

Биол. журн. Армении, 4 (55), 2003

УДК 620.193.8:681.3.016

USER FRIENDLY INTERFACE FOR MICROBIAL NOMENCLATURAL DATABASE

H.V. ASTSATRYAN, E.G. AFRIKIAN

State Microbial Depository Centre of NAS of Armenia, 378510, Abovian city

The creation of user friendly interface for microbial nomenclatural Database (DB) as well as the hierarchical cluster analysis as a general approach for revealing of groups and members similarity coefficient. A programming environment for users has been worked out functioning via WEB interface.

Հոդվածում դիտարկվում են միկրոօրգանիզմների անվանակարգային տվյալների հենքերի համար "օգտագործողին մոտ" միջավայրի մշակումը, ինչպես նաև հիերարխիկ կլաստերային անալիզը (ծառի մեթոդ) որպես գլխավոր գործիք գտնելու միկրոօրգանիզմների խմբերի և անդամների միջև հարաբերակցությունը: Այն իրականացնելու համար ստեղծվել է ծրագրային փաթեթ, որը աշխատում է WEB ինտերֆեյսի օգնությամբ:

Обсуждаются подходы по разработке дружественного интерфейса для пользователей номенклатурной базой данных микроорганизмов, а также использования иерархического кластерного анализа как главного инструмента выявления взаимосвязи между разными группами и объектами БД. Для реализации этого создан программный пакет, функционирующий WEB интерфейсом.

Microbial Databases - WEB interface - Programming Packages

The information technologies have a wide use in microbiology initiated by a huge amount of data to be treated. The implementation of it is of crucial importance for generalization of data for creation of microbial Culture Collections. The State Microbial Depository Centre (RCDM) of the National Academy of Sciences of the Republic of Armenia, established in 1993 represents the National Culture Collection of non-pathogenous microbial strains of scientific and industrial importance and at the moment maintains more than 10.000 strains. Microbial strains preserved are well-defined and supplied by detailed information on their characteristics, related to ecological, morpho-physiological, biochemical and biosynthetic patterns.

As a result of preceding researches a functional DB of microbial cultures has been developed for integration of the data kept and their interrelationships [3].

The MySQL Database programming package as DB server has been used as very fast, multi-threaded, multi-user, and robust SQL (Structured Query Language). The WEB interface for dialog between users and DB server has been applied. A managing programming package by PHP [4] language has been created. The Apache Web server has been used as a powerful, flexible, HTTP/1.1 compliant web server. PHP represents a robust open-source development language providing with tools and

flexibility to accomplish virtually any task formulated. PHP is an embedded language. It means that developers can jump between raw HTML code and PHP without sacrificing readability.

For classification of the properties of microorganisms in DB the RKC (corresponding names of the authors Rogosa, Krichevsky, Colwell) coding system was used. RKC represents the description-codes and characterizes the features (morphology, biochemistry, physiology, etc.) of microorganisms. The RKC coding is recommended by number of international organizations (CODATA, WFCC, ICSU, MSDN) as a commonly accepted standard.

The microbial DB consists of two subdatabases (Figure 1). The Secondary subdatabase serves to the Main subdatabases. The tables of Secondary subdatabase are used for Main subdatabase. The Main Subdatabase has the following format: ID (identification): Genus: Group: Species: Subspecies: RKC_code: RKC_value: Worker1: Worker2: User: Date, where main Subdatabase tables consist of following fields:

Genus: Genus Rubricator (unsigned integer, length 3)

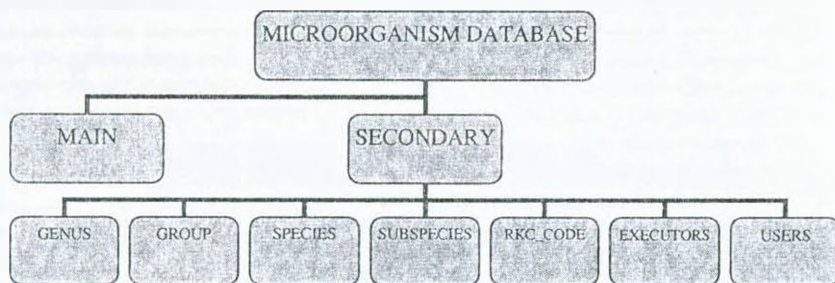


Fig. 1. The structure of microbial nomenclatural DB.

