Krzysztof NOWAKOWSKI, Przemysław J. NOWAK, Robert J. TOMCZAK, Sebastian KUJAWA, Grzegorz PRZYBYLSKI
Poznań University of Life Sciences, Institute of Biosystems Engineering
ul. Wojska Polskiego 50, 60-637 Poznań, Poland
e-mail: krzysztof.nowakowski@up.poznan.pl

DATABASE SYSTEM FOR SUPPORTING INNOVATIVE METHODOLOGY FOR PINE PROTEIN GENOTYPING

Summary

Project "Database system for supporting innovative methodology for pine protein genotyping" was aimed to create a system of representing information about the trees used in selective breeding carried out by Polish State Forests. Presented information includes data on general characteristics, such as geographic location and special features, such as genetic variation.
The project is based on Microsoft technologies:
- *ASP.NET to build a website with all the functionality*
- *SQL Server 2008 responsible for database platform.*

Key words: *Scots pine (Pinus sylvestris L.) database system, genotyping*

SYSTEM BAZODANOWY WSPOMAGAJĄCY INNOWACYJNĄ METODYKĘ GENOTYPOWANIA BIAŁKOWEGO SOSNY ZWYCZAJNEJ

Streszczenie

Projekt pt. „System bazowy wspomagający innowacyjną metodykę genotypowania białkowego sosny zwyczajnej" miał na celu wytworzenie systemu prezentacji informacji na temat drzew wykorzystywanych w hodowli selekcyjnej prowadzonej przez Lasy Państwowe. Przedstawiona informacja zawiera dane na temat cech ogólnych, takich jak położenie geograficzne i cech specjalistycznych, jak zmienność genetyczna.
Projekt powstał w oparciu o technologie firmy Microsoft:
- *ASP.NET stworzenie strony wraz z całą funkcjonalnością*
- *SQL Server 2008 program odpowiadający za platformę bazodanową.*

Słowa kluczowe: *sosna zwyczajna (Pinus sylvestris L.), system bazodanowy, genotypowanie*

1. Introduction

Using information progress many institutions and research centers are trying to reach out to the society using electronic means creating different types of technology platforms, websites, database systems, information points or other electronic forms. Main task is to facilitate the society in everyday life.

Rapid flow of information is used in many specific areas of expertise such as forestry [6]. The presented project is an example of the opportunities offered by information technology in mutual cooperation with other specializations , including a project with forest specialists. The project was carried out in the framework of their activities and Bioengineering Department of Agriculture and Forestry Research Institute. Produced by the project web-based system provides information about maternal trees (pines), which are used to the selective breeding of Polish State Forests (PSF). The information has been provided both general and highly specialized such as age or the basis of genetic variation in DNA. The project is in the assessment of the authors an innovative idea, because to this day in Poland wasn't analyzed similar issues. The described program is a proposal to fill information gaps culture of PSF.

During work authors generated database system containing information concerning maternal trees delivered for the demand of future users. The database was created using Microsoft SQL Server 2008. The next step in developing the system was to create a website with an accessible and intuitive interface. Project was created in Microsoft Visual Studio 2008, ASP.Net Framework 3.5 technologies.

In order to meet program demands a system must meet the following terms:
- Represent laboratory data.
- Display of topographic information.
- Filtrate individual components of the data in the system.
- Add records to the database.
- Edit existing records.

Creation of a system adapted to the needs of future users bridged measurable benefits of accelerating the flow of information which probably contributes to the increased intensity of research in the field of molecular genetics.

2. Material and methods

Today breeding selection is based on improving acquisition of seeds acquired from the selection objects, which are: maternal trees and seed plantations. These selection objects currently are chosen on the basis of observations of the external characteristics of a tree, while foresters estimate genetic gain only on the basis of the phenotype (appearance) of plants. It was found, however, that there is a need to increase the scope of the information available on process of selecting mother trees or creating

new seed plantations. Expected additional data is information about genetic variation, tested by the molecular markers in laboratory. The results of this research will form the basis for creating future artificial breeding population for the foreseeable genetic variation. This idea is a long-term project and the first stage of the operation requires a system for cross-correlation studies of research units involved in the project.

