by N chemical compounds, the “power of relation” is 1 : N. But if M species of fishes were affected by N chemical compounds, the “power of relation” is M : N, and etc. More about the art of DB construction one can read in [6—12, 44, 45].

Further in Fig. 4 we demonstrated how above described abstract forms can be filled with a specific meaning — the names of fishes and specific chemical substances — pollutants.

In process of further DB development the Object 1 “Fish species” and Object 2 “Chemical compound” (Fig. 1) can be transformed in the tables represented in Fig. 5. Attributes of these objects become the columns in the tables in Fig. 5. The

---

**Fig. 3.** Power of relations between two objects in process of the database construction

*Above:* “1 : N” — “one-to-many”. In framework of our task solved, this means: “1 species of fishes was affected by N chemical compounds”

*Below:* “1 : 1” — “one-to-one”. In our example, this means: “1 name of fish species in Latin correlates with 1 name of fish species in Ukrainian” (correlation between spelling of names in Latin and Cyrillic fonts)

**Fig. 4.** Logical model of DB with the data about some fish species and organic chemical compounds — water pollutants they are affected by. Fragment of ER-diagram “Fish species” and “Organic Chemical Compounds Affected Them” on the example of three fish species from Ukrainian water basins [48—50]
same procedure has to be done with the third ob-
ject in our example “Taxonomic category.”
So, if to summarize [6—8, 44, 45],
1. Objects (“entities”) of ER-diagram were trans-
formed into separate tables;
2. Attributes were transformed into the columns
(“fields”) of the tables;
3. The relations between the objects were trans-
formed into the relations between each of the
table.
Using this database, professionals and amateurs
in biology can find out what species of fishes are
under the dangerous influence of organic environ-
mental pollutants, and (with future DB develop-
ment) in which particular basin. Using such data-
bases, professional can easily perform all oper-
ations with the data, which were previously impossible:
data sorting, searching of any particular record,
data filtering, data outputting in any format, etc.
These opportunities are very convenient in pro-
fessional work because the users who works in
this field can add to the previous material new in-
formation step by step (data input into the DB),
or list the species alphabetically (sorting of the
data), or select all species from one genus (filtra-
tion procedure), or to do other operations like these.

**THE METHOD OF THE LINKS FORMATION USING THE “KEYS” FOR THE DATA INTEGRITY**

This method was invented in order to realize the
data integrity, increase the connectivity of infor-
mation in biomedical DB and its level of indi-
vidualization, as well as to prevent the losses of
information ordered in DB [6—8, 44, 45].
“Keys” are necessary “tools” for making the links (“relations”) between the objects. When “keys” were done, they guarantee that the information from the database is accessible, it cannot be lost and it can be obtained from the DB. Indeed, inaccessible data cannot be outputted! That is why the “keys” have to be elements of each relative database, each of its table. In our database too, all objects (entities) are related to each other through programmed links. After the information output from any table, it is possible to obtain consequently the information from another one associated, through a link that unites their keys. Usually, as a “key” serves a certain code from numbers or other symbols. In DB with biomedical information as a “key” in general majority of cases serves an element of a random numbers sequence. And the links between such codes from individual tables (transformed “objects”) can be done easily in many modern software environments [6—8, 44, 45].

So, it is easy to do all above described having ER-diagram where all entities are linked together. One can see the entire task in its completeness; it makes possible to organize the access to all records. So, we can be sure that information from database becomes available for everybody completely (if the data integrity is ensured). From other side, if the user does not realize this, he is unable to obtain the information from some DB sectors that are not connected through such links. However, using above described methods, programmer can recover connections based on the ER-diagram easily, through linking the objects with using the keys. In these applications of above described methods there are no differences between biomedical and technical databases.

Figure 5 shows that “keys” serve to make the links between those fields of the tables that have to be linked logically. For example, it is necessary to find from the database how any defined chemical substance (for example, phenol) influences on the fish (or fish population). Programmer knows that phenol in the water really influences on fish. So, in program the key “GenCodeCarp” from the table “Fish species” has to be connected with the appropriate key and linked with “Phenol” in the table “Chemical compound” (pollutants). And so on. Keys of chemical substances in Fig. 5 may be coded in commonly used way — by the group of randomly generated numbers (like key for “Phenol,” for example). About coding of the key “GenCodeCarp” will be written below.

