SCIENTIFIC BASIS OF INNOVATION

https://doi.org/10.15407/scine20.04.033

KLYUCHKO, O. M.[1] (https://orcid.org/0000-0003-4982-7490),
BELETSKY, A. Ya.[1] (https://orcid.org/0000-0002-3798-8150),
MELEZHYK, O. V.[2] (https://orcid.org/0000-0003-3882-7102),
and GONCHAR, O. O.[3] (https://orcid.org/0000-0003-4134-03544)

[1] National Aviation University,
1, L. Huzar Ave., Kyiv, 03058, Ukraine,
+380 44 406 7130, post@nau.edu.ua
[2] Open International University of Human Development “Ukraine,”
23, Lvivska St., Kyiv, 03115, Ukraine,
+380 67 328 2822, office@uu.edu.ua
[3] Bogomolets Institute of Physiology National Academy of Sciences of Ukraine,
4, Bogomolets St., Kyiv, 01024, Ukraine,
+380 44 256 2400, science@biph.kiev.ua

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*

Citation: Klyuchko, O. M., Beletsky, A. Ya., Melezhyk, O. V., and Gonchar, O. O. (2024). Innovation Approaches to the Construction of Bioinformation Systems with Databases: Keys Based on Genetic Codes. *Sci. innov.*, 20(4), 33—48. https://doi.org/10.15407/scine20.04.033

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, harmonization 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-

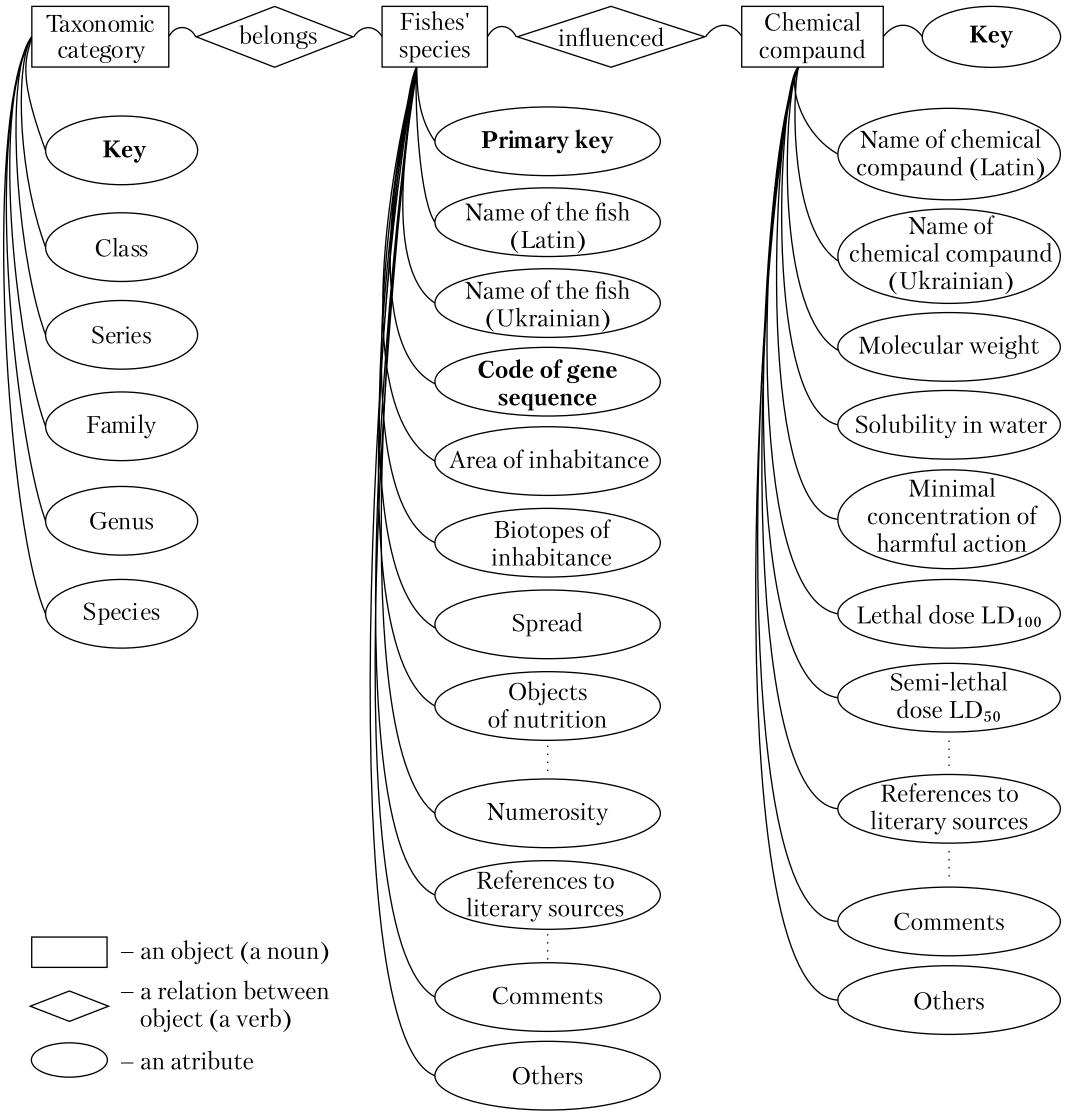


Fig. 1. Core fragment of ER-diagram "Fishes' Species and Organic Chemical Compounds Affected Them" for different fish species

teristics and relations at the first stage of relation databases construction as necessary and recommended procedure. In our case the majority of such objects are biomedical ones. Secondly they, 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 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).