Technological tool that allows correlation between scientific work and focuses the basic trend of the research is proposed in the project database. Modern science is based on the molecular analysis [1], which enter the forest sciences [3, 6], occupying one of the most important positions in the subject [8]. Genetics is based on the methods of analyzing of several fragments of the organism genome, including isoenzymes (indirect method), and microsatellites (STR) (direct method). In the project we represent report the results of the two methods of genetic analysis. They were chosen because they were selected using the knowledge in the field of population genetics. We know that the most complete genetic information can be delivered by combination of isoenzymatic protein electrophoresis and DNA sequencing based on microsatellites [1, 5].

The purpose of the project was to create a database system which is intended to represent information about the trees used in selective breeding carried out by the Polish State Forests [4]. The information includes data on general characteristics, such as geographic location and special features, such as genetic variation. Due to the scope of accumulated information the developed system is the first computer system of this type in the country.

Information presented in the project are divided in the data tables connected to each other through relationships. This assumption saves disk space and time in database work usage. The primary practical objective was to create a system that will meet objectives resulting from the present state of knowledge on the topic, there will not be limited in the future development and further expansion of the project.

The goal was achieved through the use of the module, which makes possible to use form able to editing and adding records within database tables.

The process database system creation consists of the following steps:
1. Consultation with the scientific staff of the Forest Research Institute (FRI), in order to determine the data contained in the database and the relationships between them.
2. Designing of database schema structure - the structure of the database has been developed on the basis of pictograms.
3. Gathering required information necessary to fill the database which data and data archivisation.
4. Creation of the database in Microsoft SQL Server
5. Completing the database using the SQL Server Management Studio.
6. Create a website using ASP.Net 3.5 toolset.
7. Consultation with the researchers from Forest Research Institute focused on the proposed interface and its functionality.
8. Incorporate amendments in the proposed interface - changes in the appearance of the proposed interface.

In the project authors used Microsoft SQL Server 2008. Figure 1 illustrates the ongoing relationships between tables in situation when user gave disposition to show: number of trees, species name, rdsf names, names of districts, the names of forestry, precinct and ward. Creation of functional, in accordance with the application requirements caused a need to design from scratch for each of the elements of the system. Visualization of the posted results was based on master pages mechanism [7, 10].