**PROPOSED METHOD OF USING GENETIC CODES AS “KEYS” FOR DATABASES CONSTRUCTION**

More interesting case is formation of keys for biological objects. Dr. Klyuchko O. invented it and we would like to illustrate it on the example of fish. Each biological organism, each species has its own unique biological code — genetic code, which we can express abstractly as the sequence from the letters and numbers [9]. Each such sequence is highly individual — for each species or even for each individual organism. There is only a task for us — to define the fragment of such individual code, which we would like to use for the purpose to be a “key.” Such case is a good example of so called “natural key,” which serves as “primary key.” Such novel key permits to realize the effect of “data integrity” in the DB. In table field “Code of genetic sequence” we place a sequence fragment with genetic codes of organisms expressed in alphanumeric terms (“natural key”). Because of limited space dimensions in table we suggest the examples of such fragments of codes for all three fish species separately in Fig. 6, a–c.

All other keys in Fig. 1 serve as “alternate keys”.

As it was demonstrated above, for the construction of functionally successful relation database is necessary to make ensure reliable connections between the tables (transformed “objects”) in this DB. The groups of random symbols (numbers or and letters) is the method that is used widely for these purposes for today. Such keys in our case are surrogate keys.

The disadvantage of conventional method [5—7] is that in technical systems the data integ-
rity is ensured in databases by keys (including primary keys) formed predominantly in standard way — by selecting combinations of sequences of random numbers. This method in some cases can lead to data integrity violations, to interruptions of information flows, and etc. In case of biomedical information system such integrity violations can cause the deterioration of this information system performance and, consequently, the deterioration of the quality of monitoring of environmental effects of chemicals on living organisms in the nature.

From other side, contemporary genetics demonstrates incredible successes in to-day scientific biomedical investigations. So, other type of the sequences can be suggested — genetic sequences. For today, the structures of numerous fragments of genetic sequences have been disclosed for some biological organisms, as well as for some individual persons [1—4, 6]. Such structures of fragments of genetic sequences can be drawn abstractly as sequences of numbers and letters. In many cases such data are already included into the biomedical databases, because the genetic information is quite important characteristics for each biological species as well as for individual person or animal. Sure in all listed cases it is better (or so metimes necessary!) to pick up really specific fragment of genetic code — fragment that characterizes exactly the species, or individual person, animal, tissue, etc. And contemporary genetics gives this opportunity — for numerous cases individual “characteristic” fragments can be found and used for databases construction.

Fig. 6. Three examples of the fragments of genetic codes of various fish species that can serve as keys for databases construction (as well as primary keys) [9, 51, 52]. Genetic codes are represented in form of letter and number codes (illustrative image, explanations see in text below). Consequently mean genetic codes of *Oncorhynchus mykiss* Walbaum (a), *Cyprinus carpio* L. (b) (key “GenCodeCarp” from our example, see in text), and *Barbus boryshenicus* Dybowsky (c).
Innovation Approaches to the Construction of Bioinformation Systems with Databases: Keys Based on Genetic Codes

with genetic information too, because of its importance. Such databases are used in medicine, criminology, police and military practice, and etc. But our idea is in another.

We had proposed to use the fragments of genetic codes sequences, written in the numbers and letters as keys for databases construction. Especially this is important for primary key, because it have to be based on primary information, unique for each species or person — like passport code is unique for each individual person. Realization of this idea is demonstrated on the figures below (Fig. 6).

For the realization of this idea practically we see two possible ways:

1. The code from the field “Genetic sequence” is used as the key by itself. This type of the code is shown in Fig. 5 with a dotted arrow (alphanumeric fragment of genetic code has to be moved to the field “key” — so, used as key). In this case the separated field for genetic code fragment is not necessary anymore. So, one of the fields with keys — either surrogate keys, or one that previously contained natural keys can be deleted — the database construction became more perfect.

2. The second possible way — if keys in the databases are synthetized on the base of genetic codes fragments (intelligent keys). The simplest way is to unite a group of any characters (expressed, for example, in alphanumeric form) with the fragment of genetic code (coded in alphanumeric form too). In this case both fields can be left in the database. This method has its advantages too, it will be observed in details in our future publications.