Object “Fishes' species”

Attributes:

Primary key (based on the fragment of genetic sequence)
Name of the fish (Latin)
Name of the fish (Ukrainian)
Code of genetic sequence (if it is known)
Area of inhabitance
Biotopes of inhabitance
Spread
Location
Numerosity
Reasons of numerosity changes
Features of biology
Objects of nutrition
Reproduction in captivity
Protection measures
....
References to literary sources
....
Comments
Others

Object “Chemical compound”

Attributes:

Key
Name of chemical compound (Latin)
Name of chemical compound (Ukrainian)
Molecular weight
Solubility in water
Minimal concentration of harmful action
Lethal dose LD_{100}
Semi-lethal dose LD_{50}
....
References to literary sources
....
Comments
Others

Fig. 2. Characteristics of the objects “Fishes’ species” and “Chemical compounds” [48, 50]

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”; 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].

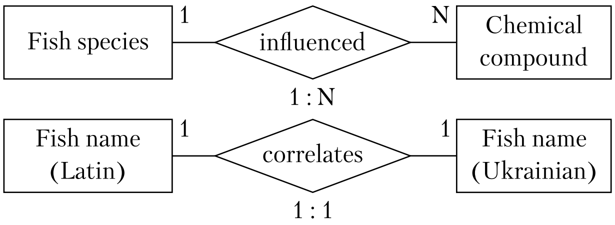


Fig. 3. Power of relations between two objects in process of the database construction
Above: “1 : N” — “one-to-many”. In framework of our task solved, this means: “1 species of fishes was affected by N chemical compounds”
Below: “1 : 1” — “one-to-one”. In our example, this means: “1 name of fish species in Latin correlates with 1 name of fish species in Ukrainian” (correlation between spelling of names in Latin and Cyrillic fonts)

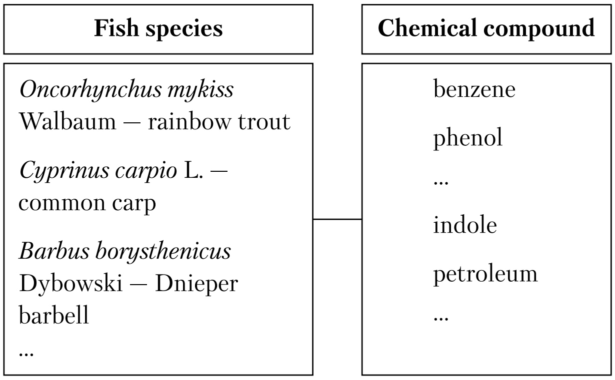
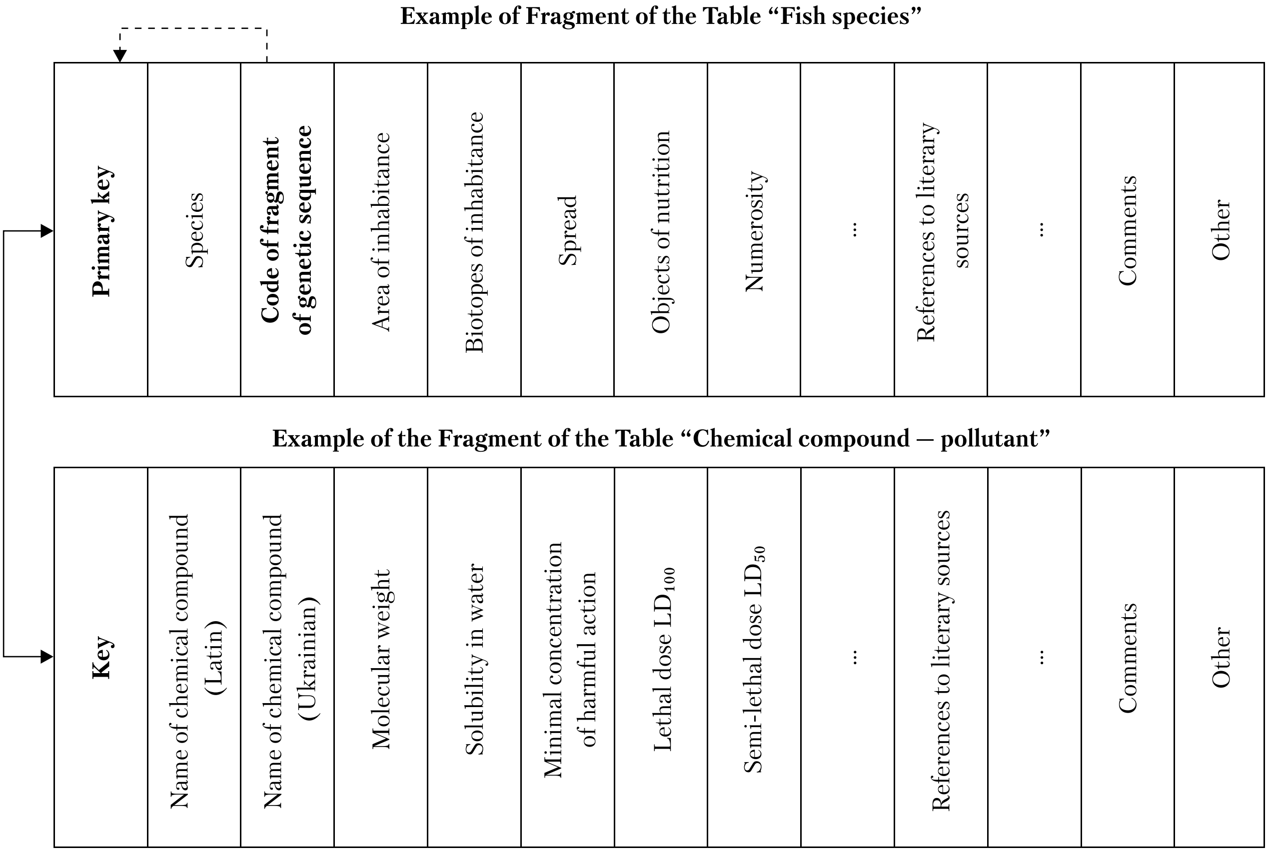