Group: Group Rubricator (unsigned integer, length 3)

Species: Species Rubricator (unsigned integer, length 3)

Subspecies: SubSpecies Rubricator (unsigned integer, length 3)

RKC_code: RKC code of microorganism (unsigned integer, length 6)

RKC_value: RKC code value of microorganism (unsigned integer, length 1)

Researcher: Researcher code (unsigned integer, length 3)

Compiler: Compiler code (unsigned integer, length 3)

Users: Web User code (unsigned integer, length 3)

Date: Input date of object (DateTime)

A cluster analysis for the screening of groups and objects similarity between microorganisms studied has been used [5]. Cluster analysis is a collection of statistical methods for identification of groups and objects that behave similarly or show similar characteristics. The simplest mechanism is the partition of the objects what defines the measurements of capturing similarity or distance between the objects. In this way, clusters and groups are interchangeable words. The clustering algorithms are broadly classified into two named hierarchical and non-hierarchical algorithms [1, 2]. In case of hierarchical procedures, possible to create a hierarchy or tree-like

structure to determine the relationship among entities. In the non-hierarchical method a position for the measurement is taken as central place and distance for measurement from such central point. Identification of right central position is a big challenge and hence non-hierarchical methods are less popular.

For realizing of the cluster analysis on a set of microorganisms, the web user first needs to establish such a unit. Users can determine a set of microorganisms taking into account the following properties: Genus, Group, Species, Subspecies, RKC_code, RKC_value, Researcher, and Compiler. Web user can give an array of values for each property of microorganism. The result of this operation is a matrix. The length of this matrix is $n \times m$, where n is a quantity of microorganisms and m is a quantity of RKC codes. The PHP distance() function transforms a matrix to a distance matrix. The function calculates the distance between given n microorganisms and so on, until the distances between all the pairs have been calculated. If we see the rows of the microorganism matrix as vectors, then each variable occupies a position in a four dimensional spaces. Thus, possible to calculate a distance between any two variables applying arithmetic vector.

In case of error the program returns error message. Then the output of distance () function becomes as input stream for draw () PHP function. This function analyses the whole structure of distance matrix and creates gif format picture in temporary directory. The managing programs visually show it on the web page. In programming environment authors use PHP, Java and HTML programming languages.

The development of Intranet, permits to realize the connection of such networks to the Internet occurs an opportunity from any workplace of organization to receive access to an information resource of the network. Taking into account the abovementioned the Web management environment (programming package) for communication links between users and DB has been created. The programming package developed is successfully functioning in the RCDM ensuring well-exploitation parameters for Culture Collection's service. Based on the results obtained the DB of synthetic polymers has to be developed. This work is done within the framework of ISTC (International Science and Technology Center) A-092.2 Project.

REFERENCES

1. *Kusiak A.* "Data Analysis: Models and Algorithms", Intelligent Systems Laboratory, 4312 Seamans Center, The University of Iowa, 2001.
2. *Mahadevan S.* "The Junction Tree Algorithm", University of Massachusetts, CMPSCI 691T: Week2 – p. 1, 1999.
3. *Pogosyan A.G., Astsatryan H.V., Adamyan M.O., Afrikian E.G.* "Relational Database for microbial degradants", V International Scientific-practical Conference, "Modern Problems of Biological Degradation of Materials (BioDegradation - 2002)", Penza, 127-130, 2002.
4. Source page: <http://www.php.net>
5. *Yaduhisa H., Takeuchi A., Inada K.* Journal of Classification, 16, 1, 1999.

Поступила 25.IX.2003