Each of these approaches has its own advantages and disadvantages. It is necessary to emphasize that in each case of genetic code fragment usage as key, it is better to use those fragment of genetic code, why either specific for each person or species (or those that demonstrate logical relations with individual person or species). And such a method, of course, should best suit the problem being solved. In any case, the use of fragments of genetic codes as keys in relational databases has good prospects for solving problems at the intersection of two huge areas of knowledge — biology and technology.

The technical result that can be obtained by carrying out this invention is that the proposed method allows for monitoring the effects of chemicals much better, more efficiently, with higher accuracy than prototypes [9–11]. The availability of the developed databases included into the information system allows collecting, order and systematizing great amount of the data about biological organisms, as well as these data storage at the different levels of processing, which can be replenished in case of new observations or during experiments. In case of developed Web-interface it provides the access to these data directly to remote scientists or other users. All these tools allow processing of large volumes of the data, using all the advantages of network technologies, and the successful operation of the information subsystem. Much better than in case of prototypes, the data integrity is provided in our invented method, where in databases the keys (including primary keys) were used being alphanumerically coded on the base of the fragments of genetic code sequences of nucleotides of fishes and other aquatic organisms. Described invented principles can also be used for other biomedical information systems development [6, 7, 9–11, 37], and in other contemporary databases like [53–54].

Indeed, the database developer may have a number of questions during his work, for example: “What exactly is considered as a key example? This whole fragment of genetic code? Or its small part? How many nucleotides?” and so on. In fact, the answers to these questions when creating each DB are the solution of beautiful task. Solution depends on the purpose of work, talent and level of professional competence of each specific developer. Yes, the developer can choose any of these options depending on: the task for which this DB is being created, the need to protect personal information in the DB and in the system (especially in the case
of the sequence of human nucleotides), his own preferences, the amount of memory computer system and a number of other factors. So, for example, in Fig. 6, the developer can choose any sequence of nucleotides or the entire sequence, which, for example, is presented in Fig. 6, c, or only its fragment. The main determining factor of the correctness of the choice should be the result — the creation of a one-to-one relationship between two relevant tables in the DB, reliable access to the information in the relevant table through this relationship and successful retrieval of complete information from it, solving the problems of information protection in the system, protecting personal data (in the case of a person, etc.), and so on. The proposed method is a powerful tool for solving this entire range of tasks.

Basing on deep theoretical analysis and generalizations, we demonstrated how the ER-diagram for two types of main objects can be designed practically — “fish species” and “chemical compounds”. As example we demonstrated this for three fish species (rainbow trout, common carp, and Dnieper barbell) and organic chemical compounds — known environmental pollutants from industrial regions of Ukraine. Further we had demonstrated how relational tables for biological databases were formed, and in what way they can be linked by “key” tools.

The first logical step in our article — the algorithm of construction of the DB with information about fishes has been described. At this step of DB construction some practical recommendations for the development of databases with information about domestic fishes may be done: maximal accuracy in data mining, standardization of input data about fishes, using the approach of object-oriented system analysis with further ER-diagram design, and so on.

Then we demonstrated the results of the work on the use of the fragments of genetic codes as keys in alphanumeric expression for the construction of DBs with biomedical information. According to it, the keys in databases can be formed on the base of the fragments of genetic codes, if to write them as alphanumerical sequences. In the best (abstract) case such sequences have to characterize, respectively, each individual person, as well as individual species of biological organisms, for example species of fish. This invented method of keys coding basing on the principles of genetic codes is really potent. This is really important for databases of individual persons or for biological species. These results can be used for the creation of appropriate information systems, including bio-information system for environmental monitoring.

We would like to emphasize the originality of results in this article. The method that we suggested and described in present article differs from the other ones. As it was mentioned above, four patents of Ukraine were obtained for the material put in base of present article [9—12]. Applying this method the integrity of information, its protection for individual objects in databases may increase; better continuity of information flows can be achieved.

Obtained results have theoretical and practical value for further development of methods of databases construction, improving some methods of data protection and can be essential for the formation of environmental and food safety. We would like to hope that practical implementation of results of our work will have good prospects.