Fig. 4. Logical model of DB with the data about some fish species and organic chemical compounds — water pollutants they are affected by. Fragment of ER-diagram “Fish species” and “Organic Chemical Compounds Affected Them” on the example of three fish species from Ukrainian water basins [48—50]

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



Example of Fragment of the Table "Fish species"

Primary key	Species	**Code of fragment of genetic sequence**	Area of inhabitance	Biotopes of inhabitance	Spread	Objects of nutrition	Numerosity	...	References to literary sources	...	Comments	Other

Example of the Fragment of the Table "Chemical compound — pollutant"

Key	Name of chemical compound (Latin)	Name of chemical compound (Ukrainian)	Molecular weight	Solubility in water	Minimal concentration of harmful action	Lethal dose LD_{100}	Semi-lethal dose LD_{50}	...	References to literary sources	...	Comments	Other

Fig. 5. Objects and tables in constructed relative database

same procedure has to be done with the third object in our example "Taxonomic category."

So, if to summarize [6—8, 44, 45],

1. *Objects* ("*entities*") of ER-diagram were transformed into separate tables;
2. *Attributes* were transformed into the columns ("fields") of the tables;
3. The relations between the objects were transformed 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 environmental pollutants, and (with future DB development) in which particular basin. Using such databases, professional can easily perform all operations 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 professional work because the users who works in this field can add to the previous material new information step by step (data input into the DB), or list the species alphabetically (sorting of the data), or select all species from one genus (filtration 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 information in biomedical DB and its level of individualization, 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 great majority of cases serves any preferred 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.

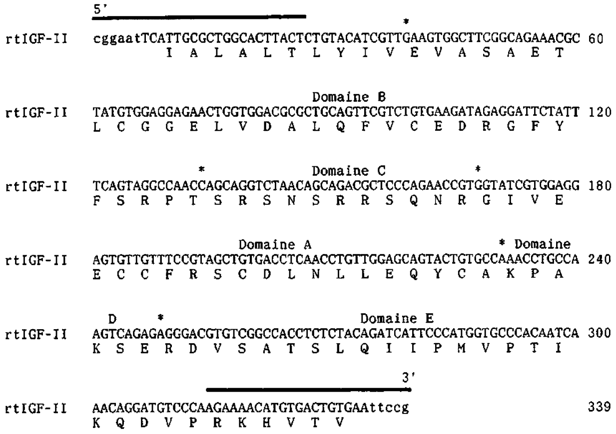
Figure 5 shows that "keys" serve to make the links between those fields of the tables that have to be linked logically. For example, it is necessary to find from the database how any defined chemical substance (for example, phenol) influences on the carp (or carp population). Programmer knows that phenol in the water really influences on carp. 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*.