Public Zone (Fig. 3) is the area of the application that is available to all users without need to log in. From this level user is able to characterize the tree view and the allocation of maternal administration. The form by which you can filter the data in the database is shown in Figure 3.
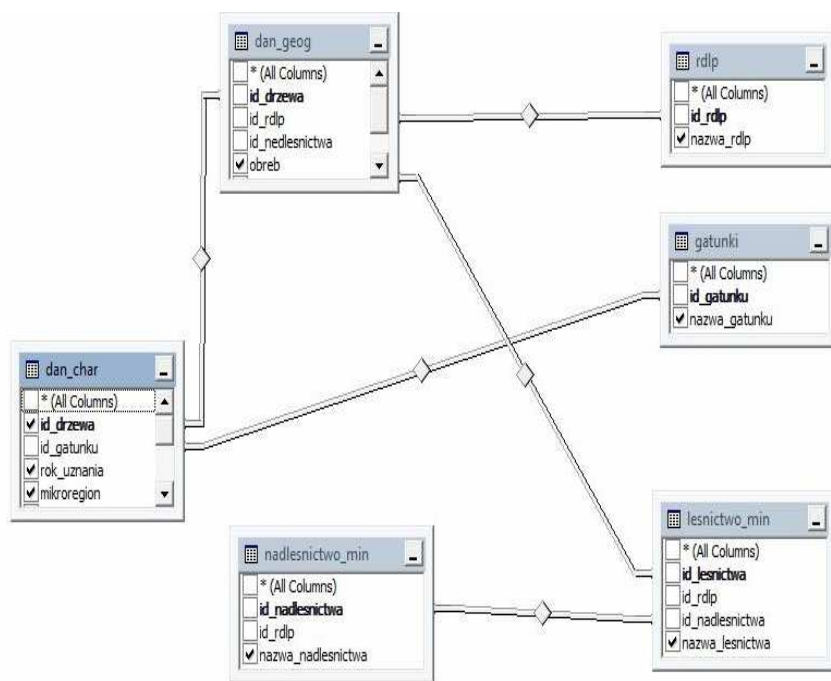
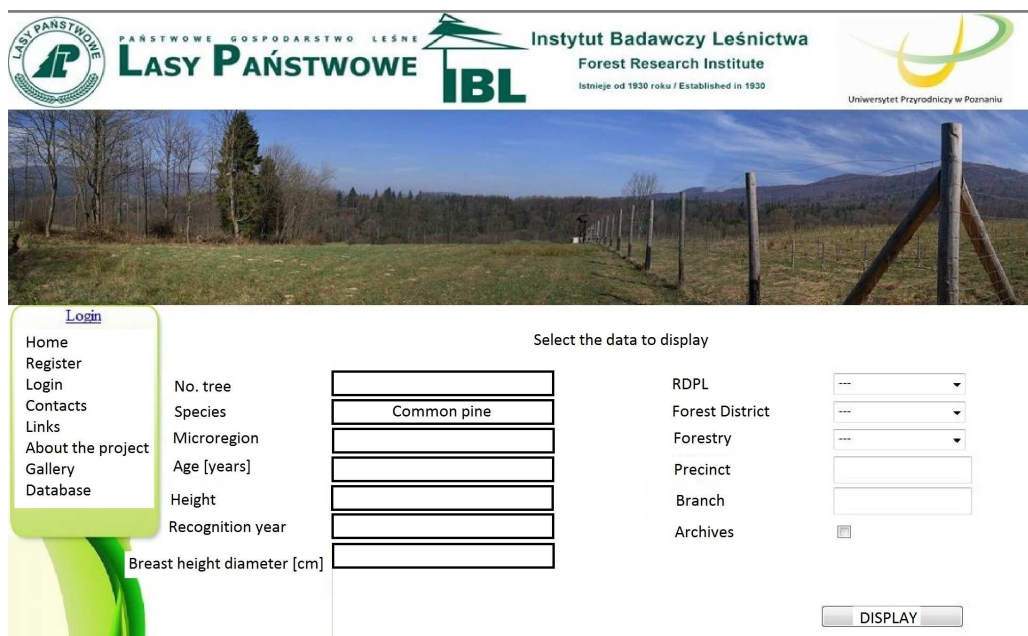