CONCLUSIONS

1. An innovative method for constructing of relational databases of biological organisms and biological material is considered, applied and described. It allows organizing of a large amount of the data on living organisms and biological materials in general, including genetic material. The article discusses in detail the construction of such database using the example of a fish database. But this method can be applied to any living organisms (including humans) and biological material. The algorithm of databases design is described.

2. Special attention is paid to the solution of the problem of keys creation basing on the genetic codes of living organisms in alphanumeric exp-
Innovation Approaches to the Construction of Bioinformation Systems with Databases: Keys Based on Genetic Codes

3. The proposed approaches of using the keys of this type have both theoretical and practical significance. The theoretical meaning is that the integrity of the information in the system and the reliability of access to it (and its retrieval) are ensured better in such a way. The high level of data individualization with using the keys based on genetic codes in such database is analyzed and substantiated.

4. The practical significance lies in the possibility of computer memory saving when applying the proposed approach, as well as improving some data protection methods. Numerous tasks of information protection in databases, computer systems in whole can be solved successfully by this suggested innovative method.

5. Another aspect of practical importance is that the obtained results can be applied to create appropriate information systems, in particular bioinformation systems, with which such databases are connected. The article examines an example of creating a database with information about fish and chemical substances that pollute environmental waters. The creation of such databases as elements of information systems can become necessary tool for the performing of the works to improve of environmental situation in Ukraine after accidents, disasters, in areas of armed operations.

6. The proposed approaches can be useful for solving the tasks of creating databases containing biomedical and genetic material for peaceful life (databases for medical institutions, for biology, police and rescuers (databases of criminal persons, consequences of accidents, disasters, etc.), as well as in wartime (with data on military personnel, their relatives, for identification of the dead and seriously wounded persons, biological materials, etc.).

REFERENCES


Innovation Approaches to the Construction of Bioinformation Systems with Databases: Keys Based on Genetic Codes


Received 12.05.2023
Revised 02.03.2024
Accepted 02.03.2024
Вступ. Застосування досягнень біології, біофізики у техніці відкриває нові можливості для інновацій, зокрема у технологіях створення релационних баз даних (БД) із біомедичними даними, сприяє вирішенню завдань, отриманню результатів на якісно новому рівні.

Проблематика. Розробка інформаційних систем із біомедичною інформацією є актуальною як у мирний час, так і під час війни. Впровадження сучасних інформаційно-комп’ютерних технологій для розробки інформаційних систем з БД у біології та медицині має свою специфіку, тому актуальним є розробка інноваційних підходів до конструювання біомедичних релационних БД з ключами із розширеними можливостями.

Мета. Розробка та конструювання біомедичних релационних БД з ключами на основі генетичних кодів організмів у буквено-цифровому вираженні із подальшим застосуванням у складі новітніх біоінформаційних систем.

Матеріали і методи. Методи об’єктивно-орієнтованого системного аналізу для побудови оптимальним чином БД з біомедичною інформацією, метод проектування ER-діаграм, методи конструювання БД.

Результати. На прикладі релационної БД з інформацією про деякі види різ розглянуто, застосовано та описано підхід об’єктивно-орієнтованого аналізу для оптимального конструювання БД, описано алгоритм їх конструювання. Особливу увагу приділено вирішенню проблеми створення ключів на основі генетичних кодів різ у буквено-цифровому вираженні, особливо як первинних ключів, що забезпечують зв’язки між окремими таблицями БД, її полісність інформації у системі, надійність доступу до неї. Проаналізовано і обґрунтовано високий рівень індивідуалізації даних при застосуванні ключів на основі генетичних кодів у такій БД.

Висновки. Результати можуть бути застосовані для створення відповідних інформаційних систем, зокрема і біоінформаційних. Вони мають як теоретичне значення для подальшого розвитку технологій розробки БД, так і практичне, уособлюючи деякі методи захисту даних, та можуть бути корисними для вирішення завдань створення БД з біоматеріалом у мирному житті та у воєнний час.

Ключові слова: інформаційні технології, інформаційна система, база даних, первинний ключ, кодування, біомедична інформація, хімічні речовини.