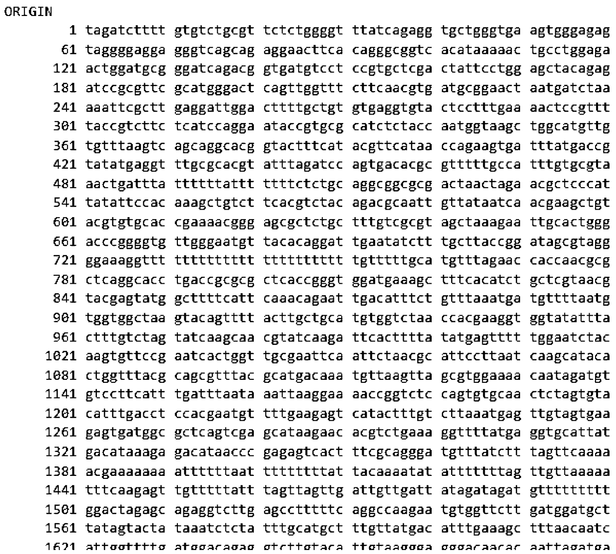
All other keys in Fig. 1 serve as "alternate keys". As it was demonstrated above, for the construction of functionally successful relation database is necessary to make ensure reliable connections between the tables (transformed "objects") in this DB. The groups of random symbols (numbers or/and letters) 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-



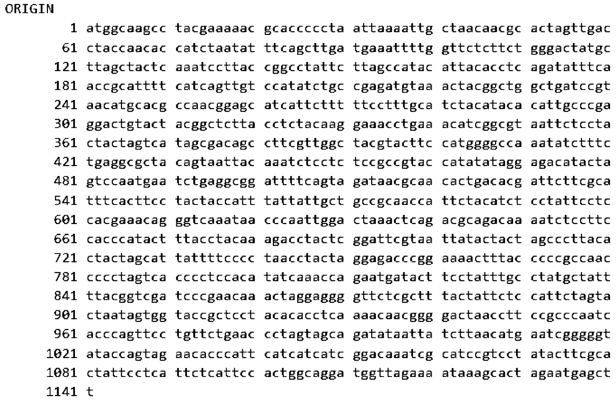
```
          5'                                       *
rtIGF-II  cggaatTCATTGCGCTGGCACTTACTCTGTACATCGTTGAAGTGGCTTCGGCAGAAACGC 60
                I  A  L  A  L  T  L  Y  I  V  E  V  A  S  A  E  T

                                         Domaine B
rtIGF-II  TATGTGGAGGAGAACTGGTGGACGCGCTGCAGTTCGTCTGTGAAGATAGAGGATTCTATT 120
          L  C  G  G  E  L  V  D  A  L  Q  F  V  C  E  D  R  G  F  Y

                    *                    Domaine C              *
rtIGF-II  TCAGTAGGCCAACCAGCAGGTCTAACAGCAGACGCTCCCAGAACCGTGGTATCGTGGAGG 180
          F  S  R  P  T  S  R  S  N  S  R  R  S  Q  N  R  G  I  V  E

                            Domaine A                     * Domaine
rtIGF-II  AGTGTTGTTTCCGTAGCTGTGACCTCAACCTGTTGGAGCAGTACTGTGCCAAACCTGCCA 240
          E  C  C  F  R  S  C  D  L  N  L  L  E  Q  Y  C  A  K  P  A

            D     *                            Domaine E
rtIGF-II  AGTCAGAGAGGGACGTGTCGGCCACCTCTCTACAGATCATTCCCATGGTGCCCACAATCA 300
          K  S  E  R  D  V  S  A  T  S  L  Q  I  I  P  M  V  P  T  I

                                             3'
rtIGF-II  AACAGGATGTCCCAAGAAAACATGTGACTGTGAAttccg                      339
          K  Q  D  V  P  R  K  H  V  T  V
```

a



```
ORIGIN
        1 tagatctttt gtgtctgcgt tctctggggt ttatcagagg tgctgggtga agtgggagag
       61 taggggagga gggtcagcag aggaacttca cagggcggtc acataaaaac tgcctggaga
      121 actggatgcg ggatcagacg gtgatgtcct ccgtgctcga ctattcctgg agctacagag
      181 atccgcgttc gcatgggact cagttggttt cttcaacgtg atgcggaact aatgatctaa
      241 aaattcgctt gaggattgga cttttgctgt gtgaggtgta ctcctttgaa aactccgttt
      301 taccgtcttc tcatccagga ataccgtgcg catctctacc aatggtaagc tggcatgttg
      361 tgtttaagtc agcaggcacg gtactttcat acgttcataa ccagaagtga tttatgaccg
      421 tatatgaggt ttgcgcacgt atttagatcc agtgacacgc gtttttgcca tttgtgcgta
      481 aactgattta ttttttattt ttttctctgc aggcggcgcg actaactaga acgctcccat
      541 tatattccac aaagctgtct tcacgtctac agacgcaatt gttataatca acgaagctgt
      601 acgtgtgcac cgaaaacggg agcgctctgc tttgtcgcgt agctaaagaa ttgcactggg
      661 acccggggtg ttgggaatgt tacacaggat tgaatatctt tgcttaccgg atagcgtagg
      721 ggaaaggttt tttttttttt tttttttttt tgtttttgca tgtttagaac caccaacgcg
      781 ctcaggcacc tgaccgcgcg ctcaccgggt ggatgaaagc tttcacatct gctcgtaacg
      841 tacgagtatg gcttttcatt caaacagaat tgacatttct gtttaaatga tgttttaatg
      901 tggtggctaa gtacagtttt acttgctgca tgtggtctaa ccacgaaggt ggtatattta
      961 ctttgtctag tatcaagcaa cgtatcaaga ttcactttta tatgagtttt tggaatctac
     1021 aagtgttccg aatcactggt tgcgaattca attctaacgc attccttaat caagcataca
     1081 ctggtttacg cagcgtttac gcatgacaaa tgttaagtta gcgtggaaaa caatagatgt
     1141 gtccttcatt tgatttaata aattaaggaa aaccggtctc cagtgtgcaa ctctagtgta
     1201 catttgacct ccacgaatgt tttgaagagt catactttgt cttaaatgag ttgtagtgaa
     1261 gagtgatggc gctcagtcga gcataagaac acgtctgaaa ggttttatga ggtgcattat
     1321 gacataaaga gacataaccc gagagtcact ttcgcaggga tgtttatctt tagttcaaaa
     1381 acgaaaaaaa atttttaat ttttttttat tacaaaatat atttttttag ttgttaaaaa
     1441 tttcaagagt tgtttttatt tagttagttg attgttgatt atagatagat gttttttttt
     1501 ggactagagc agaggtcttg agccttttc aggccaagaa tgtggttctt gatggatgct
     1561 tatagtacta taaatctcta tttgcatgct ttgttatgac atttgaaagc tttaacaatc
     1621 attggttttg atggacagag gtcttgtaca ttgtaaggga gggacaacac aattagatga
```