Source: own work / Źródło: opracowanie własne

Fig. 1. Examples of the relationships between tables in database
Rys. 1. Przykładowe relacje zachodzące pomiędzy tabelami w bazie danych
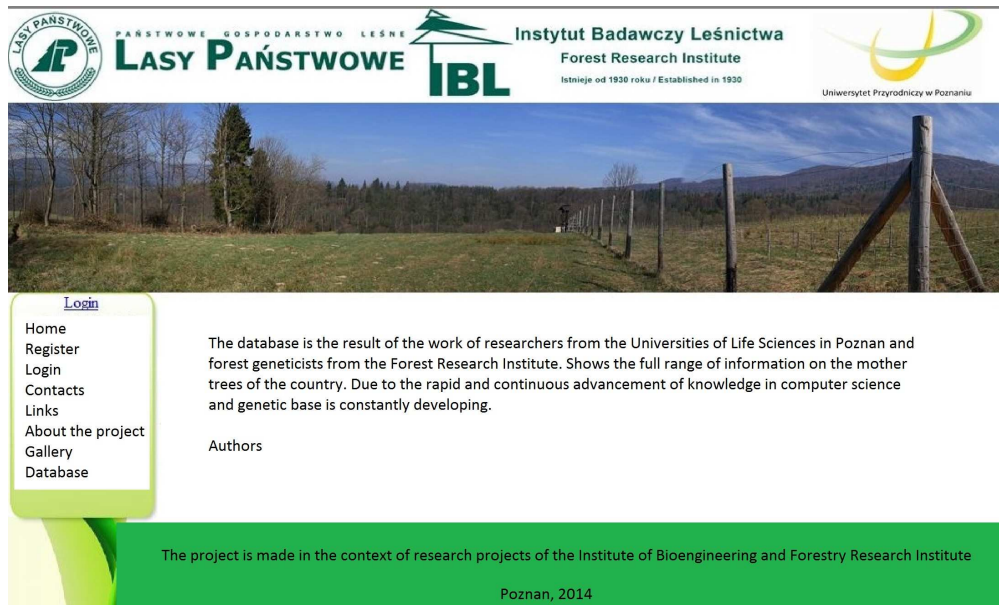
Source: own work / Źródło: opracowanie własne

Fig. 2. View of the main application window
Rys. 2. Widok okna głównego aplikacji

Source: own work / Źródło: opracowanie własne

Fig. 3. The window for selecting mother trees for the public zone
Rys. 3. Okno umożliwiające wybór drzew matecznych dla strefy publicznej

The employee module allows to view data about mother trees, just as it is in the public, but are also provides genetic information about trees. If user is logged with the rules of the employee also has the ability to obtain information about the origin of vegetative seed orchards and its generative origin. The data contained in the database relating to seed plantations cover a range of information such as:

1. The number of plantations.
2. Species occurring trees.
3. Year of establishment.
4. Area [ha].
5. Allocation of the Administration.
6. The number of mother trees from which reproductive material was collected.

A user with the assigned Administrator role in addition to browse the database has the ability to edit it [2, 9]. Opportunities user with granted administrator rights are shown on Figure 4. Deployment diagram of system components is shown in Figure 5. It identifies the computer equipment used in the system, which previously specified software components of the system are installed.

3. Summary and conclusions

The work was focused on developing innovative database system supporting protein genotyping methodology that has allowed to provide the following conclusions:
1. Developed database can serve as a system for the exchange of information between the IBL and other research centers.
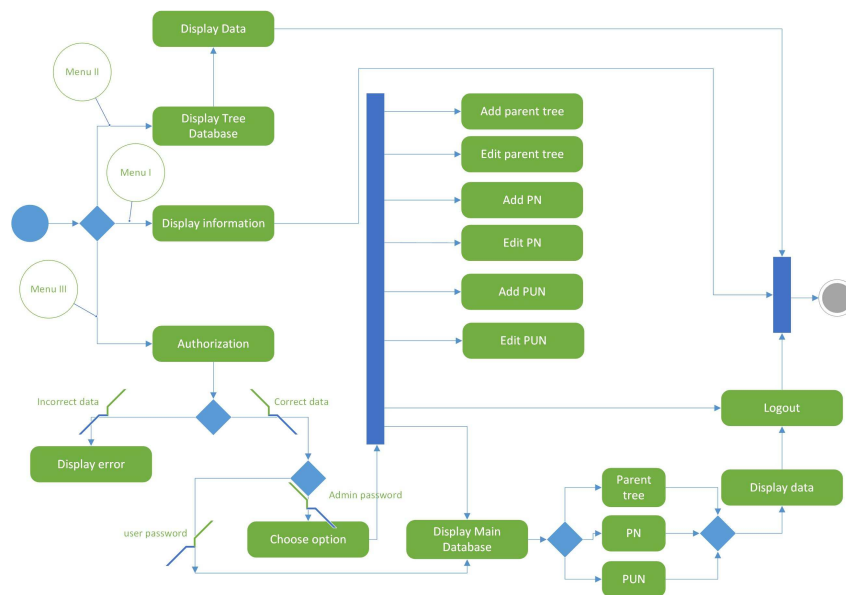
Source: own work / Źródło: opracowanie własne