b



```
ORIGIN
        1 atggcaagcc tacgaaaaac gcacccccta attaaaattg ctaacaacgc actagttgac
       61 ctaccaacac catctaatat ttcagcttga tgaaattttg gttctcttct gggactatgc
      121 ttagctactc aaatccttac cggcctattc ttagccatac attacacctc agatatttca
      181 accgcatttt catcagttgt ccatatctgc cgagatgtaa actacggctg gctgatccgt
      241 aacatgcacg ccaacggagc atcattcttt ttcctttgca tctacataca cattgcccga
      301 ggactgtact acggctctta cctctacaag gaaacctgaa acatcggcgt aattctccta
      361 ctactagtca tagcgacagc cttcgttggc tacgtacttc catggggcca aatatctttc
      421 tgaggcgcta cagtaattac aaatctcctc tccgccgtac catatatagg agacatacta
      481 gtccaatgaa tctgaggcgg attttcagta gataacgcaa cactgacacg attcttcgca
      541 tttcacttcc tactaccatt tattattgct gccgcaacca ttctacatct cctattcctc
      601 cacgaaacag ggtcaaataa cccaattgga ctaaactcag acgcagacaa aatctccttc
      661 cacccatact ttacctacaa agacctactc ggattcgtaa ttatactact agcccttaca
      721 ctactagcat tattttcccc taacctacta ggagacccgg aaaactttac ccccgccaac
      781 cccctagtca cccctccaca tatcaaacca gaatgatact tcctatttgc ctatgctatt
      841 ttacggtcga tcccgaacaa actaggaggg gttctcgctt tactattctc cattctagta
      901 ctaatagtgg taccgctcct acacacctca aaacaacggg gactaacctt ccgcccaatc
      961 acccagttcc tgttctgaac cctagtagca gatataatta tcttaacatg aatcggggt
     1021 ataccagtag aacacccatt catcatcatc ggacaaatcg catccgtcct atacttcgca
     1081 ctattcctca ttctcattcc actggcagga tggttagaaa ataaagcact agaatgagct
     1141 t
```

c

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 borysthenicus* Dybowski (*c*)

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 sometimes 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.

Sure, contemporary practice demonstrates the examples of the databases that include the fields 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. With large number of records in the database, significant amount of computer memory can be freed in such a way.

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 bioinformation 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-

ression (especially as primary keys) that provide connections between individual tables of such database.

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

1. Guidelines for Human Biobanks and Genetic Research Databases (HBGRDs). URL: https://www.oecd.org/sti/emerging-tech/guidelines-for-human-biobanks-and-genetic-research-databases.htm (Last accessed: 20.04.2023).
2. Fokkema, I. F. A. C., Kroon, M., López Hernández, J. A., Asscheman, D., Lugtenburg, I., Hoogenboom, J., den Dunnen, J. T. (2021). The LOVD3 platform: efficient genome-wide sharing of genetic variants. *Eur. J. Hum. Genet.,* 29(12), 1796—1803. https://doi.org/10.1038/s41431-021-00959-x.
3. Van Ness, Lindsey. DNA Databases Are Boon to Police But Menace to Privacy, Critics Say. URL: https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2020/02/20/dna-databases-are-boon-to-police-but-menace-to-privacy-critics-say (Last accessed: 20.04.2023).
4. Brownsword, R. Genetic Databases and the First Signs of Regulatory Opportunity (Chapter 8 in: “Rights, Regulation and Technological Revolution”) URL: https://academic.oup.com/book/5017/chapter-abstract/147524388?redirectedFrom=fulltext (Last accessed: 20.04.2023).
5. Klyuchko, O. M. (2018). Electronic expert systems for biology and medicine. *Biotechnologia Acta,* 11(6), 5—28. https://doi.org/10.15407/biotech11.06.005
6. Klyuchko, O. M., Buchatsky, L. P., Rud, Yu. P., Melezhyk, O. V. (2019). Creation of fish databases for electronic interactive map: tables and keys. *Fisheries science of Ukraine*, 50(4), 37—57. https://doi.org/10.15407/fsu2019.04.037
7. Klyuchko, O. M., Klyuchko, Z. F. (2018). Electronic information systems for monitoring of populations and migrations of insects. *Biotechnologia Acta*, 11(5), 5—25. https://doi.org/10.15407/biotech11.05.005
8. Klyuchko, O. M., Klyuchko, Z. F. (2018). Electronic databases of Arthropods: methods and applications. *Biotechnologia Acta*, 11(4), 28—49. https://doi.org/10.15407/biotech11.04.028
9. *Patent UA 143919 U.* Method of application of monitoring system with databases and keys in symbolic records of genetic codes of fish and other aquatic organisms. Klyuchko, O. M., Biletsky, A. Ya., Lizunova, A. G. [in Ukrainian].
10. *Patent UA 143918 U.* Method of applying monitoring system with databases and keys in form of images of genetic codes of fish and other aquatic organisms. Klyuchko, O. M., Biletsky, A. Ya., Lizunova, A. G. [in Ukrainian].
11. *Patent UA 143926 U.* Method of use of monitoring system with databases and keys in form of images of genetic codes of biological organisms. Klyuchko, O. M., Biletsky, A. Ya., Lizunova, A. G. [in Ukrainian].

12. *Patent UA 155203 U.* Method of application of information monitoring system with databases and keys in symbolic records of genetic codes of biological organisms. Klyuchko, O. M., Biletsky, A. Ya., Lizunova, A. G. G01N 33/00, G01N 33/50, G06F 16/00 [in Ukrainian].
13. Klyuchko, O. M. (2017). On the mathematical methods in biology and medicine. *Biotechnologia Acta*, 10(3), 31—40. https://doi.org/10.15407/biotech10.03.031
14. Klyuchko, O. M., Onopchuk, Yu. M. (2018). Some trends in mathematical modeling for biotechnology. *Biotechnologia Acta*, 11(1), 39—57. https://doi.org/10.15407/biotech11.01.039
15. Eronen, L., Toivonen, H. (2012). Biomine: predicting links between biological entities using network models of heterogeneous databases. *Bioinformatics,* 13, 119. https://doi.org/10.1186/1471-2105-13-119
16. Bor-Sen, Chen, Shih-Kuang, Yang, Chung-Yu, Lan, Yung-Jen, Chuang. (2008). A systems biology approach to construct the gene regulatory network of systemic inflammation via microarray and databases mining. *Medical Genomics,* 1, 46. https://doi.org/10.1186/1755-8794-1-46
17. Pornputtapong, N., Wanichthanarak, K., Nilsson, A., Nookaew, I., Nielsen, J. (2014). A dedicated database system for handling multi-level data in systems biology. *Source Code for Biology and Medicine,* 9, 17. https://doi.org/10.1186/1751-0473-9-17.
18. Stobbe, M. D., Swertz, M. A., Thiele, I., Rengaw, T., van Kampen, A. H. C., Moerland, P. D. (2013). Consensus and conflict cards for metabolic pathway databases. *Systems Biology,* 7, 50. https://doi.org/10.1186/1752-0509-7—50
19. van Ommen, B., Bouwman, J., Dragsted, L. O., Drevon, C. A., Elliott, R., ..., Wopereis, S. (2010). Challenges of molecular nutrition research 6: the nutritional phenotype database to store, share and evaluate nutritional systems biology studies. *Genes & Nutrition,* 5, 167. https://doi.org/10.1007/s12263-010-0167-9
20. Momin, A. A., James, B. P., Motter, T. C., Kadara, H. N., Powis, G., Wistuba, I. I. (2011). Integrating whole transcriptome sequence data and public databases for analysis of somatic mutations in tumors. *Genome Biology,* 12(1), 44. https://doi.org/10.1186/gb-2011-12-s1-p44
21. Stobbe, M. D., Houten, S. M., Jansen, G. A., van Kampen, A. H. C., Moerland, P. D. (2011). Critical assessment of human metabolic pathway databases: a stepping stone for future integration. *Systems Biology*, 5, 165. https://doi.org/10.1186/1752-0509-5-165
22. Chowbina, S. R., Wu, X., Zhang, F., Li, P. M., Pandey, R., Kasamsetty, H. N., Chen, J. Y. (2009). HPD: an online integrated human pathway database enabling systems biology studies. *Bioinformatics*, 10(11), S5. https://doi.org/10.1186/1471-2105-10-S11-S5
23. Goldstein, A. M. (2010). The NCBI Databases: an Evolutionist's Perspective. *Evolution: Education and Outreach*, 3, 258. https://doi.org/10.1007/s12052-010-0258-5
24. Albà, M. (2000). Links to molecular biology databases. *Genome Biology*, 1—4, https://doi.org/10.1186/gb-2000-1-1-reports235
25. Alfieri, R., Merelli, I., Mosca, E., Milanesi, L. (2007). A data integration approach for cell cycle analysis oriented to model simulation in systems biology. *Systems Biology,* 1, 35. https://doi.org/10.1186/1752-0509-1-35
26. Walsh, J. R., Sen, T. Z., Dickerson, J. A. (2014). A computational platform to maintain and migrate manual functional annotations for BioCyc databases. *Systems Biology*, 8, 115.
27. Tedesco, P. A., Beauchard,**O**., Bigorne, R., Blanchet, S., Buisson, L., ..., Oberdorff, T. (2017). A global database on freshwater fish species occurrence in drainage basins. *Sci. Data*, 4, 170141. https://doi.org/10.1038/sdata.2017.141
28. Maier, C. V., Long, J. G., Hemminger, B. M., Giddings, M. C. (2009). Ultra-Structure database design methodology for managing systems biology data and analyses. *Bioinformatics,* 10, 254. https://doi.org/10.1186/1471-2105-10-254
29. Tan, T. W., Xie, C., De Silva, M., Kuan Siong, Lim, Patro, C. P. K., ..., Khan, A. M. (2013). Simple re-instantiation of small databases using cloud computing. *Genomics*, 14(5), 5—13. https://doi.org/doi: 10.1186/1471-2164-14-S5-S13
30. Bouzaglo, D., Chasida, I., Tsur, E. E. (2018). Distributed retrieval engine for the development of cloud-deployed biological databases. *BioData Mining*, 11, 26. https://doi.org/10.1186/s13040-018-0185-5
31. Schnase, J. L., Cushing, J., Frame, M., Frondorf A., Landis E., Maier D., Silberschatz A. (2003). Information technology challenges of biodiversity and ecosystems informatics. *Information Systems,* 28(4), 339—345. https://doi.org/10.1016/S0306-4379(02)00070-4
32. Chen, Y., Shi, M., Cheng, Y., Zhang, W., Tang, Q., Xia Q. (2018). FVD: The fish-associated virus database. *Infect. Genet.,* 58, 23—26. https://doi.org/10.1016/j.meegid.2017.11.004
33. User Reference for Fisheries Improvement Projects Database (FIP-DB) and Query Viewer. URL: https://ru.scribd.com/document/385739269/Readme-File-for-FIP-DB#download (Last accessed: 27.12.2022).
34. Froese, R., Pauli, D. (2000). *Fish Base 2000: Concepts, designs and data sources.* ICLARM. Los Banos, Philippines.