Fig. 4. Diagram of activities
Rys. 4. Diagram aktywności
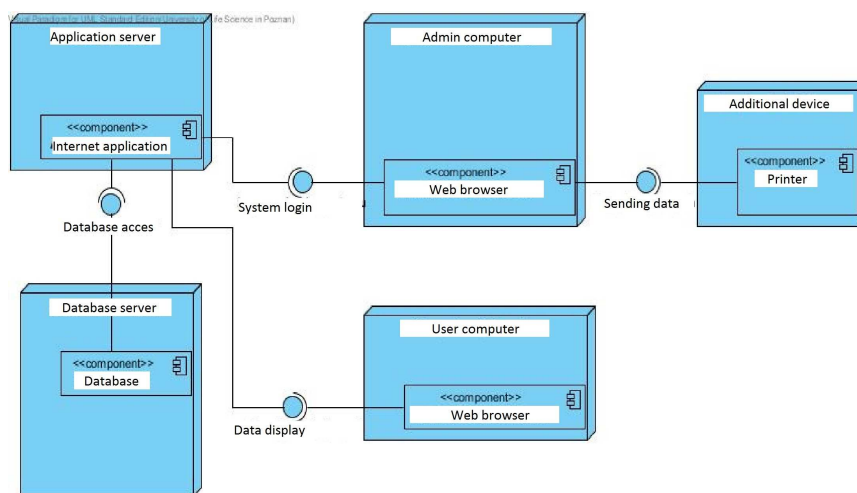
Source: own work / Źródło: opracowanie własne

Fig. 5. Diagram of system components
Rys. 5. Diagram rozlokowania komponentów

2. It was noted that in the field of forestry, there is no information systems supporting the development of this field of science.

3. The Polish State Forest data system needs to be updated. Particularly with regard to:

a) It was noted during data collection, the lack of standardization of the interface within a single company and the difficulty of obtaining data.

b) The complexity of obtaining information from the selected districts especially in the afternoon.

4. References

[1] Avise J.C.: Markery molekularne historia naturalna i ewolucja. Warszawa: WUW, 2008.
[2] Barker R., Longman C.: CASE*Method. Modelowanie funkcji i procesów. Warszawa: WNT, 2001.
[3] Berecka M.: Genetyka populacyjna drzew leśnych. Kraków: Wyd. Ak. Rol., 2006.
[4] Fonder W., Matras J., Załęski A.: Leśna baza nasienna w Polsce. Warszawa CILP (1-300), 2007.
[5] Fonder W.: Ochrona zasobów genowych rodzimych gatunków drzew i krzewów leśnych w Lasach Państwowych. Wydawnictwo Świat. Biblioteczka Leśniczego, 2001, z. 151.
[6] Matras J., Fonder W.: Założenia „Programu ochrony leśnych zasobów genowych i hodowli drzew leśnych w Polsce na lata 2011-2035. DGLP, SITLiD, Biblioteczka Leśniczego 2006, nr 234 (1-15).
[7] Muller R., J.: Bazy danych – język UML w modelowaniu danych. Warszawa: Mikom, 2000.
[8] Nowakowska J.: Zmienność genetyczna polskich wybranych populacji sosny zwyczajnej (*Pinus sylvestris* L.) na podstawie analiz polimorfizmu DNA. Prace Instytutu Badawczego Leśnictwa. Rozprawy i Monografie, Sękocin Stary, 2007.
[9] Sommerville I.: Inżynieria oprogramowania. Warszawa: WNT, 2003.
[10] Wrycza S., Marcinkowski B., Wyrzykowski K.: Język UML 2.0 w modelowaniu systemów informatycznych. Gliwice: Helion, 2006.