35. Arnot, J. A., Mackay, D., Parkerton, T. F., Bonnell, M. (2008). A database of fish biotransformation rates for organic chemicals. *Environmental Toxicology and Chemistry*, 27(11), 2263—2270. URL: https://setac.onlinelibrary.wiley.com/doi/full/10.1897/08-058.1. https://doi.org/10.1897/08-058.1 (Last accessed: 20.04.2023).
36. Del Rio, A., Barbosa, A. J. M., Caporuscio, F. (2011). Use of large multiconformational databases with structure-based pharmacophore models for fast screening of commercial compound collections. *Journal of Cheminformatics,* 3(1), P27. https://doi.org/10.1186/1758-2946-3-S1-P27
37. *Patent UA 131863 U.* Method of application of biotechnical monitoring system with expert subsystem and biosensor. Klyuchko, O. M., Biletsky, A. Y., Navrotskyi, D. A. [in Ukrainian].
38. Daoliang, Li, Zetian, Fu, Yanqing Duan. (2002). Fish-Expert: a web-based expert system for fish disease diagnosis. *Expert Systems with Applications,* 23, 311—320. https://doi.org/10.1016/S0957-4174(02)00050-7
39. Di Génova, A. D., Aravena, A., Zapata, L., González, M., Maass, A., Iturra, L. (2011). Salmon DB: a bioinformatics resource for *Salmo salar* and *Oncorhynchus mykiss. Database (Oxford).* URL: https://doi.org/10.1093/database/bar050 (Last accessed: 20.04.2023).
40. Tkachuk, R., Tkachuk, A., Stadnik, D., Yanenko, O. Ensuring high-precision testing of implants in the regulation of intra-eye pressure. (2021). *Intl. Conf. Advanced Applied Energy and Information Technologies (15—17 December, 2021, Ternopil, Ukraine),* 157—161. Ternopil.
41. Yukalo, V. G., Storozh, L. A., Datsyshyn, K. Ye., Krupa, O. M. (2018). Electrophoretic systems for the preparative fractionation of protein precursors of bioactive peptides from cow's milk. *Food science and technology*, 12(2), 26—32. https://doi.org/10.15673/fst.v12i2.932
42. Yukalo, V., Datsyshyn, K., Krupa, O., Pavlistova, N. (2019). Obtaining of β-LG, α-LA and BSA protein fractions from milk whey. *Ukrainian Food Journal,* 8(4), 788—798. https://doi.org/10.24263/2304-974X-2019-8-4-10
43. Gonchar, O. O., Maznychenko, A. V., Klyuchko, O. M., Mankovska, I. M., Butowska, K., Borowik, A., Piosik, Ja., Sokolowska, I. (2021). C_{60} Fullerene Reduces 3-Nitropropionic Acid-Induced Oxidative Stress Disorders and Mitochondrial Dysfunction in Rats by Modulation of P53, Bcl-2and Nrf2 Targeted Proteins. *International Journal of Molecular Sciences*, 22(11), 5444—5468. https://doi.org/10.3390/ijms22115444
44. Microsoft Academy: Methods and means of software engineering. URL: https://www.intuit.ru/studies/courses/2190/237/lecture/6124 (Last accessed: 24.04.2023).
45. Harrington, Jan L. (2005). *Object-oriented database design clearly explained.* USA: Academic Press.
46. Van der Laan, R., Eschmeyer, W. N., Fricke, R. (2014). Family-group names of Recent fishes. *Zootaxa Monograph.*, 3882 (1), 1—230. https://doi.org/10.11646/zootaxa.3882.1.1
47. Movchan, Yu. V. (2009). Fishes of Ukraine (taxonomy, nomenclature, remarks). *Collection of works of Zoological Museum,* 40, 47—87.
48. *Red Book of Ukraine.* Fauna (2009). Kyiv [in Ukrainian].
49. Fricke, R., Eschmeyer, W. N., R. van der Laan. (2019) Eschmeyer's catalog of fishes: genera, species. URL: http://researcharchive.calacademy.org/research/ichthyology/catalog (Last accessed: 24.04.2023).
50. Franchuk, G. M., Isaenko, V. M. (2005). Ecology, aviation and cosmos. Kyiv. 456 p. [in Ukrainian].
51. *Cyprinus carpio* isolate SPL01 chromosome A17, ASM1834038v1, whole genome shotgun sequence. URL: https://www.ncbi.nlm.nih.gov/nuccore/NC_056588.1?from=25191087&to=25198977&report=genbank&strand=true (Last accessed: 23.04.2023).
52. *Barbus borysthenicus* isolate PK-977 cytochrome b (cytb) gene, partial cds; mitochondrial gene for mitochondrial product. URL: https://www.ncbi.nlm.nih.gov/nuccore/AY331026.1 (Last accessed: 23.04.2023).
53. Human gene databases. URL: https://www.genecards.org/ (Last accessed: 23.04.2023).
54. Human genetic research databases. URL: https://www.alrc.gov.au/publication/essentially-yours-the-protection-of-human-genetic-information-in-australia-alrc-report-96/18-human-genetic-research-databases/what-are-human-genetic-research-databases/ (Last accessed: 23.04.2023).

Received 12.05.2023
Revised 02.03.2024
Accepted 02.03.2024

О.М. Ключко[1] (https://orcid.org/0000-0003-4982-7490),
А.Я. Білецький[1] (https://orcid.org/0000-0002-3798-8150),
О.В. Мележик[2] (https://orcid.org/0000-0003-3882-7102),
О.О. Гончар[3] (https://orcid.org/0000-0003-4134-03544)

[1] Національний авіаційний університет,
просп. Любомира Гузара, 1, Київ, 03058, Україна,
+380 44 406 7130, post@nau.edu.ua
[2] Відкритий міжнародний університет розвитку людини «Україна»,
вул. Львівська, 23, Київ, 03115, Україна,
+380 67 328 2822, office@uu.edu.ua
[3] Інститут фізіології ім. О.О. Богомольця Національної академії наук України,
вул. Богомольця, 4, Київ, 01024, Україна,
+380 44 256 2400, science@biph.kiev.ua

ІНОВАЦІЙНІ ПІДХОДИ ДО КОНСТРУЮВАННЯ БІОІНФОРМАЦІЙНИХ СИСТЕМ З БАЗАМИ ДАНИХ: КЛЮЧІ НА ОСНОВІ ГЕНЕТИЧНИХ КОДІВ

Вступ. Застосування досягнень біології, біофізики у техніці відкриває нові можливості для інновацій, зокрема у технологіях створення реляційних 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 актуальними є розробка інноваційних підходів до конструювання біомедичних реляційних БД з ключами із розширеними можливостями.

Мета. Розробка та конструювання біомедичних реляційних 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'єктно-орієнтованого системного аналізу для побудови оптимальним чином БД з біомедичною інформацією, метод проєктування ER-діаграм, методи конструювання БД.

Результати. На прикладі реляційної БД з інформацією про деякі види риб розглянуто, застосовано та описано підхід об'єктно-орієнтованого аналізу для оптимального конструювання БД, описано алгоритм їх конструювання. Особливу увагу приділено вирішенню проблеми створення ключів на основі генетичних кодів риб у буквено-цифровому вираженні, особливо як первинних ключів, що забезпечують зв'язки між окремими таблицями БД, цілісність інформації у системі, надійність доступу до неї. Проаналізовано й 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 й біоінформаційних. Вони мають як теоретичне значення для подальшого розвитку технологій розробки БД, так і практичне, удосконалюючи 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, бази даних, первинний ключ, кодування, біомедична інформація, хімічні речовини.